Recent Trends in Intensive Computing

Edited by
M. Rajesh
K. Vengatesan
M. Gnanasekar
Sitharthan R
A.B. Pawar
P.N. Kalvadekar
P. Saiprasad

39

ADVANCES IN PARALLEL COMPUTING

IOS Press

Recent Trends in Intensive Computing

In a world where computer science is now an essential element in all of our lives, a new opportunity to disseminate the latest research and trends is always welcome.

This book presents the proceedings of the first International Conference on Recent Trends in Computing (ICRTC 2021), which was held as a virtual event on 21 – 22 May 2021 at Sanjivani College of Engineering, Kopargaon, India due to the restrictions of the COVID-19 pandemic. This online conference, aimed at facilitating academic exchange among researchers, enabled experts and scholars around from around the globe to gather for the discussion of the latest advanced research in the field despite the extensive travel restrictions still in place. The book contains 134 papers selected from 329 submitted papers after a rigorous peer-review process, and topics covered include advanced computing, networking, informatics, security and privacy, and other related fields.

The book will be of interest to all those eager to find the latest trends and most recent developments in computer science.

RECENT TRENDS IN INTENSIVE COMPUTING

Advances in Parallel Computing

Parallel processing is ubiquitous today, with applications ranging from mobile devices such as laptops, smart phones and in-car systems to creating Internet of Things (IoT) frameworks and High Performance and Large Scale Parallel Systems. The increasing expansion of the application domain of parallel computing, as well as the development and introduction of new technologies and methodologies are covered in the *Advances in Parallel Computing* book series. The series publishes research and development results on all aspects of parallel computing. Topics include one or more of the following:

- Parallel Computing systems for High Performance Computing (HPC) and High Throughput Computing (HTC), including Vector and Graphic (GPU) processors, clusters, heterogeneous systems, Grids, Clouds, Service Oriented Architectures (SOA), Internet of Things (IoT), etc.
- High Performance Networking (HPN)
- Performance Measurement
- Energy Saving (Green Computing) technologies
- System Software and Middleware for parallel systems
- Parallel Software Engineering
- Parallel Software Development Methodologies, Methods and Tools
- Parallel Algorithm design
- Application Software for all application fields, including scientific and engineering applications, data science, social and medical applications, etc.
- Neuromorphic computing
- Brain Inspired Computing (BIC)
- AI and (Deep) Learning, including Artificial Neural Networks (ANN)
- Quantum Computing

Series Editor:
Professor Dr. Gerhard R. Joubert

Volume 39

Recently published in this series

Volumes 1–14 published by Elsevier Science.

Recent Trends in Intensive Computing

Edited by

M. Rajesh

Department of Computer Science and Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra. India

K. Vengatesan

Department of Computer Science and Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra. India

M. Gnanasekar

Department of Computer Science and Engineering, Sri Venkateswara College of Engineering Chennai, Tamil Nadu, India

Sitharthan.R

School of Electrical Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, India

A.B. Pawar

Department of Computer Science and Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra. India

P.N. Kalvadekar

Department of Computer Science and Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra. India

and

P. Saiprasad

Department of Computer Science and Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra. India

Preface

We are glad to introduce you to the 2021 1St International Conference on Recent Trends in Computing (ICRTC 2021), which was successfully held on May 21–22. Different from the previous four times, ICRTC 2021 was carried out in the form of a virtual conference due to the impact of COVID-19. Because there is a worldwide travel restriction, we held this flexible online conference to gather experts and scholars around the globe with the aim to continue disseminating the latest advanced research in the field of Recent Trends in Computing and developing the academic exchange among researchers.

The committee of ICRTC 2021 consists of about hundred and thirty-four experts in the area of Computing Science from home and abroad. With its professional and powerful influence, we are honored to have invited five renowned experts as our keynote speakers. Prof. Fadi Al-Turjman, is a Professor at Near East University Applied Artificial Intelligence Research Centre, Turkey., held a speech on recent computing trends in artificial intelligence. In addition, Mr.Nimit Shishodia, CEO, Ecode Networks, UK., performed his speech entitled A 5G and Research Opportunities. Dr. Parikshit Mahalle, BOS Member, SPPU Professor & Head, Computer Engg. Dept, SKN COE, Pune., performed his speech entitled Future of Data Science. Mr. P. Vinoth Kumar, Software Quality Analyst, Glencore Information Services Pvt Ltd, UK., performed his speech entitled Emerging Trends in the area of Extended Reality. In addition, Mr. R. Sairam, CFO, Cybertheronai Pvt. Ltd, Singapore., performed his speech entitled Security using ML and ELK. The keynote speeches were the first session of the conference, which was each delivered in about 30 minutes via Google meet.

We are glad to share with you that we still received lots of submissions from the conference during this special period. Hence, we selected a bunch of high-quality papers and compiled them into the proceedings after rigorously reviewing them. These papers feature the following topics but are not limited to Advanced Computing, Networking, Informatics, Security and Privacy, and other related fields. All the papers have been through rigorous review and process to meet the requirements of international publication standards.

We would like to express our sincere gratitude to the Shri Amit Kolhe, Managing Trustee, Dr. A. G. Thakur, Director, and Dr. D. B. Kshirsagar, Professor and Head, Department of Computer Engineering Sanjivani College of Engineering, Kopargaon, India, the distinguished keynote speakers, as well as all the participants. We also want to thank the publisher for publishing the proceedings. May readers gain some valuable knowledge from the book. We are expecting more and more experts and scholars from all over the world to join this international event next year.

<div align="right">

The Book Editors
Prof. Dr. M. Rajesh, and Prof. Dr. K. Vengatesan
Department of Computer Science and Engineering
Sanjivani College of Engineering, Kopargaon, Maharashtra. India.

</div>

Prof. Dr. J. M. Gnanasekar,
Department of Computer science and engineering
Sri Venkateswara College of engineering Chennai, Tamil Nadu, India

Dr. Sitharthan R
School of electrical engineering,
Vellore Institute of Technology, Vellore, Tamilnadu, India.

Dr. A. B. Pawar, Dr. P. N. Kalvadekar, and Dr. P. Saiprasad
Department of Computer Science and Engineering
Sanjivani College of Engineering, Kopargaon, Maharashtra. India.

Contents

An Image Encryption Technique Based on Logistic Sine Map and an Encrypted Image Retrieval Using DCT Frequency

Rajana Kanakaraju [a, 1], Lakshmi V [b], Shanmuk Srinivas Amiripalli [c], Sai Prasad Potharaju [d], R Chandrasekhar [b]

aDepartment of Computer Science, Gayatri Vidya Parishad College for Degree and PG Courses, M.V.P Campus, Visakhapatnam, AP, India

bDepartment of Computer Applications, Gayatri Vidya Parishad College of Engineering(A), Madhurawada, Visakhapatnam, AP, India

cDept of CSE, GITAM Univeristy, Visakhapatnam, AP, India
dDept Computer Engineering, Sanjivani College of Engineering, Kopargaon, MH, India

Abstract. In this digital era, most of the hospitals and medical labs are storing and sharing their medical data using third party cloud platforms for saving maintenance cost and storage and also to access data from anywhere. The cloud platform is not entirely a trusted party as the data is under the control of cloud service providers, which results in privacy leaks so that the data is to be encrypted while uploading into the cloud. The data can be used for diagnosis and analysis, for that the similar images to be retrieved as per the need of the doctor. In this paper, we propose an algorithm that uses discrete cosine transformation frequency and logistic sine map to encrypt an image, and the feature vector is computed on the encrypted image. The encrypted images are transferred to the cloud picture database, and feature vectors are uploaded to the feature database. Pearson's Correlation Coefficient is calculated on the feature vector and is used as a measure to retrieve similar images. From the investigation outcomes, we can get an inference that this algorithm can resist against predictable attacks and geometric attacks with strong robustness.

Keywords. Cloud Computing, confidentiality, encryption, decryption, feature vector, Pearson's correlation, discrete cosine transformation, image retrieval, logistic sine map.

1. Introduction

Nowadays, cloud computing technology is mostly used by many medical institutions to store medical data. Images occupy more space when compared to text messages and become a big issue if the data need to be shared, then the medical institutions are using

---

[1] Lakshmi V,Department of Computer Science, Gayatri Vidya Parishad College for Degree and PG Courses, Email: lakshmi.vkl@gvpce.ac.in.

untrusted third-party cloud platforms to store and share medical data, nevertheless Whenever the user stores their image data in the third-party cloud server, the users no Longer have the direct control over the data, which makes them tense about their privacy and confidentiality [1-4]. Thus, before storing data on the cloud server, the data should be converted into a not understandable manner. To encrypt images there are several approaches, the encryption algorithms in the spatial domain are instinctive and easy to implement and which makes use of the matrix structures of the image data, but due to high algorithm complexity and distortion of the image makes these encryption algorithms inefficient. In comparison, the encryption algorithms in the frequency domain are quite steady and can counterattack interfering. A combination of discrete cosine transformation and chaotic maps are used to encrypt images, there are many variants of chaotic maps, the most popular chaotic map used is logistic map but this map has lost its importance due to its drawbacks such as the non-uniform distribution of output sequences and the chaotic sequences exist only in a limited range, these drawbacks can be covered by another variant of a chaotic system called logistic sine map system which is a non-linear grouping of two different one dimensional chaotic maps [5].

To diagnose a disease, the doctor needs to do so much analysis. In some cases, the doctor uses medical images of past medical cases to do more precise diagnosis and also to take proper medical care. Hence, it is essential to have an effectual and operative medical image retrieval system. Old-style methods for retrieving similar images are meta search retrieval and image retrieval based off of content (CBIR), which mainly uses features as content, in particular text [6], color [7], composition, shape [8], and texture [9] as a feature vector. Feature vector plays a crucial role in image retrieval. However, the data in the cloud platform is encrypted, so that CBIR technology is not helpful in the image retrieval process as there is no use if the content is decrypted before retrieving and also increases a lot of unnecessary calculations. After a while, there is a novel and homomorphic encryption [10-12]. Homomorphism encryption is used in favor of doing arbitrary computations over the encrypted medical image data, with the identical outcome as the novel calculation of the original medical image [13-14].In the present study, we are proposing an algorithm for Encrypted Medical Image Retrieval which uses of Discrete Cosine Transformation (DCT) and Logistic Sine map. The algorithm uses DCT transformation with a Logistic sine map for encryption and uses Pearson's Correlation Coefficient between encrypted images to retrieve similar images [15-20]. The algorithm can resist against conventional and geometric attacks with strong robustness.

2. Preliminaries

Discrete Cosine Transform is used to translate the matrix data into a summation of a sequence of cosines fluctuating at different frequencies. It is like Discrete Fourier Transform (DFT), but discrete cosine transformation uses of just cosines and real coefficients. In contrast, Fourier transformation uses both cosines and sines and also requires complex number coefficients. Figure 1 Discrete cosine transformations simple to compute. Both discrete Fourier transform and discrete cosine transformation converts data into the frequency domain from a spatial domain, and the inverse

functions of DCT and DFT transform data into the original data. After the DCT transforms, the low-frequency part contains a highly concentrated attribute of energy and natural signals. The mathematical expressions for Forward DCT transformation and Inverse DCT transformation are as defined in Eq (1) and Eq (2) respectively. The pixel value in the frequency domain and spatial dominate represented by F(u, v) and f (x, y) respectively. The medical image size is represented by M and N. In Logistic Sine Maps Chaotic maps are highly used in image encryption because of their non-convergence, chaotic property, input sensitivity, and state ergodicity. The complex chaotic behavior is generated by using simple dynamic equations. The logistic sine map is a 1D chaotic map. It is a non-linear grouping 1D Logistic map and 1D Sine map. The equation to represent the Logistic Sine map, defined in equation (3). Where r is the input branch constraint with a scope of [0,4] and Xn is the chaotic output with a range of [0,1].

3. Proposed Algorithm
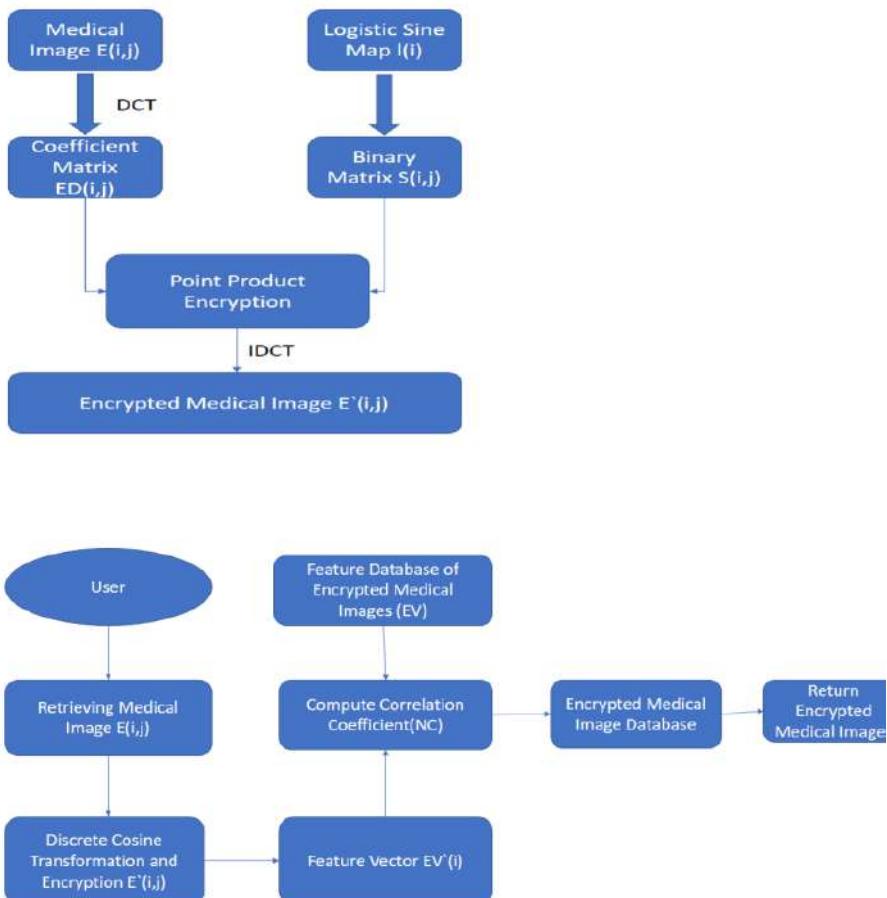
Figure 1. Phases of proposed model

Proposed Algorithm

Step1: *Initially, process the original medicinal image F (i, j) using DCT (Discrete Cosine Transformation), gaining a coefficient matrix of complex number i.e. FD (i, j);*
Step2: *Generate a 1D chaotic vector bl(j) by using logistic sine map after setting the initial value ofX0;*
Step3: *Create a binary matrix S(i,j) which consists of -1 and 1. The 1D chaotic vector bl(j) is used to construct a binary matrix by using a threshold routine sign(l).*

Step4: *Then, do dot product between binary matrix S(i,j) and the complex matrix FD(i,j) which outputs a matrix L(i,j);*

Step5: *To acquire an encrypted image E(i,j), the L(i,j) is processed using IDCT (Inverse Discrete Cosine Transformation).*

Step6: *At first the encrypted image E(i,j) is processed using DCT, then attaining a complex number matrix ED(i,j);*

Step7: *Consider ED``(i,j) upper left corner's 8 X 8 low frequency matrix of ED(i,j);: By using a sign function, create a binary sequence that consists of 1's and 0'sas a feature vector from step2.*

Step 8. *The features FV={FV1, FV2…, FVN} are calculated for encrypted images in the encrypted database E={E1, E2…, EN}.and Upload FV={FV1,FV2…,FVN} into feature database.*
Step 9: *Encrypt the medical image E (i,j) which is uploaded by user, results in an encrypted image E`(i,j) and Extract feature vector FV`(j) for encrypted medical image E`(i,j);*

Step 10: *Calculate Cross-Correlation between FV`(j) and feature vectors in feature database FV(j) using the equation (9) and Retrieve the encrypted medical image, which has the maximum Cross-Correlation value (Near to 1.0)Where mx is the mean of the vector x and my is the mean of the vector y.*

---

4. Results and Discussion

In this study, the model used for simulation is Python 3.7, we pick the MRI information, and the motivation behind the analysis is to examine the algorithm under various attacks and to test whether the retrieval algorithm is retrieving the images or not and also to prove the robustness and homomorphism of the algorithm The outcomes of the experiments are as follows To demonstrate that the feature obtained by the technique mentioned in this paper is a significant feature of the encrypted image. We conducted experimentations among various medical images (made known in figure 2) and their encrypted images (made known in figure 3). By noticing table 1 to table 7 it can be found that the higher the correlation value, higher the similarity, and lower the

correlation value, lower the similarity. It also found that the correlation value between the original medical image and the decrypted medical image is also 1.0 so that the encryption algorithm is lossless and also found that the encryption algorithm is maintaining homomorphic features. Peak Signal to Noise Ratio (PSNR) is broadly used to assess the quality of the picture. In this study, we use PSNR to impartially assess the quality of the picture to be retrieved after the attacks. An experiment is conducted to retrieve encrypted medical images from cloud platforms by passing a medical image to the system.

The system retrieved the images which are having a cross-correlation value of less than 0.5, and then the results obtained are as follows. To prove that the algorithm is retrieving encrypted images even in interference of Gaussian noise, we did experiments on images by increasing the intensity of Gaussian noise to an image. The algorithm has strong anti-Gaussian ability as the image is retrieved even with Gaussian intensity of 20% and the cross-correlation value is 0.86. The results of the experiment are as follows. To prove that the algorithm is retrieving encrypted images against JPEG compression attacks, the experiments were conducted on a medical image while compressing the image up to 50%. The results proved that the algorithm is good enough against the JPEG compression attacks and the results are as follows. The research on median filtering attacks was conducted on a medical image by altering the median filter size and by the number of repeats.

The results have shown that the algorithm has strong sturdiness against median filter assaults. The experiments are conducted up to a size of 3 x3 to 7 x 7. The results of the experiments concluded that the algorithm is strong enough to handle rotation attacks. The experiment goes through various rotations of the image, the retrieval system successfully retrieved the images which are rotated up to 10o. The algorithm has undergone various experiments and proved that it is durable against scaling attacks as the experimental results proved that the retrieval system was successfully retrieved images even with a scaling factor between 0.2 and 4.0. The investigation outcomes proved that the algorithm is tough against cropping attacks and the cropping has done against the medical image on both x and y axes. The cropping ratio for each axis is from 2 % to 20.

Table 1. Cross-correlation between the feature and vector of different encrypted medical images

	a	b	C	d	e		Ea	Eb	Ec	Ed	Ee
a	1.0	0.26	0.03	0.05	-0.12	Ea	1.0	0.19	0.09	0.13	-0.03
b	0.26	1.0	0.35	0.07	-0.29	Eb	0.19	1.0	0.41	-0.12	0.03
c	0.03	0.35	1.0	-0.09	0.06	Ec	0.09	0.41	1.0	-0.28	0.06
d	0.05	0.07	-0.09	1.0	-0.09	Ed	0.13	-0.12	-0.28	1.0	-0.03
e	-0.12	-0.29	0.06	-0.09	1.0	Ee	-0.03	0.03	0.06	-0.03	1.0
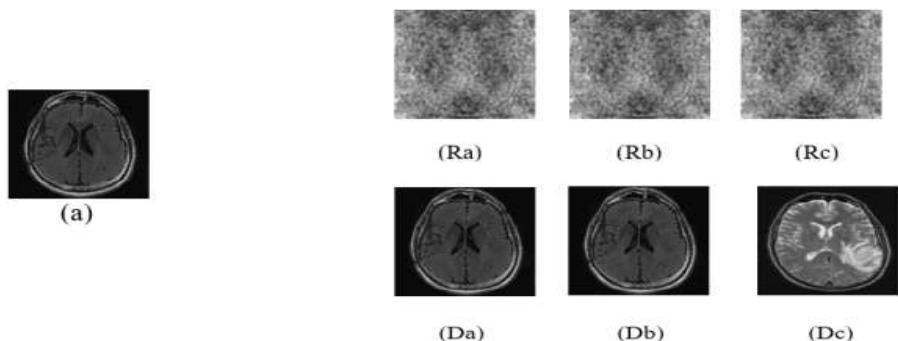
Figure 2. (a) Medical Images (b) Retrieved encrypted medical images (c) decrypted retrieved medical images
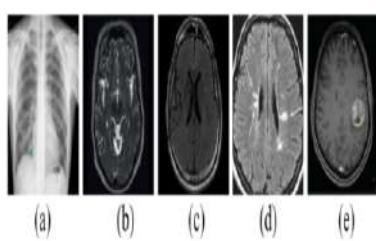
Figure 3. a Medical Images   Figure 3. b Encrypted Medical Images
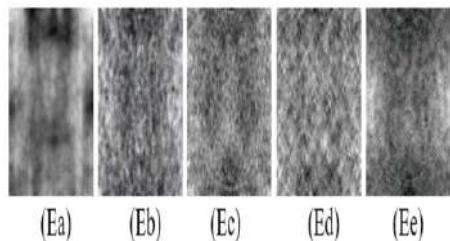
Table 2. Experimental results of Gaussian noise

Noise Intensity	1	4	6	8	12	16	20
PSNR (dB)	21.33	16.64	14.86	13.62	12.91	11.31	10.06
Cross Correlation	0.98	0.95	0.92	0.91	0.89	0.88	0.86

Table 3. Experimental results of JPEG compression attacks

Percentage (%)	5	12	17	22	27	35	42	50
PSNR (dB)	21.22	24.67	25.55	26.56	26.80	27.67	27.73	28.58
Cross Correlation	0.98	0.98	0.97	0.98	0.98	0.98	0.98	0.98

Table 4. Experimental results of median filtering attacks

	Size of Median Filter								
	[3 x 3]			[5 x 5]			[7 x 7]		
Repeating times	2	5	10	2	5	10	2	5	10
PSNR (dB)	28.41	27.46	25.85	25.55	24.71	22.50	23.99	24.32	21.23
Cross Correlation	1.0	1.0	1.0	1.0	0.93	0.90	0.93	0.93	0.85

Table 5. Experimental results of rotation attacks

Degree (o)	-10o	-5o	-1o	0o	1o	5o	10o
PSNR (dB)	14.16	16.01	22.79	103.49	22.80	16.14	14.18
Cross Correlation	0.60	0.74	0.92	1.0	0.90	0.75	0.50

Table 6. Experimental results of rotation attacks

	Ratio (%)	2	4	6	8	10	12	14	16	18	20
X	Cross-Correlation	0.81	0.79	0.78	0.70	0.63	0.47	0.47	0.44	0.37	0.44
Y	Cross-Correlation	0.91	0.84	0.75	0.75	0.79	0.78	0.62	0.59	0.59	0.56

Table 7. Experimental results of scaling attacks

Scaling Factor	0.2	0.4	0.6	1.00	1.2	1.5	2.0	4.0
Cross Correlation	0.75	0.97	0.97	1.00	1.00	1.00	1.00	1.00

5. Conclusion

In this research paper a strong algorithm to retrieve encrypted images is proposed, which is a combination of logistic sine map, 2D discrete cosine transform, and feature extraction in the encrypted field. The investigational efforts demonstrate that this algorithm has perfect sturdiness against Median filtering, cropping attacks, Rotation attacks, Gaussian noise, JPEG compression and scaling attacks. This algorithm is also used to secure medical image data. Moreover, this algorithm has a fast retrieve speed and good operability.

References

[1]   Xia, Z., Xiong, N. N., Vasilakos, A. V., & Sun, X. (2017). EPCBIR: An efficient and privacy-preserving content-based image retrieval scheme in cloud computing. Information Sciences, 387, 195-204.
[2]   Index, G. W. (2013). Instagram tops the list of social network growth.
[3]   Ferretti, L., Pierazzi, F., Colajanni, M., & Marchetti, M. (2013, August). Security and confidentiality solutions for public cloud database services. In The Seventh International Conference on Emerging Security Information, Systems and Technologies (pp. 36-42).
[4]   Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. Journal of network and computer applications, 34(1), 1-11.
[5]   Zhou, Y., Bao, L., & Chen, C. P. (2014). A new 1D chaotic system for image encryption. Signal processing, 97, 172-182.
[6]   Simpson, M. S., You, D., Rahman, M. M., Xue, Z., Demner-Fushman, D., Antani, S., & Thoma, G. (2015). Literature-based biomedical image classification and retrieval. Computerized Medical Imaging and Graphics, 39, 3-13.
[7]   Chatzichristofis, S. A., & Boutalis, Y. S. (2008, May). CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. In International conference on computer vision systems (pp. 312-322). Springer, Berlin, Heidelberg.
[8]   Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.

[9] Vo, A., & Oraintara, S. (2010). A study of relative phase in complex wavelet domain: Property, statistics and applications in texture image retrieval and segmentation. Signal Processing: Image Communication, 25(1), 28-46.

[10] Marwan, M., Kartit, A., & Ouahmane, H. (2016, October). Applying homomorphic encryption for securing cloud database. In 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt) (pp. 658-664). IEEE.

[11] Liu, J., Han, J. L., & Wang, Z. L. (2016, July). Searchable encryption scheme on the cloud via fully homomorphic encryption. In 2016 Sixth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC) (pp. 108-111). IEEE.

[12] Xu, Y., Gong, J., Xiong, L., Xu, Z., Wang, J., & Shi, Y. Q. (2017). A privacy-preserving content-based image retrieval method in cloud environment. Journal of Visual Communication and Image Representation, 43, 164-172.

[13] Rivest, R. L., Adleman, L., & Dertouzos, M. L. (1978). On data banks and privacy homomorphisms. Foundations of secure computation, 4(11), 169-180.

[14] Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.

[15] Zhang, C., Li, J., Wang, S., & Wang, Z. (2017, June). An encrypted medical image retrieval algorithm based on DWT-DCT frequency domain. In 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA) (pp. 135-141). IEEE.

[16] Naidu, J. L., Gorakala, A. C., & Amiripalli, S. S. (2020). Hash functions and its security for Snags.

[17] Amiripalli, S. S., & Bobba, V. (2019). An Optimal TGO Topology Method for a Scalable and Survivable Network in IOT Communication Technology. Wireless Personal Communications, 107(2), 1019-1040.

[18] Amiripalli, S. S., & Bobba, V. (2019). Impact of trimet graph optimization topology on scalable networks. Journal of Intelligent & Fuzzy Systems, 36(3), 2431-2442.

[19] Jitendra, M. S., Amiripalli, S. S., Kollu, V. V. R., Chowdary, P. R., & Rao, R. V. (2020). Analysis of Airline Connectivity System using Graph Theory. International Journal of Control and Automation, 13(4), 77-84.

[20] Amiripalli, S. S., Kumar, A. K., & Tulasi, B. (2016, February). Introduction to TRIMET along with its properties and scope. In AIP Conference Proceedings (Vol. 1705, No. 1, p. 020032). AIP Publishing LLC.

Speech to Indian Sign Language Translator

Hemang Monga [a, 1], Jatin Bhutani [a], Muskan Ahuja [a] Nikita
Maid[a] and Himangi Pande [a]

[a] *School of Computer Engineering and Technology, MIT*
World Peace University, Pune, India

Abstract. Indian Sign Language is one of the most important and widely used forms of communication for people with speaking and hearing impairments. Many people or communities have attempted to create systems that read the sign language symbols and convert the same to text, but text or audio to sign language is still infrequent. This project mainly focuses on developing a translating system consisting of many modules that take English audio and convert the input to English text, which is further parsed to structure grammar representation on which grammar rules of Indian Sign Language are applied. Stop words are removed from the reordered sentence. Since the Indian Sign Language does not support conjugation in words, stemming and lemmatization will transform the provided word into its root or original word. Then all the individual words are checked in a dictionary holding videos of each word. If the system does not find words in the dictionary, then the most suitable synonym replaces them. The system proposed by us is inventive as the current systems are bound to direct conversion of words into Indian Sign Language on-the-other-hand our system aims to convert the sentences in Indian Sign Language grammar and effectively display it to the user.

Keywords. Speech to text, Sign Language Translation, NLP**.**

1. Introduction

Sign language is used worldwide to reduce the communication gap for people with hearing or speech impairment which depends primarily on sign language for everyday communication; currently, we do not have efficient models that convert text to Indian Sign Language. Proper and effective audio-visual support for oral communication is still lacking. There has been significant progress in detecting sign languages in other countries using computers, but very little research has been done to detect ISL. Most of the work done on this topic has focused more on American Sign Language (ASL) or British Sign Language, but for Indian Sign Language, there are hardly any systems developed. The underlying architecture for most of the systems are based on: Direct Translation, Statistical Machine Translation, Transfer-based Architecture The system focuses on Indian Sign Language. For each word, its relevant video will be gathered. For words that are not present in this dictionary, they would be exchanged by other words having similar meanings, considering the duplication of words and their parts of speech. Each language has its own rules and grammar, and therefore translation between two spoken languages is a tedious task. Furthermore, when one language is a spoken language while the other is a sign language, then the translation complexity increases.

---

[1] Nikita Maid, School of Computer Engineering and Technology, MIT World Peace University, Pune, India
 Email: nikitamaid23@gmail.com.

The users of this system are mainly individuals with hearing impairment, as communication for them in everyday places is very problematic because not everyone understands sign language. Therefore, no conversation can occur between a person who speaks and a person with a hearing impairment. Thus, this system is for people who want to learn sign language or reduce communication gaps.

2.   Literature Survey

In India, there is far less research done on sign language and especially Indian Sign Language. Each state in India has its regional language, and therefore India is also known as a multilingual nation. It is also important to note that each state and its language also have its sign language. As India was once a British Empire colony, there are many similarities between the two nations' sign languages.

The ISLRTC (Indian Sign Language Research & Training Center) seeks to balance the principles of program and application of ISL. A widespread practice for deaf people in India is to learn ASL since it is very easily readable and then later learn sign language. 98 percent of 1.1 million deaf people in India are illiterate. Consistent with oralist philosophy, deaf schools try to intervene at an early stage with hearing aids, but this, in particular, does not apply to the community. There have been multiple attempts to provide a new solution to this root cause by researchers worldwide. The techniques for producing gloss in this system are aligned with the Indian sign language grammar. The TSA's data collection method has been carefully analyzed to determine the strategies used in this project.

The primary concern with the transition from Spoken English to Sign Language is the formation of phrases that help translate by giving the correct ASL grammar phrases. This project implements this by using the rule-based technique for building a set of phrases. Each sentence is converted into tokens, and then for each token, we identify the part of speech with which it binds, after which specific actions are taken (Each word is transformed to its root word, and auxiliary verbs are removed). All the words from Spoken English, which are not a part of ASL Gloss, are removed. Then proper nouns (names of countries, cities, people, etc.) are outputted in a different format to display sign language, and other words are processed according to various rules such as verb correction, etc. Each proper noun is stored as alphabets separated by a hyphen, and each alphabet is shown as a sign on display. For Example, Mumbai (name of the city) is stored as M-U-M-B-A-I, and India, in a similar manner, will be stored as I-N-D-I-A, corresponding to the use of conventional sign language. The system is rigid and can only lead to destructive behavior if there is a human error. The human error can be seen while creating an ASL gloss or while storing datum for each sign manually. We can see an increase in efficiency and reduction in time by using a phrase-based algorithm to eliminate repeating words, but it still takes much time in rendering, and the other issue is portability from ASL to ISL [1].

Proposed a model that uses the Hidden Markov Model to convert speech to English text language. The generated text generates Indian Sign Language Gloss using pre-processing and Wordnet, an American Linguistic Database. Pre-processing methods like punctuation removal, selective stop word removal, tokenization, and stems are used. The similarity measurement is done using Leacock Chodorow similarity. This system has a 68% accuracy however faces an issue when word similarity in Wordnet is more remarkable with another word that does not suit the meaning of the statement or

finds a noun for which no other synonym can be found in the ISL dictionary. It has no provision to remove repetition in the ISL gloss, which leads to redundancy [2].

They are considering the most commonly used NLTK libraries and creating a robust and impactful application. They are first converting audio input into text using Google text to speech API. The proposed system performs tokenization and uses rules to convert English to ISL gloss followed by lemmatization and stemming. Then the resultant output is passed to Hamburg Sign Language Notation System, a database retrieval tool. Further, it is passed to the SIGML representation tool to represent the sign language in the form of avatars. Introducing a communication system for people with disabilities for effective communication. Converts speech (English) into three-dimensional avatar animation displays Hindi (Indian) language symbols instead of GIFs, images, or videos for effective memory management [3].

IBM Model 1: The order of words in the target text may not be a correct sentence in syntax and semantics. In order to make a correct sentence from these frames, the arrangement of lexemes is required. This is done using IBM Model -1. For example, "Passengers are" is converted to "passenger," "requested to" is converted to "please," etc.

IBM Model 2: In the IBM Model-2, the design possibilities of glosses depend on their positions as the words in the sentence reside somewhere according to the syntax. Nevertheless, instead of following the target grammar, we will consider the conditional probability of the word in a sentence concerning other words.

IBM Model 3:The inconsistent name of the targeted sentence will be associated with the invalid source name, e.g., Sometimes, when words are changed using the above models, the whole sentence may not have a meaning, so model 3 helps to arrange the words to make a logical sentence. In the first step, they participate in a translation model where the words of the introductory sentences are translated into corresponding words of the desired language, and the order of these words is converted into a valid target sentence. The rules and grammar of both languages are ignored. The work can also be improved by introducing a sentence-based translation machine as word-based translation can be done successfully in sentences with a few words. The work can be improved by using a sentence-based machine translation process. This research presents an architectural framework of Pakistan sign language recognition tool developed to convert given PSL (Pakistan Sign Language) into voice and text. Using the given tool, sign language can be converted into text using the CNN model and TensorFlow. The system accepts input in the form of an image from the camera and translates the symbol into text and voice. The currently developed translation tool can only translate a few Pakistani sign languages into text and speech because it is an example, and there is a need to build an app or portable device to overcome the barrier of communication while using this user-friendly translation tool. In developing the portable version of the tool, the system on chip devices (SoC) is considered.

3.   System Analysis Proposed Architecture

Algorithmic Design:
1. Pyaudio is used to obtain user speech and convert it to text.
2. The sentence is restructured using stop words removal and lemmatization.

3. Final ISL gloss is generated by converting text in context free grammar.
4. YouTube IFrame API is used to display the output in the web application.
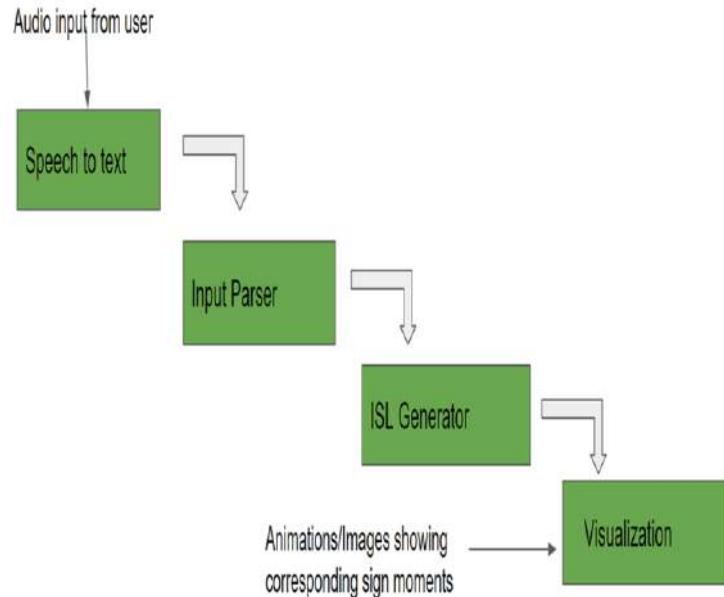
Figure 1. Low-Level Design

Speech to text phase: Input to the module can be given by an inbuilt microphone, an external source on any Personal Digital Assistant (PDA). Audio given by the user will be recognized using PyAudio, which is a Python binding for PyAudio in Figure 1. We can record and play audio on various platforms including, Microsoft Windows, Apple Mac, and Linux. The module uses Google Cloud Speech for audio to text. The output is given as an English text string. The module will be developed in python script

Table 1. POS Tags used for Stopwords removal

POS Tag	POS Tag full-form
CC	Coordinating Conjunction
NNS	Noun, Plural
RP	Particles
NNPS	Proper Noun, plural
SYM	Symbols
DT	Determiners

Input Parser: The input section is made into tokens of sentences. Using Machine Learning and Natural Language Processing Tools, each sentence is marked. The issuance of this section is a list of tokens in each line of text with appropriate punctuation marks in Table 1.

ISL Generator: This module aims to convert the input text with the grammar of English into text with the grammar of ISL. After the tokenization, we will first POS tag defined as the process where each word is matched with its part of speech. Also commonly known as POS tagging. Part of Speech includes adjectives, conjunctions, verbs, adverbs, nouns, pronouns, and sub-categories. Once the POS tagging is completed, we can go ahead and eliminate all the stop words. Stop words in a simple language are words which does not add much meaning to the sentence. Since they do not add much meaning, they can be removed, and the sentence's meaning will be unchanged.

Furthermore, after removing them, we will use lemmatization to eliminate the inflectional ending and return the root form of the word called the lemma. If the token saw is to be lemmatized, it will return either saw or see depending on whether the token used a noun or verb in the sentence.

Visualization: Input to this module is the ISL text string. In this module, generators map each token of text to a database to fetch the link. For every word present in the database, there is a corresponding file present, and for the words that are not there, we will find their synonym, and if the synonym is also not there, we will display each alphabet individually.

4.   Result and analysis

For a given English audio, the system aims at generating its equivalent sign language depiction. Our system generates these outputs in the following ways:

Video generation - Output from the ISL conversion phase of the input sentence is passed to the video generation phase, wherein each of the words in the sentence is looked up in the database for its corresponding video file. Then these files are all concatenated to produce a more structured, informative, and easy-to-understand visual depiction of Indian Sign Language.

For example, we provide an audio input "I am going to the garden" the input parser converted this audio into text and tokenized this sentence as "I", "am", "going", "to", "the", "garden" , then ISL generator converted this tokenized sentence into ISL Gloss "garden I be go" then video output for the ISL gloss is played by extracting the video of the words from the database and playing it using Youtube IFrame API.

Figure 2. Result output

In comparison to existing approaches, we are using a YouTube player in our project for visualization, which makes us feel like we are communicating with a person instead of an avatar or animation. Figure 2. Our output gives results with the help of ISL Gloss sentence, which increases its understanding and reduces sentence complexity for generating signs.

5.   Conclusion

Here, we have attempted to create a model that will allow people with disabilities to express themselves distinctly, which will help them blend with the rest of the world without any difficulty. Our proposed model will successfully convert the given input audio into an animation. Many improvements along this route can be made as and when the ISL Dictionary grows. The words in the ISL are small, so many improvements can be made by adding new words to their dictionary to increase their breadth. In addition, text-and-speech integration can be done on a project to enable

better communication techniques that will allow users to convert Text into Indian Sign Language by hand-input text.

References

[1] F. Shaikh, S. Darunde, N. Wahie, and S. Mali, "Sign Language Translation System for Railway Station Announcements," 2019 IEEE Bombay Section Signature Conference (IBSSC), 2019, pp. 1-6, DOI: 10.1109/IBSSC47189.2019.8973041.

[2] K. Saija, S. Sangeetha and V. Shah, "WordNet Based Sign Language Machine Translation: from English Voice to ISL Gloss," 2019 IEEE 16th India Council International Conference (INDICON), 2019, pp. 1-4, DOI: 10.1109/INDICON47234.2019.9029074.

[3] B. D. Patel, H. B. Patel, M. A. Khanvilkar, N. R. Patel and T. Akilan, "ES2ISL: An Advancement in Speech to Sign Language Translation using 3D Avatar Animator," 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 2020, pp. 1-5, DOI: 10.1109/CCECE47787.2020.9255783.

[4] Mishra Gouri Sankar, Ashok Kumar Sahoo, and Kiran Kumar Ravulakollu, "Word-based statistical machine translation from English text to Indian sign language." "*ARPN Journal of Engineering and Applied Sciences* 12.2 (2017)".

[5] 1. Abbas A, Sarfraz S. Developing a Prototype to Translate Text and Speech to Pakistan Sign Language With Bilingual Subtitles: A Framework. Journal of Educational Technology Systems. 2018;47(2):248-266. doi:10.1177/0047239518794168

[6] Advaith Sridhar, Rohith Gandhi Ganesan, Pratyush Kumar, and Mitesh Khapra. 2020. INCLUDE: A Large Scale Dataset for Indian Sign Language Recognition. In Proceedings of the 28th ACM International Conference on Multimedia. Association for Computing Machinery, New York, NY, USA, 1366–1375. DOI:https://doi.org/10.1145/3394171.3413528

A Proposed Methodology to Mitigate the Ransomware Attack

Salunke M. D [a, 1], Kumbharkar P. B [b] and. Dr. Pramod kumar [c]
[a] *Research Scholar, Shri JJT University, India*
[b] *Professor, RSCOE, Pune, India*
[c] *Professor Shri JJT University, India*

Abstract. Now a day's network security becomes more important to organizations, government offices. With the fast advancement of the innovation, assaults throughout the years have turned out to be both progressively various and modern. Ransomware attack becomes one of the most popular weapons for network attackers that ransomware attack is increased rapidly year by year. The study shows that the ransomware attack is one of the top attacks that are most attacked malware by the attackers. In this paper, we focus on mitigation techniques that can be used to recover and mitigate the ransomware attack. The mitigation or recovery approach is very difficult as ransomware is depending upon cryptographic algorithms which are very difficult to crack.

Keywords. Network Security, Attacks, Ransomware, malware, Mitigation, Sniffing.

1. Introduction

Attack is anything which imposes the harm on the system. It can be of two kinds in general. Active Attack: An Attack that may change the information or damage the System. Passive Attack: A type of Attack in which the Attackers goal is to obtain information. The Attack doesn't modify or harm the system. Attack is an assault on system framework security that gets from a keen risk [1].
The Three Goals of Security:
1: Confidentiality: hiding information from unauthorized access.
2: Integrity: preventing information from unauthorized modification.
3: Availability: should be easily available to authorized users.
        The four classes of attacks that violate different security properties of the computer Networks are Interruption, Interception, Modification and Fabrication. [2] Ransomware is type of malware used by attacker to attack or locks user's system or data and ask user to pay ransom to gain access to data or system. Ransomware are real threat to systems that attacker can inject to user's system and encrypt the user data or system by using encryption algorithms like AES, RSA or some modern-day ransomware also uses combination of symmetric and asymmetric algorithms for

---

[1]Mangesh Dilip Salunke, Research Scholar, Shri JJT University,India
Email: salunkemangesh019@gmail.com.

encryption of user data or system and then ask user to pay ransom for accessing system or data by displaying the message about how you can get back your data, files or system [3].

Network defense techniques: The network defense techniques are those tools that can be used to protect network from being attacked or that are used to monitor malicious activities of attacker or process in network. Sometime these techniques allow us to avoid the malicious activities by providing strong security policies for organizations.

Cryptography: It includes protection of data and Data Hiding through the implementation of various algorithms and techniques. While Data Hiding the simple text is converted to a ciphertext, which is a combination of simple text and a certain key. This key is developed through different methods. Basically, there are two types of keys those are Symmetric key and public key. Symmetric key is uses same key for encryption and decryption of data which is kept private. Public key is using a public key known to all to encrypt data and a private key to decrypt data.

Cryptographic attack: Cryptographic attacks are those attacks that are created to evade the current network security systems. These attacks are designed by expertise attacker who knows how to compromise the security systems. These attacks can easily stop actions, productivity and revenue of well expanded business and cost millions of losses. Ransomware attack is one of the types of attack of cryptographic attack that allows digital extortion to the attacker.

2. Ransomware types

There are two types of Ransomware attack, one is Locky Ransomware, and another is Crypto Ransomware. These two classes of ransomware that defines the way of attack, one is Lock Screen Ransomware that simply locks or encrypt the user system or desktop or input system and keeps the user inaccessible by simply showing extortion message is also known as Locker Ransomware. And another is Crypto or Encryption Ransomware that encrypts the user data or files and asks to pay ransom is also known as cryptographic ransomware. Ransomware attacker changes their targets from individuals to organizations such as banking, government offices, and hospitals and may more. Attacker uses different ways to penetrate the user system with ransomware attack that is known as deployment location of attack like spam mails, compromised web sites, Downloading/Opening any malicious file, Log-into any already infected PC, Installing Pirated software are some examples.

The very first Ransomware attack happened in 1989 known as PC Cyborg also popularly known as AIDS Trojan that demanded $189 or $378 to the 20000 victims that work for AIDS research related journal. For the distribution attack vector used by attacker Dr. Josef Popp is in the form of Floppy disk. After that the attack becomes more popular of attacker as Internet grows the attack grow with it. The use of internet for financial transaction, digital communication, data transfer, and social interaction makes attack more spread and big concern of organizations [3].

3. Literature Review

Authors developed HelDroid a lightweight small emulation system on android platform that is a real time detection system. The system uses the features of natural language processing technique to detect locking mechanisms or encryption detectors with the help of static and dynamic analysis for Android Ransomware samples. The system is tested on a large dataset near about hundreds of thousands APK's that contains scareware, Malware, Good ware, and known ransomware and unknown Ransomware samples. Result shows that the system has near about zero false positive rate and 99% capability of recognizing new Ransomware samples on Android platform [4].

Ransomware attacks and mitigation or awareness strategies of attack with conducting interview and survey of victimized and non-victimized people and the result of survey were analyzed by using statistical analysis by using different factors such as age, education, awareness that are dependable upon causes of ransomware attack. And the result shows that it is irresponsibility and dependency for attack on the IT department of other employee to make the attack happen [5].

Author creates a novel solution in the form of an easy to use script that runs on Windows 7 platform to recover from Crypto ransomware. They executed their system on a secure platform such as Virtual machine and tested by attacking that virtual machine by renaming vssadmin.exe file which is used for backup purpose in virtual machines to prevent encryption of backup with theses crypto ransomware attack. The system shows that by using proper preventive measures such as updated antivirus, updated operating system and software, proper backup and well configured firewall one can easily restore the system to the normal state as it was before the encrypted state. The analysis is done on most recent and common 4 crypto Ransomware samples at that time [6].

In this authors implement a technique that monitors suspicious activity over a locally virtualized environment for a file system that is dynamic system name as POSTER. Their solution is based on the behavior of Ransomware on Windows 7 Platform to mitigate the effect of ransomware with four modules. The result shows the possibility to detect old as well as new variants of the Ransomware family. To analyze the gap of previous existing systems they have developed three new variants of ransomware such as Zero replacement, Friend indeed behavior, Ledger Manager Behavior based on their behavior [7].

Author studied and describes the crypto ransomware attack in terms of history and timeline of some popular ransomware attack that describes the journey of attack for making awareness among internet users about the attack and potential harm of attack. Static analysis is done on captured packets of ransomware from various sources to understand the working of ransomware attack, C&C server and contribution of crypto currency to ransomware attack. They have performed their static analysis on virtual machine environments. For that they have used two VMware machines one is for a fake DNS server as REMnux operating system and another is victim's machine that is a Windows 7 machine. Sample of Cerber ransomware was used to analyze a Wireshark traffic analysis tool [8].

Authors discussed crypto virology, the technique in which an attacker merges the cryptography with malware such as ransomware attack. They have discussed the symmetric and asymmetric key cryptography technique. The attacker uses the asymmetric key to launch the ransomware attack and how the public key of the attacker

is of no use to the victim and also how randomly generated symmetric key received to victim is unused to others [9].

They analyzed ransomware attack and provided a report of their analysis and study of ransomware attack with the help of history and evolution of ransomware attack and provide some prevention techniques that are necessary to counter the ransomware attack for different organizations. To understand and to analyse the working of attack they have created a sample model of attack as a demo ransomware in a controlled environment. Their demo model starts encrypting files with AES encryption algorithm after searching to only some specific types of files such as *.cpp extension files that they used for the target file in their model. In their result they state that the CPU usage increased from 3% to 28 % after execution of attack [10].

Author discussed the types of ransomwares such as Goldeneye, WannaCry, RAA ransomware, Cerber, Crysis, petya, Locky, CryptoWall, PowerWare with the help of case study of attack with most popular cyber-attack such as Google China, WannaCry, Heart bleed and PlayStation server. Author stated that the ransomware attack for encryption of files it generally use RSA 2048 encryption algorithm and because of security loophole of SMB that is samba server vulnerability which uses EternalBlue protocol which used by attacker as a exploit kit which mostly affect the windows operating systems such as windows 7, windows xp, windows 8, windows 10 and windows server which use AES-128 encryption algorithm for encryption process. Lastly the author described some prevention tips and security tips to users to avoid the ransomware attack [11].

Author discussed previous ransomware detection systems with the help of literature survey. Author also evaluates both stated and implicit assumptions on each and every detection system they have surveyed as their literature study. They gave name to working of attack as ransomware attack kill chain as a Identify and recon, Initial Attack, Command & Control, Discover & Spread and Extract & Exfiltrate are the phases of ransomware attacks on Android system [12].

4. Mitigation Approach

As ransomware attack is based on cryptographic algorithm such as AES, RSA and other algorithmic techniques are used to create this most effective attack. In cryptography the plain text is converted into unreadable cipher text format with the help of Cryptographic key. The key is required to convert this cipher text to again into plain text that is readable by humans. Attackers' usage these approach to attack systems or files to encrypt the users system or files. As the key is required to convert the files or system back to working stage attacker ask the user to pay for that key. Also these cryptographic algorithms are very strong in nature one cannot crack the key for decrypting the files which are locked due to ransomware attack.

Figure 1 describes the architecture of proposed systems mitigation module, in which it describes the process of mitigation or recovery from ransomware attack. In this left part shows the working of mitigation module and right part shows the working of decryption process of mitigation module.

(i) Packet Capturing: Packet capturing is the process of capturing the network packets for monitoring and detecting the possibilities of malicious code or attachments. There are various packet capturing techniques are available for analyzing and detecting the

packets for possible attack. Wireshark, Pcap and TCPDUMP are some examples of packet capturing and analyzing tools.
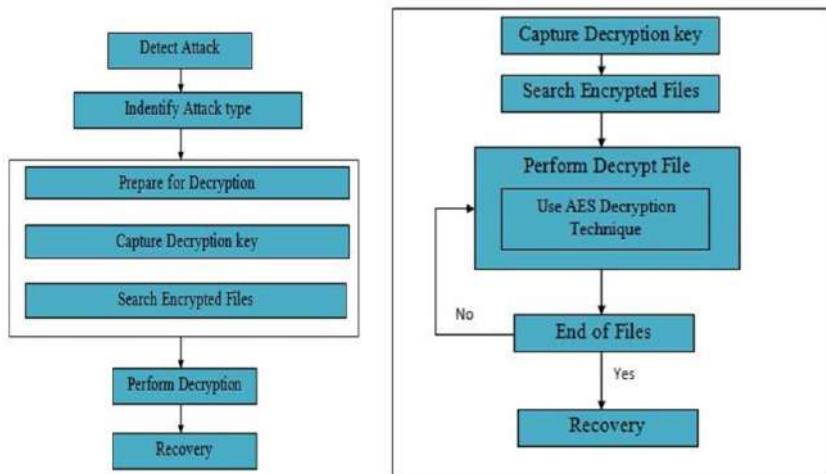
Figure 1. Architecture of proposed system mitigation module.

Wireshark: It is one of the most popular and widely used network packet analyzer tools for monitoring and analyzing the computer network for checking the possibilities of network attack in Figure 1.

PCAP: It is nothing but a Packet Capture. A tool for capturing the network traffic with the help of Application programming interface (API).

(ii) Sniffing: Sniffing or sniffer used to sniff the network packets with the help of packet capturing and analyzing tools. The tools are also known as Packet Sniffer. It is also known as Network Analyzer or Packet Analyzer which is useful for monitoring the computer network for network attack.

(iii) HTTP Traffic analysis: It is the process in which HTTP traffic from any network can be analyse, that is it capture the HTTP packets that are transfer from one end to another end of network and these packets are then analyzed for checking the possibility of malicious activity. Several tools can be used to capture and analyze the HTTP traffic.

5. Result and discussion

 In proposed approach of ransomware mitigation technique, it is not even possible to crack or perform dictionary attack on AES or RSA to get the encryption key. So only way to get back data or files that are encrypted with ransomware attack is either pay the ransom or sniff the packets that are going outside into the direction of command-and-control server. With the help of this sniffing technique, we can able to capture and analyze the packet data that is going towards the command-and-control server after encryption process and that key of decryption is send to that server.

The proposed methodology for mitigation or recovery of ransomware attack data shall be implemented on Virtual machine and Windows 10 Machine. AES

cryptographic algorithm is used for encryption and decryption process of data in attack. For analyzing and monitoring the network packets Wireshark tool is best used.

6. Conclusion

The ransomware attack is very vital and most dangerous attack for computer users and organizations, so it is very important to counter such kind of attack. The biggest concern while designing the countermeasure system for countering the ransomware attack is the mitigation part as it allows users to recover from attack and get back data or system of users. One has to take care about recovery or mitigation module while building the ransomware countermeasure systems. Because of which in proposed system this countermeasure technique is includes mitigation as a one of the modules which is an important part of proposed system and works as a defending the most popular and powerful weapon of attacker that is ransomware attack.

References

[1] Saiyed Kashif Shaukat, Vinay J. Ribeiro," RansomWall: A Layered Defense System against Cryptographic Ransomware Attacks using Machine Learning", 2018 10th International Conference on Communication Systems & Networks

[2] Ibrahim Nadir, Taimur Bakhshi, "Contemporary Cybercrime: A Taxonomy of Ransomware Threats & Mitigation Techniques", 2018 International Conference on Computing, Mathematics and Engineering Technologies – iCoMET 2018, 978-1-5386-1370-2/18/$31.00 ©2018 IEEE.

[3] Kul Prasad Subedi, Daya Ram Budhathoki, Dipankar Dasgupta, "Forensic Analysis of Ransomware Families using Static and Dynamic Analysis", 2018 IEEE Symposium on Security and Privacy Workshops, DOI 10.1109/SPW.2018.00033

[4] Nicol´o Andronio, Stefano Zanero, and Federico Maggi, "HELDROID: Dissecting and Detecting Mobile Ransomware", Springer International Publishing Switzerland 2015, RAID 2015, LNCS 9404, pp. 382–404, 2015. DOI: 10.1007/978-3-319-26362-5 18

[5] Rhythima Shinde, Pieter Van der Veeken, Stijn Van Schooten, Jan van den Berg," Ransomware: Studying Transfer and Mitigation", 2016 International Conference on Computing, Analytics and Security Trends (CAST) College of Engineering Pune, India- 2016 IEEE

[6] Mattias Wecksten, Jan Frick, Andreas Sjostrom, Eric Jarpe,"Novel Method for Recovery from Crypto Ransomware Infections",2016 2nd IEEE International Conference on Computer and Communications

[7] Manish Shukla, Sutapa Mondal ,Sachin Lodha, "POSTER: Locally Virtualized Environment for Mitigating Ransomware Threat", ACM ISBN 978-1-4503-4139-4/16/10. DOI: http://dx.doi.org/10.1145/2976749.2989051

[8] Shaunak Sanjay Ganorkar and Kamalanathan Kandasamy,   "UNDERSTANDING AND DEFENDING CRYPTO-RANSOMWARE", ARPN Journal of Engineering and Applied Sciences, VOL. 12, NO. 12, JUNE 2017

[9] Adam L. Young and Moti Yung, "Cryptovirology: The Birth, Neglect, and Explosion of Ransomware", COMMUNICATIONS OF THE ACM JULY 2017, VOL. 60

A Relative Investigation of Various Algorithms for Online Financial Fraud Detection Techniques

C Pallavi [a, 1], Girija R [b], Vedhapriyavadhana R [b], Barnali Dey [c], Rajiv Vincent [b]

[a]*School of Electronics and Communication, VIT University, Chennai Campus, India*
[b]*School of Computer Science and Engineering, VIT University, Chennai Campus, India*
[c]*Department of Information Technology, SMIT, SMU, Sikkim, India*

Abstract. Online financial transactions play a crucial role in today's economy. It becomes an unavoidable part of the business and global activities. Transaction fraud executes thoughtful intimidations to e-commerce spending. Now-a-days, the online contract or business is fetching additional sound by knowing the types of online transaction frauds associated with, these are raising which disturbs the currency accompanying business. It has the capability to confine and encumber the contract accomplished by the intruder from an honest consumer's credit card information. In order to avoid such a problem, the proposed system is established transaction limit for the customers. Efficient data is only considered for detecting fraudulent user action and it happens only at the time of registration. Transaction which is happening for any individual is not at all known to any FDS (Fraud Detection System) consecutively at the bank which mainly issues credit cards to customers. To speak out this problem, BLA (Behaviour and Location Analysis) is executed. The FDS tracks at a credit card provided by bank. All the inbound business is directed to the FDS aimed at confirmation, authentication and verification. FDS catches the card particulars and matter to confirm that the operation is fake or genuine. The pick-up merchandises are unknown to Fraud Detection System. If the transaction is assumed to be fraud, then the corresponding bank declines it. In order to verify the individuality, uniqueness or originality, it uses spending patterns and geographical area. In case, if any suspicious pattern is identified or detected, the FDS system needs verification. The information which is already registered by the user, the system identifies infrequent outlines in the disbursement method. After three invalid attempts, the system will hinder the user. In this proposed system, most of the algorithms are checked and investigated for online financial fraud detection techniques.

Keywords. Online fraud, Secure payment, Location analysis, Fraud detection system.

1. Introduction

Online scams charge billions of dollars all over the world. Applying machine learning algorithms to predict the possibility of a transaction being a fraud transaction is one of the efficient methods. In our proposed system we will take credit card transactions, analyze the data, create the features and labels and finally apply one of the ML

---

[1] Rajiv Vincent, School of Computer Science and Engineering, VIT University, Chennai, India
E-mail: rajiv.vincent@vit.ac.in.

algorithms to judge the nature of the transaction as being a fraud or not. Then check the accuracy, precision, and score of the model chosen. Fraud detection is one of the major priorities for banks and commercial Organizations. This can be lectured with the use of algorithms in machine learning for calculating the accuracy. Fraud detection methods are updating promptly permitted to familiarize in latest deceitful approaches globally. But, the expansion of novel deception detection procedures makes more complicated due to the undecorated restriction of the designs and methods of conversation in fraud detection [1]. The number of fraudulent dealings is frequently found in an actual little percentage when compared to the total transactions. Hence the task of detecting fraudulent connections in a precise and effectual manner is fairly difficult and challenging. Therefore, the development of efficient methods which can discriminate occasional fraud happenings from billions of legitimate transactions seems essential. Credit card fraud happens when an unauthorized person gains access to your credit card information and uses it to make purchases. Some of the most common ways are high-level hacking of the bank account details, through Phishing attacks, when card details are overseen by some other person, lost or stolen cards and fake phone call convincing the target to share the details.

There are certain main challenges which complicate the credit card detection frauds are:

1.1. Development of Fraud designs over period

It is very toughest to report meanwhile the fraudsters are continuously find all the inventive methods to increase admittance to the design to get credit card details. Hence, it becomes very significant for all the machine learning models which is to be restructured with respect to developing arrangements in order to detect suspicious configurations. The report is showing decrease in accuracy and efficiency. Hence, all the machine learning models must be reorganized otherwise; they will fail in their objectives.

1.2. Imbalance

In the fraud detection models, there is a disparity in the classification. It makes the system harder. For genuine clients, the drawback of this challenge is worst experience. Sometimes, imbalance occurs, by weakening the truthful transactions then only fraudsters will get caught.

1.3. Model Interpretations

Model interpretations contest is accompanying with the conception of explainability because all the learning approaches calculates a score for identifying the transaction is fraud or genuine.

1.4. Feature generation are time-consuming

It has been identified and reviewed that specialists might need longer periods of time to produce a perfect feature established which reduces down the fraud recognition process.

1.5. Dissimilar misclassification significance

Various misclassification mistakes have been coming under the category of fraud detection. Misclassification of a reasonable purchase as deception is not as damaging when compared to noticing a fraud transaction as a legitimate one.

1.6. Overlying data

Various transactions might be reflected duplicitous, which may be a false positive, and in reverse, it may be a false negative. Hence, in order to obtain a truncated rate of false positive and false negative and this is the major challenge in fraud detection systems.

1.7. Absence of adaptability

It is very tough for the classification of machine learning algorithms to distinguish innovative styles of legitimate or dishonest patterns.

There are various measures to solve these challenges are:

1.8. Human-in-the-loop

This method solves the imbalance problems as well as decreases the time for detecting the features [2]. It also comprises persons supporting the replicas by providing information to detect novel patterns, features, and many dimensions of fraud.

1.9. Ensemble approach

In order to encounter the continuously emerging fake outlines, Ensemble approach chains numerous representations for a solo mission such as fraud revealing. Collaborative with respect to machine learning can detention fraud patterns to exploit outcomes and increase accurateness.

1.10. Explain ability

The concept of understandable machine learning can deliver explanations for complimentary or deteriorating transactions as fraudulent, therefore resolving the exemplary elucidation experiment. There are techniques such as surrogate modelling, maximum activation analysis, and others that provide all these benefits.

There are countless categories of credit card deceptions. Some of those deceptions are Application frauds, Electronic or Manual Credit Card Trajectories, CNP (Card Not Present), Counterfeit Card Fraud, Lost and Stolen Card Fraud, Card ID Theft, Mail Non-Receipt Card Fraud, Account Takeover and Merchant Collusion. Application Frauds is the impostor improvements contact to the presentation classification by retrieving worker particulars similar password and username and making a bogus explanation by person's particulars. This is called individuality stealing when the swindler relates for a new credit card in the board's designation. The fraudster bargains searching and essential papers in order to maintenance their deceitful presentation [3]. Electronic or Manual Credit Card Trajectories is nothing but once the swindler gains

admittance to the info which is positioned on the magnetic stripe of the card which is actual personal and it can be used to acquisition in the future. Once the impostor gets contact to the account numeral and finishing date of the card, that particular card can be used for purchasing without its original presence, and then it is CNP (Card Not Present). Process of skimming and counterfeiting credit card are interrelated. A false magnetic swipe card clamps all the particulars of the unique card and it is a operational card and can be used to purchase in the future. Once the genuine cardholder mislays their card and if it becomes into the hoaxer's hand and it can be used to brand expenditures. It is tough to do this a machine is in need of PIN, but online dealings are relaxed sufficient for the swindler and comes below group of lost and stolen card fraud. Card ID Stealing is the fraudster gets the particulars of the genuine card to use the card or to produce a new version. When a user smears for a card, it proceeds selected time to appearance all the rules and regulations. If a swindler may record the card in their designation and use it to make acquisitions by interrupting in the intermediate of the conveyance. It comes under the never received issue fraud and it arises below the grouping of Mail Non-Receipt Card Fraud. Account Takeover is the furthermost communal method of deception in which an impostor strength improvement admittance to the card particulars of the inventive cardholder and several profound forms. They can even contact the credit card company and act as if they are the original cardholder and might level ask them to adjustment the address [4]. By the way of the fraudsters have all the specifics which they got done hacking or else manipulating the original cardholder, they can deliver them by way of resilient. The replacement card resolve at that moment be directed to the place assumed by the fraudster. False Merchant Sites is like a phishing violence anywhere then operator becomes stuck in a counterfeit webpage, fashioned through the swindler, which appearances actual analogous to a recognized and honest website to influence the user similar a reductions page promising the consumer to acquisition the merchandises. As soon as the compensation is finished, all the evidence is composed and the swindler usages it for upcoming procurements. When the merchant or shopkeeper licenses on the information connected to the user's card deprived of the cardholder knowing is nobody but Merchant Collusion.

2. Methodology

2.1. Platform

For Python coding (Using Anaconda Navigator)-JupyterLab is a user interactive environment used for development purposes. JupyterLab is flexible to implement for our Project and Python Coding as well. The packages included for this comparison are Numpy, Pandas and Seaborn. NumPy adds support for matrices and 3-dimensional arrays with a huge collection of mathematical expressions to work on these arrays. In data-intensive calculation, Numpy provides a range of methods that make data manipulation in Python less complicated. Since Python is slower in execution compared to other languages during looping, Numpy helps speed up the operations by converting repetitive code to the compiled form.Pandas offers operations for operating numerical tables and time series. Seaborn is used for data visualization and gives an interface for creative and knowledgeable statistical graphics. Visualization is the major part of Seaborn which helps in understanding data. Seaborn offers the following functionalities such as Dataset-oriented API to regulate the relationship between

variables, Automatic calculation and plotting of linear regression plots. it supports a high-level concept for multi-plot grids, Visualizing univariate and bivariate distribution. Matplotlib- Matplotlib library is used for plotting and getting the right plot is often achieved through trial and error. It gives an API for implanting plots into web apps using tools like Tkinter, Qt,etc and there is a methodological "pylab" interface which depends on a state machine, created closely resembling MATLAB. Pydot- Pydot is a Graphviz interface and it can analyze the DOT language. It has been developed using Python. Sklearn- The library consists of a lot of important tools for machine learning like regression, clustering,etc [5]. Ipython - It is a command shell library for various programming languages. It was originally created for Python that offers shell syntax, introspection, tab completion, and history.

2.2. Comparison Algorithms

2.2.1. Isolation Forest Classifier

Isolation Forest algorithm is a tree-based model used to detect outliers and it is separated with arbitrary breaches than a model consisting in an even class, as outliers are repeated fewer than usual results and must have prices beyond the dataset [6].

Succeeding this concept, this classifier selects a feature and at that point selects a value within the sort of this article as the split significance. By means of the previous stage, the algorithm continuously creates a tree. The quantity of essential arbitrary separations to separate a model is the tree penetration. The separation numeral, be an average of concluded a forest of such unsystematic trees, is an evaluation of ordinariness and our pronouncement purpose to distinguish outliers. Random splitting creates acknowledgeable tinier tree depths for outliers and lengthier tree depths for the break of the data models. Hence, when a forest of random trees creates smaller pathway lengths for a specific document point, this is expected to be an outlier.

2.2.2. Random Forest Classifier

Random forest is an algorithm based on ensemble learning. It is an algorithm in which the predictions are obtained by aligning different models or similar models' multiple times. The random forest algorithm functions in the same way and uses various algorithms i.e., multiple decision trees, concluding in a forest of trees, hence that's how the algorithm is named as "Random Forest" which may be used for both regression and classification purposes. Some of the benefits of this classifier are that this algorithm is not biased and it is based on the group of trees where every single tree is trained separately based on the data, therefore biasedness is lowered, it's a very strong and firm algorithm. Even if a new data point is added to the dataset it doesn't affect the overall algorithm but affects only a single tree. Thus, using this Random forest algorithm and decision trees gives the accurate percentage from the given dataset by studying its behaviour [7].

2.2.3. Decision Tree Classifier

It is the most accurate model used in data, and machine learning. This is mainly based upon decision tree which is used to look at an any item and conclude the item's target value. The Ordering tree is a model anywhere the goal variable can yield a vivid set of values from the tree structures. Regression trees are an example of decision trees where

a specific variable can yield uninterrupted values. Decision trees are unique of the best machine learning procedures and are a simple representation of classifying data. It is Supervised Machine Learning where the data is split according to certain parameters [8-10].

2.2.4. Naive Bayes Classifier

Naive Bayes classifiers are founded on relating Bayes theorem through individualistic expectations amongst the structures. They are one of the easiest Bayesian network models. But they could be added with Kernel density and achieve high accuracy levels.Naive Bayes classifiers are very accessible, requiring a particular number of conditions linear in the number of features in a learning problem. Highest training can be done by calculating a closed-form form by considering linear time, rather than by repeated approximation which is used for other types of classifiers as well [11]. It has real-valued attributes estimated by assuming a Gaussian distribution and is easiest to work with, and only need mean and std from training data by calculating mean and std of input values(X) for each class to summarize the distribution.

2.2.5. Logistic Regression

Logistic regression uses a logistic expression to model a multi-variable that is dependent. In regression analysis, logistic regression is used to evaluate the criteria of a logistic model. In Spite of the name being regression, logistic regression is used for clustering for detecting binomial and multinomial results, with the aim of estimating the values of coefficients using the sigmoid function.

2.2.6. K Nearest Neighbours

The k-nearest neighbour algorithm is a non-parametric method created by Thomas Cover for regression and classification of data. The input includes all the k nearest training examples in the set. An object is grouped by the decision of its neighbours, with the object being assigned to the class which is usual among its neighbours. When k = 1, the object is allocated to the group of a nearest neighbour. In k-NN regression, the result is the data for the object.

2.3. Statistical Evaluation tools

2.3.1. Confusion matrix

The confusion matrix is comparatively simple to understand, but the terminology can be confusing. The confusion matrix represents the True Positive values, which means a class of the data matches the predicted class of the data. False Positive represents that the actual class of the data was 1 but the model predicted it to be 0. False Negative represents that the actual value of the class was 0 but it was predicted to be 1. True Negative represents that the actual value was 1 and the predicted value is also 1. Figure 1. Represents the confusion matrix.

Actual

Predicted	Positives (0)	Negatives (1)
Positives (0)	TP	FP
Negatives (1)	FN	TN

TP = True Positives, FP = False Positive

FN = False Negative, TN = True Negative

Figure 1. Confusion matrix

Accuracy is a measure for scoring the models. It is the part of results the model got right. It is been calculated by the below equation.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{1}$$

Precision is the proportion of accurately detected positive observations to the total detected positive observations. The formula for evaluating Precision is Equation 2.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

The F1 score is calculated using Precision and Recall. They are considered as weighted average for F1 score. Hence, this F1 score includes the false positive and False Negative. Always, accuracy is the foremost parameter, but in our proposed model, F1 score is more important than accuracy. The above case is principally for an uneven dissemination. It is calculated using the equation3.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{3}$$

Recall score is the one whose ratio is decorously anticipated positive opinions to all explanations in the genuine class. Recall has been calculated by the following equation.

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

Matthews Correlation Coefficient is the coefficient considered a valid account TP, TN, FP and TN. Generally, it is regarded as a well-adjusted portion that is in use, even there are classes with different sizes. It yields a value between −1 and +1 since it is a correlation coefficient between the predicted and pragmatic binary classifications. If it is a well and perfect prediction, the coefficient value is +1. In case, if it is a random

prediction, then the coefficient value is zero. The last case, if it completes mismatch between and disagreement between prediction and observation, then coefficient value is -1.

Cohen's kappa coefficient (κ) is a statistic, which is used to extend both Inter-rater dependability and also Intra-rater reliability for qualitative and categorical items. As κ considers, the opportunity of the arrangement happening by chance is generally to be a more robust degree than simple percent agreement calculation. But there is a difficulty in interpreting indices of agreement and this is the controversy for Cohen's kappa. Many researchers have suggested very good thoughts for evaluating disagreement between items.

3.   Implementation

3.1.   Gathering Data

While using machine learning, the initial step is to know the problem. According to the problem definition, data should be collected. For machine learning, a dataset can be created or there are data that already exists. There are many platforms that provide the collection of datasets to solve machine learning problems.

3.2.   Pre-processing the data

After the data is collected, the data needs to be processed. Without pre-processing the data or providing raw data to the model, it does not provide the expected results. Use the techniques which can provide the best form of the data which increases the accuracy of the model. If the dataset is skewed, try to balance it, perform feature selection, feature extraction, transferred learning.

3.3.   Split the dataset

After cleaning the data, divide the dataset. Data can be split into train test ration, train-test-validation ratio, or use cross-validation. By splitting the dataset, the training dataset for the training of the model and the remaining for evaluating the model can be provided. This helps to avoid the over fitting of the model.

3.4.   Choosing a model

After dealing with data, select the model according to the dataset, and the type of task needed to be performed like classification, clustering. Choosing an appropriate model is very important or else results will not be achieved.

3.5.   Evaluate the model

After the training of the model, predict the results on the unseen dataset. If the prediction metrics provide the results which are expected, then the model is said to be ready for classifying the data. If the results are not satisfactory, retrain the model and change the parameters, fine-tune them, until the achieved results are satisfactory.

.

4.   Results and Discussions

From the figure 2, it has been clearly observed, fraud representation of data. Credit card fraud occurs mostly in the time of transaction. Figure.3. represents the comparison between fraud cases and genuine cases in an hour.
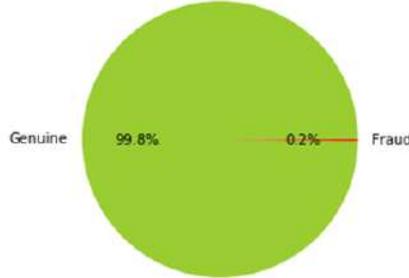
Figure 2. Pie chart representation of data
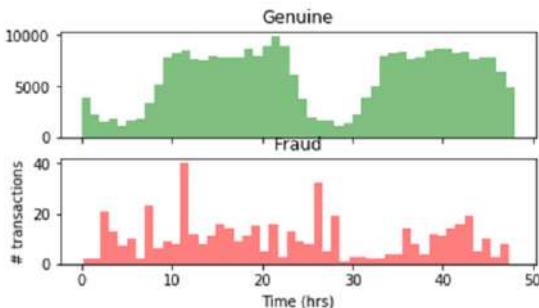
Figure 3. Transactions over time (in hrs)

Detection cases are identified, and it has been separated in such a way as false detection cases and True detection cases. It has been displayed in the table. 1.

Table 1. No. of actual and false transactions and other statistical data

```
False Detection Cases
---------------------
count      492.000000
mean       122.211321
std        256.683288
min          0.000000
25%          1.000000
50%          9.250000
75%        105.890000
max       2125.870000
Name: Amount, dtype: float64

True Detection Cases
---------------------
count   284315.000000
mean        88.291022
std        250.105092
min          0.000000
25%          5.650000
50%         22.000000
75%         77.050000
max      25691.160000
Name: Amount, dtype: float64
```

The correlation matric has been calculated and it is shown in figure. 4.
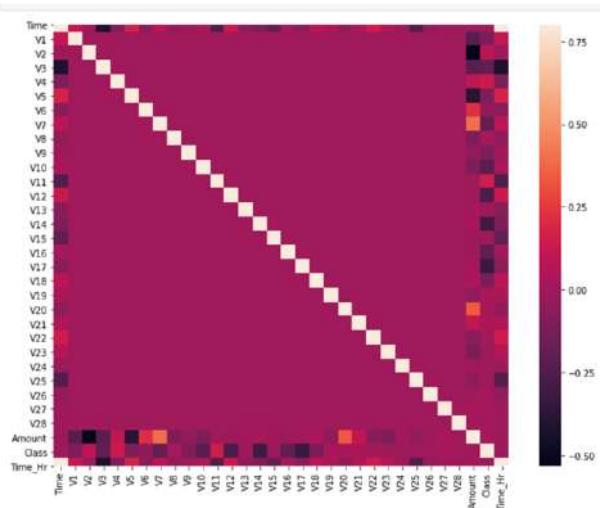
Figure 4. Correlation matrix.

5. Comparative Analysis

We compared the performance of classification models by calculating all the metrics for statistical evaluation that are generated by the algorithm in table 2. True Negative is a number of negative results that are valid and also classified as valid. True Positive is a number of positive outputs that are considered fraudulent and by the system too. False Positive is the number of positive results that are mentioned as valid but are incorrectly mentioned as fraudulent. False Negative is the number of negative results that was mentioned as fraudulent but is incorrectly mentioned as valid by the system.

Table 2. Comparison Chart

Techniques	Accuracy	Precision	Recall	F1-Score	Matthews Correlation Coefficient	Cohen Kappa
Isolation forest	99.7%	34.06%	31.6%	32.8%	32.7%	32.7%
Random forest	99.9%	96.05%	74.4%	83.9%	84.5%	83.9%
Decision tree	99.9%	96.05%	74.4%	83.9%	84.5%	83.9%
Naive Bayes	98.3%	8.4%	87.7%	15.4%	26.9%	15.2%
Logistic Regression	99.9%	80.8%	56.1%	66.2%	67.3%	66.2%
K- Nearest Neighbor	99.9%	92.1%	71.4%	80.4%	81.08%	80.4%

6.   Conclusion and Future Work

After comparing all the classification algorithms, it is concluded that the random forest/decision tree classifiers are the most suited to this kind of application as the basic classifiers such as logistic regression, Naive-Bayes and K-nearest neighbours all have similar accuracy but significantly lower values in the other metrics. Thus, we can conclude that a fraud detection system that is built based on the random forest algorithm will work well as compared to other algorithms. Future work on this proposed method can enhanced to test/compare more algorithms by building a classifier from multiple models to find a superior one to increase the metric scores.

References

[1]   Jain, Y. & Tiwari, N. & Dubey, S. & Jain, Sarika, "A comparative analysis o f various credit card fraud detection techniques,"*International Journal of Recent Technology and Engineering.*vol. 7, 2019.

[2]   T. R. C.Sudha, "Credit card fraud detection in the internet using k nearest neighbour algorithm," *IPASJ international journal of computer science*, vol. 5, 2017.

[3]   E. D. Yusuf Sahin, "Detecting credit card fraud by ann and logistic regression," 2011.

[4]   F. Carcillo, Y.-A. Le Borgne, O. Caelen , Y. Kessaci, F. Oblé, and G.Bontempi, "Combining unsupervised and supervised learning in credit card fraud detection,"*Information Sciences*, doi: 10.1016/j.ins.2019.05.042, 2019.

[5]   A. S. Wheeler R, " Multiple algorithms for fraud detection.Knowledge-Based Systems," no. S0950-7051(00)00050-2, 2000.

[6]   A. O. A. S. A. O. John o. Awoyemi, "Credit card fraud detection using machine learning techniques: A comparative analysis," in *International conference on computing networking and informatics,* 2010.

[7]   S. K. M. A. Masoumeh Zareapoor, "Analysis of credit card fraud detection techniques: based on design criteria," in *International journal of computer applications*, 2012.

[8]   S. Mittal and S. Tyagi, "Performance Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection,"*2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India*, 2019, pp. 320-324, doi: 10.1109/CONFLUENCE.2019. 8776925.

[9]   Jain, Y., NamrataTiwari, S., & Jain, S. (2019). "A comparative analysis of various credit card fraud detection techniques,"*International Journal of Recent Technology and Engineering*, 7(5), 402-407.

[10] Duman, E., & Ozcelik, M. H. (2011). "Detecting credit card fraud by genetic algorithm and scatter search,"*Expert Systems with Applications*, 38(10), 13057-13063.

[11] Kumar, V. A., Kumar, V. A., Malathi, S., Vengatesan, K., & Ramakrishnan, M. (2018). Facial recognition system for suspect identification using a surveillance camera. Pattern Recognition and Image Analysis, 28(3), 410-420.

A Comprehensive Survey on Aspect Based Word Embedding Models and Sentiment Analysis Classification Approaches

Monika Agrawal [a,1] and Nageswara Rao Moparthi [a]

[a] Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur- 522502, AP, India

Abstract. Sentiment Analysis includes methods and techniques for businesses to understand and analyze customer reviews, feedback and opinion on a particular product or service. Sentiment Analysis uses Natural Language Processing (NLP) tools to analyze feelings or emotions, attitudes, opinions, thoughts, etc. behind the words. Sentiments such as positive, negative and neutral are associated with a particular product. Sentiment analysis is applicable in multi- domains such as customer feedback for a particular product, movie reviews, social and political comments. This survey basically focuses on different aspect-based word embedding models and aspect-based sentiment classification techniques, where the goal is to extract key features from the sentences and classify sentiment on entities at document level. Aspect Based Sentiment Analysis (ABSA) is a technique that concentrates not only the entire sentence but analyses key terms explicitly to predict the polarity as a whole. ABSA model accepts aspect categories and its corresponding aspect terms to generate sentiment corresponding to each aspect from the text corpus. This article provides a comprehensive survey on different word embedding models under CNN framework for aspect extraction and different machine learning techniques applicable for sentiment classification purpose.

Keywords- Aspect sentiment, filtering, classification, polarity.

1. Introduction

As the use of internet is growing rapidly, number of users is exchanging their thoughts and opinions on numerous issues on ecommerce and other websites such as twitter, LinkedIn, Amazon, eBay, etc. This social Websites allows users to give their feedback or opinion on use of different products and services, stock exchange data etc., which was very hard to collect desired data and analyze them in an easy way in past time, is now readily available. Through these sites people are sharing their views, feelings and opinion anytime on a particular product or service expressed in positive, negative, or neutral way. Sentiment Classification, is a technique to detect the polarity of given text, is one of the basic tasks in NLP. Figure 1, illustrates General Sentiment Classification process.

---

[1] Monika Agrawal, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, India
Email: monika.agrawal1986@gmail.com.
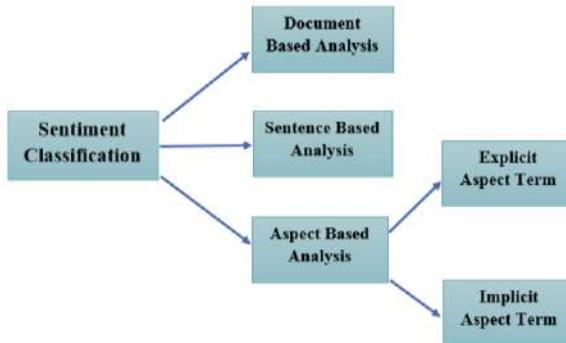
Figure 1. General Sentiment Classification Process

In Document based analysis, Sentiment classification technique extracts all opinion words from the entire document to give the opinion/ sentiment of entire document as a whole. In Sentence based perspective, Sentiment Analysis focusses mainly on polarity detection of complete sentence as positive opinion or negative opinion without considering in to account the important key features of the sentences. Next, Aspect Based Sentiment analysis aims to extract essential aspect related key features from the text and then assigns polarity or classifies sentiment of aspect terms to be positive negative or neutral. In ABSA, first aspect extraction process takes place in which key features are extracted for the given entity and then key features are subjected to machine learning classification approaches to determine the polarity of text [1].

2.   Methodology

Aspect based sentiment analysis (ABSA) works on three different phases. Figure 2 shows the flowchart of ABSA framework. Consider input data set such as Twitter Dataset, Amazon Product Data, Movies Review Dataset, Paper Reviews dataset etc. First phase begins with Pre-processing or filtering input data using standard NLP library. Filtered data of the phase one is given to the second phase for essential features extraction. In the second phase, word embedding models are designed and implemented on the filtered data for aspect features identification [2].

In this framework, different hybrid word embedding methods such as TF-ID word2vec, GLOVE and skip gram are used to perform feature selection in the deep learning CNN/ RNN framework [3]. The features that are marked in the second phase are given to third phase for text classification process. In the last phase, ML classification techniques are used on high dimensional feature space to detect polarity of given input sentences [4].

3.   Sub Task-2: Aspect Based Feature Extraction

3.1. Word2vec

Word2Vec is a word embedding deep learning model which accepts the preprocessed sentences into tokens and identifies the context of words [5]. This model works on large set of databases including number of documents from different domains like twitter, Facebook, ecommerce websites and returns a huge vector space of several

dimensions. In this model, each word in the text corpus will be defined by a unique vector in the vector space [6].
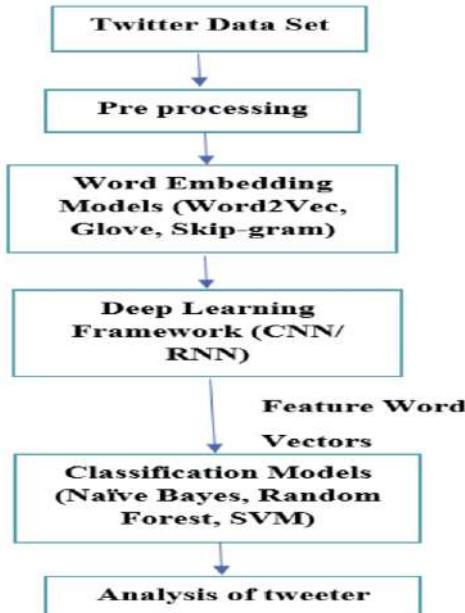
Figure 2. Aspect Based Sentiment Analysis Framework

3.2. Glove

Glove generates word embedding by training a model based on global co-occurrence counts of words, global statistics and uses mean squared error as the loss function. The generated word embedding with such a model preserves word relationships and similarities. A co-occurrence matrix in table 2 for a given sentence tells us how often a given pair of words appear together. Each element in the matrix is the count of the pair of the words occurring together [7].

3.3. Convolution Neural Networks (CNN)

Convolutional Neural Networks (ConvNets) have in the past years shown break-through results in some NLP tasks, one particular task is sentence classification, i.e., classifying short phrases (i.e., around 20~50 tokens), into a set of pre-defined categories. A Convolutional Neural Network typically involves two operations, which can be thought of as feature extractors: convolution and pooling [8].

4. Sub Task-3: Aspect Based Sentiment Classification

Most vital activity in sentiment analysis is to identify best suitable classification technique in order to detect polarity of the input text [9]. This section defines, compares and summarizes most of the sentiment classification techniques. Machine learning classification models are one of the best techniques for analyzing huge amount of text data to classify sentiments either as positive, negative or neutral. ML

Classification algorithms can be categorized as Supervised and Unsupervised learning models. These models gain popularity due to its adaptability and accuracy feature [10].

4.1. Naïve Bayes

Naive Bayes is a simple probabilistic technique in which each word of given input text data has probability associated of being positive or negative polarity. In mathematical terms, P(A/B) probability of A, if B is true, is defined as P(B/A) times the probability of A being true P(A), divided by the probability of B being true P(B)

4.2. Support Vector Machines

SVM is a supervised machine learning algorithm can be used for two group classification problems, regression and outlier detection [11]. SVM works on input trained datasets by plotting the points in 2D plane for each category and then draw a hyperplane that best separates the class labels.

4.3. Neural Networks

Artificial Neural networks are inspired by human brain neurons, which works on multiple layers and at each layer a number of processing units are defined. The output of first layer is fed as an input to the next consecutive layer and so on. Neural network can learn to perform tasks (e.g., classification) by modifying the connection weights between neurons, resembling the learning process of a biological brain [12].

4.4. Random Forests

Random Forest is a Supervised learning technique which works on ensemble learning approach. In this approach, we can either construct a hybridization of similar algorithms number of times or different algorithms can be ensembled. This will result in a dense forest by adding multiple times decision tree algorithms. Random forests are well suited for regression or classification analysis [13].

5.   Evaluation and Comparison of Results

Table 1, describes the efficiency of the aspect sentiment classification of the different classification models on the input training data. Random forest achieves better efficiency than the other conventional models on training dataset.

Table 1. Comparative analysis of different classification models and its accuracy on aspect sentiment classification

Test Data	NN	Naïve Bayes	SVM	Random Forest
Aspect Test 1	0.82	0.86	0.82	0.92
Aspect Test 2	0.8	0.86	0.84	0.91
Aspect Test 3	0.83	0.84	0.86	0.91
Aspect Test 4	0.81	0.85	0.86	0.91
Aspect Test 5	0.84	0.86	0.85	0.93
Aspect Test 6	0.84	0.86	0.87	0.92
Aspect Test 7	0.84	0.86	0.85	0.92
Aspect Test 8	0.81	0.87	0.88	0.91
Aspect Test 9	0.82	0.84	0.86	0.92
Aspect Test 10	0.81	0.86	0.85	0.92

Table 2, describes the efficiency of the aspect sentiment classification accuracy of the different classification models on the input training data. Random forest achieves better accuracy than the other conventional models on training dataset.

Table 2. Comparative analysis of different classification models and its recall on aspect sentiment classification

Test Data	NN	Naïve Bayes	SVM	Random Forest
Aspect Test 1	0.83	0.86	0.87	0.92
Aspect Test 2	0.82	0.86	0.83	0.92
Aspect Test 3	0.81	0.87	0.88	0.92
Aspect Test 4	0.82	0.84	0.85	0.92
Aspect Test 5	0.81	0.84	0.83	0.91
Aspect Test 6	0.82	0.84	0.86	0.91
Aspect Test 7	0.83	0.85	0.83	0.91
Aspect Test 8	0.81	0.85	0.85	0.92
Aspect Test 9	0.84	0.84	0.83	0.91
Aspect Test 10	0.8	0.84	0.82	0.91

Figure 3, describes the efficiency of the aspect sentiment classification recall of the different classification models on the input training data. Random forest achieves better recall rate than the other conventional models on training dataset.
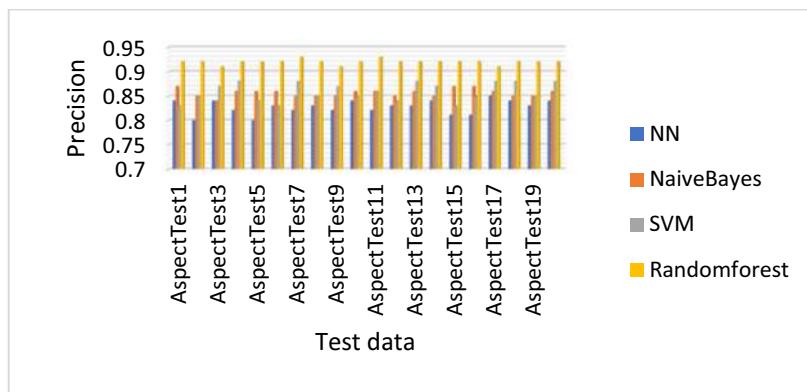
Figure 3. Comparative analysis of different classification models and its precision on aspect sentiment classification

Figure 4, describes the efficiency of the aspect sentiment classification F-measure of the different classification models on the input training data. Random forest achieves better F-measure rate than the other conventional models on training dataset.
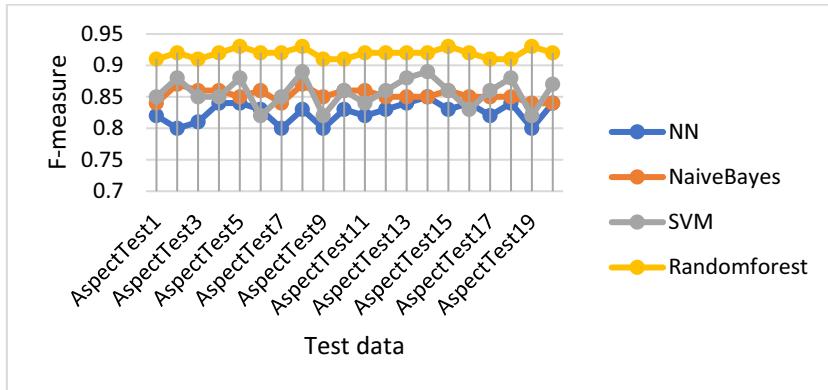
Figure 4. Comparative analysis of different classification models and its F-measure on aspect sentiment classification

6.   Evaluation and Comparison of Results

In this paper, different aspect sentiment classification models are studied on the input training data for statistical analysis. Most of the conventional classification models are difficult to find the classification efficiency due to problem of feature extraction or noise elimination problem. In this paper, different classification models are tested by integrating the word embedding measures for data classification problem. Experimental results proved that the random forest achieves better accuracy, recall and error rate than

References

[1] María del Pilar Salas-Zárate, José Medina-Moreira etal, "Sentiment Analysis on Tweets about Diabetes: An Aspect-Level Approach", Hindawi Computational and Mathematical Methods in Medicine Volume 2017, Article ID 5140631, 9 pages https://doi.org/10.1155/2017/5140631

[2] Naveen Kumar Laskari, Suresh Kumar Sanampudi, "Aspect Based Sentiment Analysis Survey", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 18, Issue 2, Ver. I (Mar-Apr. 2016), PP 24-28

[3] Nipuna Upeka Pannala, Chamira Priyamanthi Nawarathna, J.T.K.Jayakody, "Supervised Learning Based Approach to Aspect Based Sentiment Analysis", 2016 IEEE International Conference on Computer and Information Technology.

[4] M.Sivakumar, Dr.U.Srinivasulu Reddy, "Aspect Based Sentiment Analysis of Students Opinion using Machine Learning Techniques", Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017) IEEE Xplore.

[5] Alson Cahyadi and Masayu Leylia Khodra, "Aspect-Based Sentiment Analysis Using Convolutional Neural Network and Bidirectional Long Short-Term Memory", 978-1-5386-4804-9/18/$31.00 @2018 IEEE

[6] Ganpat Singh Chauhan, Yogesh Kumar Meena, "An Unsupervised Multiple Word-Embedding Method with Attention Model for Cross Domain Aspect Term Extraction", 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE2020), 07-08 February 2020

[7] Walaa Medhat, Ahmed Hassan, Hoda Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, 2014

[8] Lei Zhang, Shuai Wang, Bing Liu, "Deep Learning for Sentiment Analysis: A Survey", Wiley, January 2018, DOI: 10.1002/widm.1253

[9] Toqir A. Rana, Yu-N Cheah, "Aspect extraction in sentiment analysis: comparative analysis and survey", Published in Springer 2016, DOI 10.1007/s10462-016-9472-z.

[10] A.M.Abirami, V.Gayathri, "A Survey On Sentiment Analysis Methods And Approach" 2016 IEEE Eighth International Conference on Advanced Computing (ICoAC)

[11] Oscar Araque, Ganggao Zhu, Manuel Garcia-Amado and Carlos A. Iglesias, "Mining the Opinionated Web: Classification and Detection of Aspect Contexts for Aspect Based Sentiment Analysis", 2016 IEEE 16th International Conference on Data Mining Workshops.

[12] Ruijun Liu, Yuqian Shi, Changjiang Ji, And Ming Jia, "A Survey of Sentiment Analysis Based on Transfer Learning", IEEE Access, 2019 DOI: 10.1109/ACCESS.2019.2925059

[13] Soujanya Poria, Erik Cambria, Alexander Gelbukh, "Aspect Extraction for Opinion Miningwith a Deep Convolutional Neural Network" Knowledge-Based Systems Published by Elsevier, 2016 DOI: 10.1016/j.knosys.2016.06.009

Impact of Internet of Things (IoT) in Smart Agriculture

O. Vishali Priya [a,1], Dr.R. Sudha [a]

[a]*Department of Computer Science,PSG College of Arts & Science, Coimbatore, Tamilnadu, India*

Abstract. In today's world, technology is constantly evolving; various instruments and techniques are available in the agricultural field. And within the agrarian division, the IoT preferences are Knowledge processing. With the help of introduced sensors, all information can be gathered. The reduction of risks, the mechanization of industry, the enhancement of production, the inspection of livestock, the monitoring of environment conditions, the roboticization of greenhouses, and crop monitoring Nearly every sector, like smart agriculture, has been modified by Internet-of-Things (IoT)-based technology, which has shifted the industry from factual to quantitative approaches. The ideas help to link real devices that are equipped with sensors, actuators, and computing power, allowing them to collaborate on a task while staying connected to the Internet, dubbed the "Internet of Things" (IoT). According to the World Telecommunication Union's Worldwide Guidelines Operation, the Internet of Things (IoT) is a set of sensors, computers, software, and other devices that are connected to the Internet. The paper is highly susceptible to the consequences of its smart agriculture breakthrough.

Keywords.: Internet of Things, Smart Farming, Smart Agriculture, Application of IOT, Benefits of Agriculture, Implications of implementation, Working of Agriculture

1. Introduction

This theory stems from advancements including the Internet of Things, Big Data, and Cloud Computing, among others, which have given rise to the concept of smartness. Farming IoT is a network of monitors, cameras, and computers that can all function together to help a farmer perform his job more effectively. This computer would be self-sufficient so that they will be able to interact with one another without the need for human intervention. To put it another way, the gadgets are pre-programmed with the knowledge of the moment and the motives for communicating with other instruments in the scheme.

Several agricultural industries have switched to IoT engineering for smart farming to increase productivity, performance, global market, and other features such as minimal human interference, time, and cost, among others. The advancement in technology ensures that sensors are becoming smaller, more modern, and more affordable. The

systems are also easily accessible and all-inclusive, allowing smart farming to be carried out with complete confidence. The answer to the challenges that this sector is currently experiencing is keen cultivation, which focuses on empowering advancement to agriculture. Much of this can be achieved with the aid of mobile phones and IoT gadgets. A farmer may obtain any need data or information, as well as monitor his agricultural field.

2.   Internet of Things (IoT)

The Internet of Things (IoT) is the most productive and essential methodology for designing solutions to problems. IoT grows from a number of building blocks, such as sensors, applications, network elements, and other electronic devices. Furthermore, it improves the effectiveness of knowledge. IoT allows data to be exchanged across a network without the need for human intervention.

In the Internet of Things, be prepared to converse with objects in a natural fashion, rather than in the manner of a regular human being, such as a sensor, a car driver, and so on. This object has been given an address so that it can transmit data across a network. According to Gather's study, there would be a 30% increase in connected computer checks by the end of 2016 relative to 2015. He predicts that by 2020, this figure will rise to 26 billion.[1].

Because of the following considerations, IoT technology is more efficient:

- Access to the internet from every laptop.
- As little physical effort as possible
- Improved Access
- Time Management
- Communication that effective

3.   Smart Agriculture Using IoT

Agriculture is India's economic development's most important pillar. Climate change is the most important boundary that conventional farming faces. Overwhelming flooding, the most extreme hurricanes and warm winds reduced rainfall, and other climatic shifts are among the many consequences. As a consequence of these considerations, performance suffers greatly. Climate change often has natural effects, such as periodic shifts in plant lifecycles.

In order to increase productivity and reduce boundaries in the farming sector, creative creativity and Internet of Things strategies were needed. The Internet of Things (IoT) is now shifting its focus to the agricultural sector, enabling farmers to face the immense obstacles they face. Using IoT, farmers may gain access to a wealth of data and knowledge regarding future trends and innovation.

Since global agriculture is becoming more industrialized, it is critical to establish agricultural interdepartmental cooperation at the same time. Agricultural intercolumniation has resulted in a positive change in global agriculture.

As far as agricultural improvement is concerned, rural intercolumniation can be a major impediment to agricultural progress and reform and a cornerstone for sustaining stable and sustainable economic growth. We've been concentrating on agricultural data gain and framework improvement for a while now. Exceptional results had been shown in rural system improvement after several years of challenging efforts [2].

Smart agriculture is a broad term that refers to agricultural and food production practises that make use of IoT, big data, and advanced analytics. The Internet of Things refers to the integration of sensing, automation, and analytics technology into present agricultural processes. The most common IoT applications in smart agriculture are as follows:

- ✓ Sensor-based systems for monitoring crops, soil, fields, animals, storage facilities, and just about anything else that has an impact on productivity.
- ✓ Smart agricultural vehicles include drones, autonomous robots, and actuators.
- ✓ Connected agricultural settings include smart greenhouses and hydroponics.
- ✓ Data analytics, visualization, and management systems.

The Internet of Things, like other industries, offers hitherto impossible efficiency, resource and expenditure savings, automation, and data-driven operations in agriculture. However, in agriculture, these benefits aren't improvements; they're cures for an entire sector beset by a host of serious problems [3].

3.1. Extremely efficient

Agriculture is become a competition. With deteriorating soil, dwindling land supplies, and increasing weather unpredictability, farmers must produce more. IoT-enabled agriculture allows farmers to monitor their products and conditions in real time. They have fast perceptions, can predict issues before they arise, and make well-informed decisions on how to avoid them. Agriculture IoT solutions include automated features such as demand-based watering, fertilisation, and robot harvesting.

3.2. Expansion

By the time we reach 9 billion people, 70% of the world's population will be living in cities. IoT-based greenhouses and hydroponic systems, which should be able to feed these folks with fresh fruits and vegetables, enable short food supply chains. Thanks to sophisticated closed-cycle agricultural systems, food may be produced in supermarkets, on the walls and rooftops of buildings, in shipping containers, and, of course, in the comfort of everyone's home.

3.3. Resources are in short supply.

Many agricultural IoT solutions seek to make the most efficient use of resources like water, power, and land. Precision farming is based on data acquired from a variety of sensors in the field, allowing farmers to precisely allocate exactly adequate nutrients to one plant.

3.4. Hygienic Procedure

Smart farming using IoT is a tried-and-true method of reducing pesticide and fertiliser consumption. Precision farming not only saves water and energy and makes farming more ecologically friendly, but it also drastically decreases the need of pesticides and fertilisers. In compared to traditional agricultural methods, this technique yields a cleaner, more organic end product.

3.5. Agility

One of the benefits of using IoT in agriculture is the increased adaptability of operations. Thanks to real-time monitoring and forecasting technologies, farmers can react quickly to any significant change in weather, humidity, air quality, or the condition of any crop or soil in the field. Agriculture specialists can now save crops in the face of extreme weather changes thanks to new skills.

3.6. The quality of the product has improved.

Data-driven agriculture aids in the production of more and better commodities. Using soil and crop sensors, overhead drone monitoring, and farm mapping, farmers may gain a better understanding of the complex relationships between the environment and the quality of their crops. Using connected systems, they can replicate ideal conditions and increase the nutritional value of the goods.

The whole handle revolves around the gathering of data for farmers and other collaborators to use. Usually, the most critical feature of the activity. Sensors, microphones, and fawning pictures are among the gadgets included. The second segment consists of an agreement that will assist in sharing knowledge provided by the machines, as previously mentioned. Various types of network advancements, such as GSM, LTE, Wi-Fi, and 3G, may be used based on connectivity and requirements. The third section, similar to Cloud administrations, consists of data processing and computing creativity.

Cloud servers may be made open without any limitations, rendering them suitable for IoT frameworks. On such servers, knowledge may be stored and processed. Cloud administrations may be accessed on a pay-per-use basis, although they are becoming more common for this purpose. The system's final component will be the Big Data analytics tools, which will sift through the vast amounts of data generated and stored on cloud storage to find important trends and patterns.

The steps in working of Agriculture IOT:

Smart mobile data collection

Network-based data transfer

Cloud-based data collection and computing

Use large data software to analyze data

4.   Major Applications

Each aspect of conventional farming strategy can be modified on a very simple level by implementing the most recent detecting and IoT advances in agriculture practices. Right now, consistent convergence of wireless sensors and the Internet of Things in smart agriculture will propel agriculture to previously unimaginable heights[4]. IoT may help advance the solutions to various typical farming problems, such as dry spell reaction, surrender optimization, arrive reasonableness, water method, and bother regulation, by adopting the tenets of savvy agriculture [5].

4.1.   Climate Change: Agriculture is significantly impacted by climate change. Furthermore, a lack of understanding of climate has a substantial influence on agricultural production quantity and quality. IoT technologies, on the other hand, allow you to track weather conditions in real time. Sensors have been deployed both within and outside of agriculture regions. They collect data from the environment, which is then used to determine which crops are most suited to growing and surviving in the given climatic conditions. Sensors are used across the Internet of Things ecosystem to accurately monitor real-time weather factors such as humidity, rainfall, temperature, and more. There are a variety of sensors available to monitor and customise all of these factors to fit your smart farming demands. These sensors monitor the health of the crops as well as the weather conditions around them. When unexpected weather conditions are found, an alarm is dispatched. The necessity for human presence during inclement weather is removed, boosting output and allowing farmers to enjoy additional agricultural advantages.

4.2.   Precision Farming: One of the most well-known IoT applications in agriculture is precision farming, often known as precision agriculture. Animal monitoring, vehicle tracking, field observation, and inventory monitoring are examples of smart farming applications that serve to make farming more precise and controlled. The goal of precision farming is to assess data supplied by sensors and respond appropriately. Precision farming allows farmers to collect data from sensors and analyse it to make educated and timely decisions. Irrigation management, livestock management, vehicle tracking, and other precision agricultural procedures are all vital in increasing efficiency and effectiveness. Precision farming allows you to evaluate soil conditions as well other pertinent data in order to increase operational efficiency. Not only that, but you may also check water and nutrient levels by monitoring the associated devices' real-time operational conditions.

4.3.   Smart Greenhouse: The Internet of Things has allowed weather stations to change climatic conditions automatically in response to a set of instructions, allowing us to make our greenhouses smart. In greenhouses, the integration of IoT has eliminated the need for human intervention, making the entire process more cost-effective while also increasing accuracy. Solar-powered Internet of Things sensors, for example, may be utilised to build modern, low-cost greenhouses. These sensors collect and transmit real-time data, allowing for precise real-time greenhouse monitoring. Thanks to the sensors, the water use and greenhouse state may be tracked via emails or SMS messages. Irrigation is carried out automatically and intelligently via the Internet of Things. Pressure, humidity, temperature, and light levels may all be measured with these sensors.

4.4.   Data Analytics: IoT device data necessitates more storage than a conventional database system can supply. Cloud-based data storage and an end-to-end IoT platform are critical components of the smart agriculture system. These systems are expected to play a key role in facilitating the execution of better tasks. In the IoT era, sensors are the primary source of large-scale data collecting. The data is analysed and transformed into valuable information using analytics tools. Data analytics may be used to analyse weather, livestock, and agricultural conditions. Using technological improvements, the knowledge obtained enables for better decision-making. You may learn about the real-time status of your crops by collecting data from sensors utilising IoT devices. Predictive analytics can help you gain insight into harvesting decisions and make better ones. Farmers can utilise trend analysis to forecast upcoming weather and crop harvesting circumstances. IoT has assisted farmers in preserving crop quality and soil fertility, resulting in increased production volume and quality in the agriculture business.

4.5.   Aerial Drones in Agriculture: Agricultural operations have almost totally altered as a result of technology advancements, with the most recent disruption being the introduction of agricultural drones. Drones are used for agricultural health checks, crop monitoring, planting, crop spraying, and field analysis on the ground and in the air. Thanks to suitable strategy and planning based on real-time data, drone technology has given the agriculture industry a boost and makeover. Drones with thermal or multispectral sensors detect areas where irrigation has to be adjusted. As the crops develop, sensors assess their health and calculate their vegetation index. The environmental impact of smart drones was finally reduced. As a consequence, the quantity of chemical that reaches the groundwater has decreased significantly [6].

Agriculture facilitated by the Internet of Things has assisted in the adoption of cutting-edge technology solutions to age-old knowledge. This has helped to close the gap in output, quality, and yield. Quick reaction and minimum crop damage are ensured by data obtained by collecting and importing data from multiple sensors for real-time use or storage in a database. Because to end-to-end intelligent operations and improved business process execution, produce is processed faster and reaches supermarkets in the shortest time feasible [7].

5.   Implications of Implementation

Agriculture IoT implementation on a wide scale is possible, as it were, with the government's support [8]. It will aid in the decision-making process by delivering user-friendly plans and strategies. It can have subsidized gadgets and bases that farmer can't handle in any other way. Agriculture supply chain flaws must be addressed [9]. To offer the most severe benefits to farmers and consumers, the position of an agent should be re- examined and handled [10].

5.1 The apparatus

To build an IoT solution for agricultural, you must first choose the sensors for your equipment (or create a custom one). The kind of data you want to collect and the ultimate objective of your solution will guide your selection. In any case, the precision and dependability of the data obtained will decide the success of your product, therefore the quality of your sensors is important.

5.2 The brain

Data analytics should be part of any smart agriculture solution. It'll be pointless if you can't make sense of the data you've acquired. You'll require advanced data analytics abilities, as well as prediction algorithms and machine learning, to derive valuable insights from the acquired data.

5.3 Performing routine maintenance

Maintaining your hardware is a big challenge for IoT devices in agriculture because sensors are frequently used in the field and can be easily destroyed. As a result, you must make certain that your hardware is both durable and repairable. You'll have to replace your sensors more frequently than you'd want if you don't.

5.4 Movement ability

Agricultural applications should be particularly built for use in the field. For a business owner or farm manager, the information should be available on-site or remotely by smartphone or desktop computer.Each linked device should also be self-contained and have enough wireless range to communicate with other devices as well as provide data to the central server.

5.5 smart-farming-app-development infrastructure.

To ensure that your smart farming application functions well, you'll need a solid internal architecture (and can manage the data load). Furthermore, your internal systems must be protected. If you fail to effectively safeguard your system, breaking in, stealing your data, or even gaining control of your autonomous tractors becomes more possible.

5.6 Interdependence

The necessity for data to be sent across several agricultural locations continues to be a roadblock to smart farming adoption. Naturally, the connectivity connecting these facilities must be dependable enough to withstand adverse weather and offer continuous service [11]. Despite current efforts to develop universal standards in this area, IoT devices currently use a range of communication protocols. The development of 5G and technologies like space-based Internet will, hopefully, help to resolve this issue.

5.7 The amount of data collected on a regular basis

Because there are so many different forms of data in the agriculture industry, finding the optimal data collection frequency may be challenging. Data from field-based, aerial, and environmental sensors, applications, devices, and equipment, as well as processed analytical data, are all subject to restrictions and laws. The secure and timely delivery of this data, as well as its interchange, is one of the most difficult aspects of smart farming.

6.   Conclusion

As a result, smart agriculture is needed. The Internet of Things would assist in the development of smart agriculture. IoT is used in various agricultural domains to

improve time efficiency, water conservation, crop monitoring, soil management, bug spray and pesticide safety, and so on. It also eliminates human labor, deconstructs agricultural methods, and creates a difference in smart farming implementation. Agriculture is a profession that has relied on traditional practices and experiences up to now. However, the passage of time has affected rural traditions, and they have started to adapt to the flow of change. The usage of the Internet of Things (IoT) in agriculture would increase yields and help control all cultivate operations. Agriculture must undergo significant improvements since a substantial portion of the population relies on it for survival.

Reference

[1]   Babu Loganathan, Ganesh. "Smart Agriculture System With E–Carbage Using Iot." International Journal of Modern Agriculture 10, no. 1 (2021): 928-931.
[2]   Haseeb, Khalid, Ikram Ud Din, Ahmad Almogren, and Naveed Islam. "An energy efficient and secure IoT-based WSN framework: An application to smart agriculture." Sensors 20, no. 7 (2020): 2081
[3]   Miles, Badreddine, El-Bay Bourennane, Samia Boucherkha, and Salim Chikhi. "A study of LoRaWAN protocol performance for IoT applications in smart agriculture." Computer Communications 164 (2020): 148-157.
[4]   Ratnaparkhi, Sanika, Suvaid Khan, Chandrakala Arya, Shailesh Khapre, Prabhishek Singh, Manoj Diwakar, and Achyut Shankar. "Smart agriculture sensors in IOT: A review." Materials Today: Proceedings (2020).
[5]   Kassim, Mohamed Rawidean Mohd. "IoT Applications in Smart Agriculture: Issues and Challenges." In 2020 IEEE Conference on Open Systems (ICOS), pp. 19-24. IEEE, 2020.
[6]   Kumar, Dr A. Senthil, Dr AR, L. Ganesh Babu, and Dr G. Suresh. "Smart Agriculture Robo With Leaf Diseases Detection Using IOT." European Journal of Molecular & Clinical Medicine 7, no. 11 (2022): 2462-2469.
[7]   Marcu, Ioana, George Suciu, Cristina Bălăceanu, Alexandru Vulpe, and Ana-Maria Drăgulinescu. "Arrowhead technology for digitalization and automation solution: smart cities and smart agriculture." Sensors 20, no. 5 (2020): 1464.
[8]   Rasooli, Mohammad Wasi, Brij Bhushan, and Nagesh Kumar. "Applicability of wireless sensor networks & IoT in saffron & wheat crops: A smart agriculture perspective." Int. J. Sci. Technol. Res 9, no. 2 (2020): 2456-2461.
[9]   Prasad, Shitala, Sateesh K. Peddoju, and Debashis Ghosh. "Energy efficient mobile vision system for plant leaf disease identification." In 2014 IEEE Wireless Communications and Networking Conference (WCNC), pp. 3314-3319. IEEE, 2014.
[10] Dhaware, Chaitali G., and K. H. Wanjale. "A modern approach for plant leaf disease classification which depends on leaf image processing." In 2017 International Conference on Computer Communication and Informatics (ICCCI), pp. 1-4. IEEE, 2017.
[11] Singh, Vijai, and A. K. Misra. "Detection of unhealthy region of plant leaves using image processing and genetic algorithm." In 2015 International Conference on Advances in Computer Engineering and Applications, pp. 1028-1032. IEEE, 2015.

Real Time Tracing and Alerting System for Vehicles and Children to Ensure Safety and Security Using Labview

Dr. R. Deepalakshmi [a,1], Dr. R. Vijayalakshmi [a], S. Lavanya [a], T.K. Rakshitha Rasmi [a], S.B. Sathiya [a]

[a] *Department of Computer Science and Engineering, Velammal College of Engineering and Technology, Madurai, Tamil Nadu, India*

Abstract – The Absolute time monitoring, detecting and Alerting System for vehicles and children is required to trace and transmit the collected information at regular intervals to ensure safety and security of children. The illustration of the Realtime detecting and warning System consists of two units: Tracing Unit that traces the location information, transfers to the monitoring area, records the data in the database and takes the help of these data to locate the exact point of area of the vehicle with Google/other maps. The second unit is Alerting Unit that tracks the students using active Radio Frequency Identification Devices (RFID)which will be placed on student ID card. radio- wave trans-receiver transmits a common radio wave which is received by the RFID in the ID card. This radio-wave is modified by the RFID's coil and resent to the receive RFID tags are also used for attendance which is updated directly to the database and displays the other student information.

Keywords: Tracking, Laboratory Virtual Instrument Engineering (LabVIEW), Position, Radio Frequency Identification (RFID), Alerting, RFID Tags, RFID reader.

1. Introduction

Safety and security are the most important aspects for students and the management to prevent students from abduct, mislaid etc., In the existing passive tracking system, a sensor device is placed in the vehicle that records position, speed etc., and also when vehicle backs to the specified position the device is deleted and data transferred to the system. It also includes automatic downloading option which transfers information through radio link but were not absolute time. Passive system does not help to be more helpful to trace vehicles to prevent accidents. Absolute time tracing system was needed to send the stored information of the vehicle at regular intervals or at least transmit the information when required by the monitoring space. In this system, a device is fixed in the vehicle which will be interlinked with LabView to track the vehicle in real-time, locate it on the Google/other map and the alarming system included with alerting the client about the vehicle location only. This does not concentrate on the passengers in the vehicle. This led to the making of active modules. Hence this automatic vehicle tracing system will help you to track the particular target vehicle as well as the students in real time and also alerts the parents and school management at regular intervals of time by using the

---

[1] Deepalakshmi, R, Professor, Velammal College of Engineering and Technology, India
Email: rdl@vcet.ac.in.

software called Lab View for vehicle tracking and RFID for student tracking. The active system is developed for the real time tracking of particular target vehicle at regular interval of time. By using LabView, it captures the image of the vehicle and starts tracking it. The vehicle information is always stored in the database such as time, speed, location which is connected with Google Maps. The students who are travelling in the vehicle are also tracked using RFID. Each student contains a tag that has a unique code. RFID system uses an electromagnetic field to transmit data from RFID to tracker. RFID, indeed provides accurate and real time tracking data for fixed and mobile assets. The vehicle and student information are stored in the database. Alerting system alert the parents and the management by getting information from the database linked whenever required.

2. System Requirements

- • Browser: Firefox, Edge, Chrome, etc.,
- • Operating System: Windows 10 with core i5 processor
- • IDLE: LabVIEW version 18.0 (32-bit)

3. Architecture of the project

The domain of project consists of two modules: The first module is Vehicle tracking, this module tracks the target vehicle in real time by using the software LabView and stores the location information, vehicle video and stores the collected information in database which is interlinked with the software. Vehicle tracing is the process of pointing a moving material in using a lens/camera. The algorithm senses the frames and outputs the position of material within the frame. School Vehicle Tracking System uses LABVIEW software that is used to track the vehicle.

This is the Algorithm of Vehicle tracking using LabVIEW simulation tool:

Step 1: Capture and save the image of the vehicle using the LabVIEW simulation tool

Step 2: Check if the captured vehicle is same as the target vehicle.

- a) If YES, track the vehicle until the destination is reached.
- b) If NO, capture and check for target vehicle until found.

Step 3: Once the target vehicle is found and tracked, check if the location information is available.

- a) If YES, store it in the database server.
- b) If NO, again continue with step 2b.

Step 4: Make sure the location information stored serves the needful and use them track the vehicle in Google maps.

The role of the tracing algorithm is to analyze the frames to calculate the moving measurements. These measurements characterize the position of the target vehicle. They help to identify other factors such as speed, number of route changes, total time moving and also information about size of the target vehicle.
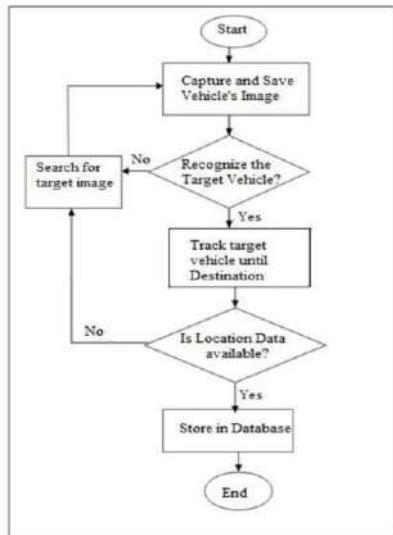
Figure 1. Vehicle tracking system flow diagram
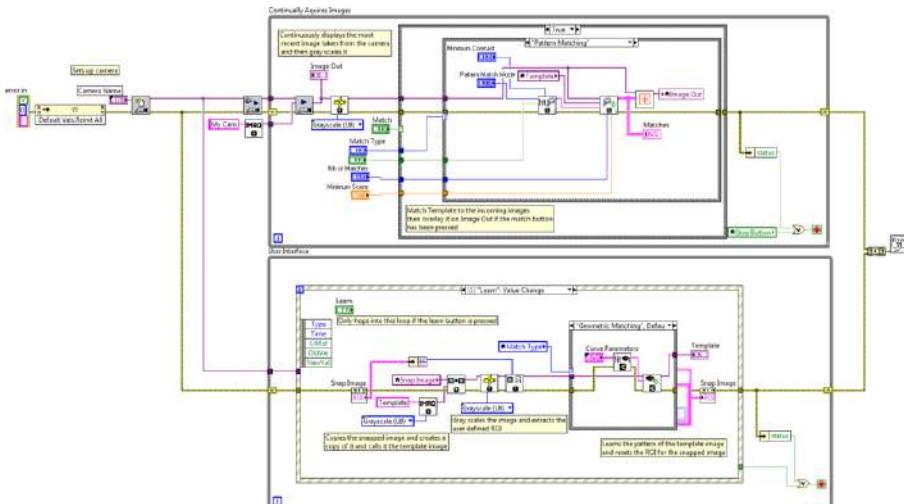
Figure 2. Block Diagram of Vehicle Tracking

The second module is student tracking system. It recognizes the student who enters the vehicle at the source location by reading the tag holds by the student using RFID (Radio Frequency Identification), and monitors the presence and absence of students and stores the information in the database. By integrating this system with DBMS (Database Management System), a database can be maintained to monitor the speed and the location at which the vehicles had travelled. The alerting system can also be provided in case of any speed violation by the traveler. The speed limit can be set by the owner of the vehicle and they can receive notifications through mail in case the vehicle travels above the speed limit specified by the user.

The following is the Algorithm for student tracking using RFID Tag:

Step 1: When the student enters the bus, the RFID tag details are recognized by the sensor and the data is sent to the RFID reader.

Step 2: This reader decrypts the message and sends it the Attendance application.

Step 3: The Attendance application compares the information with the database.

Step 4: Checks if the data matches with the Database

  a)   If YES, Attendance is marked, saved and acknowledged.

  b)   If NO, Absence is marked and notification is sent to the parents and alerts them.

Step 5: Parents are acknowledged on students reaching the destination.
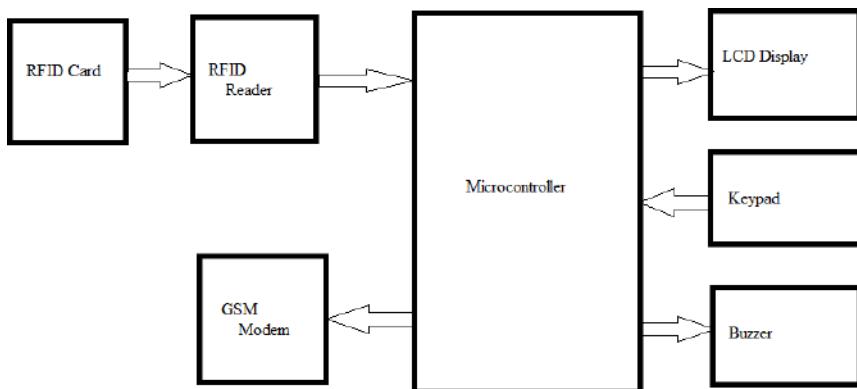
Figure 3. Student Tracking and Attendance System

4. Module

  1.   Vehicle Tracking using LabVIEW.
  2.   Student Tracking using RFID.
  3.   Alerting System

4.1. Vehicle Tracking using LabVIEW

This module tracks the vehicle using LabVIEW. LabVIEW is just a simulation tool. LabVIEW (Laboratory Virtual Instrument Engineering Workbench) is a model-creating platform. It is created by National Instruments (NI). LabVIEW is used for automated manufacturing, test, validation, take charge and monitoring of a system. For real time tracking of vehicle, LabVIEW captures the image of the target vehicle by using the in- built camera and then highlights the captured image using rectangular red box. Once, when the LabVIEW diagram and the vehicle image gets simulated it starts tracking the vehicle until it reaches the destination. While tracking the vehicle, information like vehicles location and time information gets stored in the database.
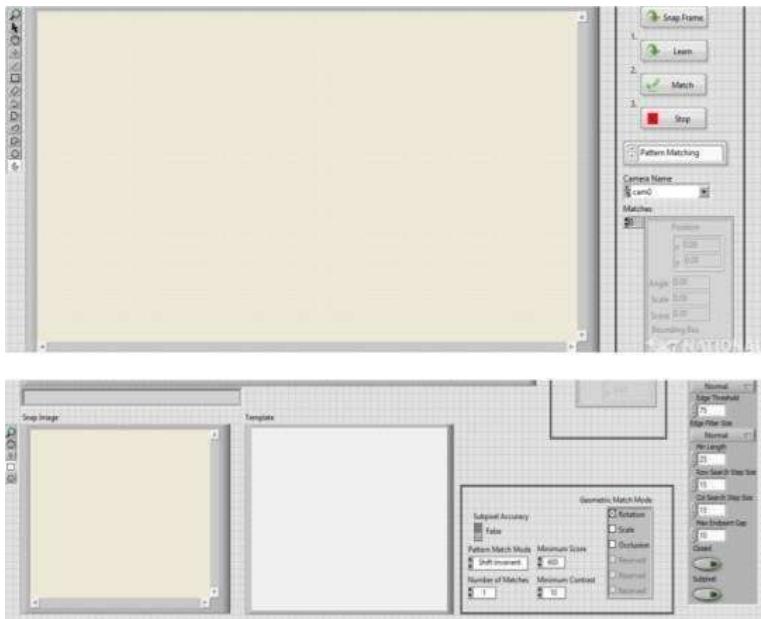
Figure 4. LabVIEW Implementation

4.2. Student tracking using RFID

Parents need ways to monitor the safety of their children during their journey to/from school. In this module, Student get tracked by using the technology RFID. RFID is used for data collecting that includes automatic identification of materials using tags, an antenna, an RFID reader and a transceiver.

Step 1: Wearable RFID tags are given to all students.

Step2: RFID reader are installed in the vehicle. When the student enters the bus, it automatically scans the reader.

Step3: RFID identifies the tag number and stores the student information such as students source location, absence of students and the entering time the database.

4.3. Alerting system

Collected information from the first module is stored in the vehicle database, and the information about the students are stored in relevant ID. In the module, a solution is provided using Radio Frequency Identification (RFID) technology in order to track students journey to/from school thus notifies the parents about the status of children. This reads the children's RFID tag and automatically sends different generated messages to their parents. The proposed solution has a set of capabilities, such as notifying the parents that the children has arrived to the school safely, and if any children are absent without an authorized permission then an immediate message is sent to their parents.

5. Conclusion

The system mainly aims and concentrates on making a system to trace the location of vehicle and the passenger in absolute time, point it on the Google/other maps and alarming module added to heed user of the module if diversion/interruption in route takesplace as well as when students are missing/absence of students. The system could yet be used to trace vehicles and alert passengers from misfortune, also provide emergency/fast services in case of any such situations. This system can be expanded to identify vehicles which looks similar to each other by using their number plate, the student tracing can be done more precisely to avoid misusing of tags by students and also calculating the Body Mass Index of an individual for identifying the particular student.

References

[1]   Dao, D., Rizos, Ch., Wang, J., (2002).Location-based services: technical and business issues, GPS Solutions Vol.6 .

[2]   David Moore,(June 5,2003). A real-world system for human motion detection and tracking, California Institue of Technology.

[3]   Prawat Chaiprapa, Supaporn Kiattisin and Adisorn Leelasantitham ,(2003). A Real-Time GPS Vehicle Tracking System Displayed on a Google- Map-Based.

[4]   Guochang Xu,(Springer 2007).GPS Theory,Algorithms and Applications.

[5]   F. M. Franczyk, and J. D. Vanstone, (22 April 2008). Vehicle warning system, Patent number: 7362239.

[6]   Thuong Le-Tien, Vu Phung, (July,2009).Routing and Tracking System for Mobile Vehicles in Large Area,Vietnam.

[7]   ThuongLe-Tien, Vu Phung, (IEEE,2010).Routing and Tracking System for Mobile Vehicles in Large Area,Vietnam.

[8]   Reddy, S. S., (2011).Trip Tracker Application on Android, San Diego State University, USA.

[9]   Dinkar, A. Sh., Shaikh, S. A., (2011). Design and Implementation of Vehicle Tracking System Using GPS, Journal of   information Engineering and Applications.

[10] Harshadbhai, P. K., (2013).Design of GPS and GSM Based Vehicle Location and Tracking System, International Journal of Science and Research,Vol. 2,No. 2,pp. 165-168.

[11] Lee, S., Tewolde, G., Kwon, J.,(IEEE 2014) ,.Design and implementation of vehicle tracking system using GPS/GSM/GPRS technology and smart phone application.

Air Quality Monitoring System

G. Vijay Kumar [a ,1], K. Kiran Kumar [a], Bravara Manohar [a], Chintha Venkatesh [a], Shaik
Chand Basha [a], Shaik Javeed Akthar [a]

[a] *Department of Computer Science & Engineering, Koneru Lakshmaiah Education
Foundation, Greenfields, Vaddeswaram, Guntur-522502, India*

Abstract. Safeguarding and checking of air quality has becoming a most important
practice in various modern and metropolitan zones now a days. The behaviour of
air is unfavourably influenced by various types of contamination that we get by the
transportation in production of power, natural resources utilizes and so for the
emission of different poisonous gases is making a major danger for the personal
life in various urban communities. The contamination of air is getting expanded
day by day. Our main task is to provide a effective air contamination prediction
observing techniques by which we can gather the data about infectious or
poisonous gases present in every zone and gives us an idea of air contamination in
each zone. Therefore, air quality measuring has become amongst important
technique. The shape of air is influenced by multi-dimensional elements contains of
area, time and unsure factors. As of now, various specialists get started to use the
very large information investigation methodology because of various ways in large
information applications and using the resources of nature detecting organisations
and information sensed by sensor. The examination researches different various
information and air quality evaluation methods using artificial intelligence
methods. Moreover, it identifies and predict the future explanation needs.

Keywords. Air Quality Evaluation, Air Quality Prediction, Data Mining, Random
Forest, Decision Tree.

1.   Introduction

Air is the very important resources that we get from the nature. By inhaling this air
various living organisms survive their life on this planet. All the living organism's life
in all forms it is directly attached on the air for survival. Living organisms need
sufficient amount of fresh air from emission of poisonous gases on this planet. Based
on a survey in the past decades amount of air pollution has moved above 20% by the
three analyses. Global warming is one most severe problem of the environment. In the
atmosphere CO2 level and greenhouse effect increasing day to day. These drastic
changes increase radiation and the sunlight makes the planet hotter, this change in
atmosphere is called greenhouse effect, it causes to rise in earth's mean temperature.
Air contamination is one the major concerning issue in smart cities.

---

[1] G Vijay Kumar,Department of Computer Science & Engineering, Koneru Lakshmaiah Education
Foundation Greenfields, Vaddeswaram, Guntur-522502
Email: gvijay_73@kluniversity.in.

Emission of poisonous gases in to the environment or releasing of poisonous gases in to the atmosphere causes direct or indirect problems to many living organisms and other species of the eco system. Contamination of air has major severe problem of air by various ways of emission contaminants in to the atmosphere. The substances which get pollute air and also leads to serious health issues and environment threat problems [1].

Smoke venting from vehicles is the reason for air contamination. The petroleum products burning are generous wellspring of sulphur dioxide. The raised degrees of sulphur dioxide from power plant and autos are the astounding impact of corrosive downpour. This downpour brings about consumption of metals and disintegration of structures. The essential presence of environments is through Ozone layer [2]. It is liable for safeguarding environments from poisonous bright (UV) beams. The uplifted environmental chlorofluorocarbons cause contamination in atmosphere. All in all, are responsible for arrangement off a few imperfections in birth, respiratory issues, heart issue, malignancy and nerve harm. By Air pollution, every year seven million people are dying worldwide. The data with WHO shows that 9 out of 10 people breathe air that exceeds WHO guideline limits containing high levels of pollutants or dust particles, with low and middle-income countries are suffering from the greater exposures. WHO is supporting countries to help on air pollution?

2.   Related Work

Many researches are conducted in this field to improve the quality of air and to measure the amount of fresh air by using artificial intelligence and data mining. The problem of Quality of Air contamination is a big issue globally due to increase of industries and urbanization. Pollution levels around the globe are increasing day by day which is alarming. Different types of Air Quality system are proposed and are designed for monitoring pollution [3] levels at different levels. The major problem of the system's present for monitoring is that they require continuous power supply and Wi-Fi connectivity to work. The expansion of smart cities requires intelligent devices with rapid connectivity [4]. The usage and power of Wi-Fi is increasing, which can connect different physical device to the Internet. IOT provides access of internet to the object-flow of information. Earlier, WSN (Wireless Sensor Network) and RFID (Radio Frequency Identification) are used as targeted technologies in IOT. There is a possibility of finding regular patterns [5, 6] in air pollution system. There are many designs on surveillance of the issue through WSN. IOT enabled sensor networks were deployed where transmission node composed of sensing entity and connects to receiver node in wireless mode. The sensors which are fixed indifferent regions in different locations gathers data and next the data is transmitted to station where status report is generated [7].

3.   Problem Statement

The main objective in this paper is, we are going to measure of infectious or poisonous gaseous particles that obtained in environment and as well we have to predict the coming days atmosphere conditions utilizing data mining & AI procedures with iot. By using machine learning& data mining techniques for future prediction is possible and we can predict the climatic conditions. Data mining, ML these are the present booming technologies across globe for the coming up generation.

4.   Methodology

Air quality prediction system contains various data collected with huge amount of monitoring devices by the EPA, Control agencies or offices. AQS contains geological, about informative to each controlling station, and data quality control. This data is very useful to evaluate quality of air, non-compliance, assess regional and as well as national plan implementation in urban areas, perform modelling for required analysis &other type of qualities of air management functions. According to the "CLEAN AIR ACT", where AQS information is helpful to reports. Random forest algorithm uses technique where several decision trees constructed from subset of data and a summary of predictions. Use of this method is used to determine air quality for sensing system. The city data generated in sensing includes weather report, road data information, such real-time conditions. The use of Chinese standard that AQI is measured from six atmospheric gas levels, namely nitrogen dioxide (NO2), sulphur dioxide (SO2 and suspended small particles which have an aerodynamic diameter is less than 10 μm, which is to be not greater thanPM10 suspended particle.

4.1 Data Mining and Machine Learning Techniques

AI& ML are the fields where a lot of immense rises in usage in last year. The technology of Artificial Intelligence in which machine makes decision on its own, rather than running most effective with the aid of using orders given by programmer as conventional programming works, progressively began out influencing all components of our life. Starting from early-stage, start-up companies and finishing to big companies, for all of them, Artificial intelligence &it's part machine learning is becoming the key areas. Learning technique in where machine which implements artificial intelligence gathers data from sensors in an environment and learns& decide show to act. One of the reasons why we have opted ML to predict the quality of air index, was this ability of adapting of machine learning algorithms.

Data mining is a method of extracting information and finding patterns in data sets and methods using database &machine learning systems. Data mining is a computer science and statistics interdisciplinary subfield with the aim of extracting useful data sets and transforming them into a human-readable structure.

4.2 Data Mining Algorithms

Decision Tree Algorithm: The idea of decision tree algorithms is to divide data into classifications so that an algorithm can predict where new data points will land. The values of independent & dependent variables are used for these classifications.

Random Forest Algorithm: Random Forest is process having multiple trees, a forest of trees. Those trees can be of equal type or the forest can be made of merging of trees (algorithms). Here are few interesting further metaphorical thoughts that describe how the forest acts (decides).

4.3 Air Quality Evaluation

Air quality prediction is most used &important monitoring technique and evaluating quality of air. Influence for supply of gas characteristics is its adaptability for specified purposes. Several air pollutants are called Air pollutants. These contaminations can affect the health &endanger your health. Current standard the pollutants are:

- Particulate matter (PM)
- Sulphur Dioxide (SO2).
- Ozone (O3) ,Nitrogen Dioxide (NO2)
- Lead (Pb)
- Carbon Monoxide (CO)

AQPS stores the data of air pollution received from EPA, state, local agency from thousands of air sensors. AQPS also stores the descriptive weather data Information about each station and quality of data Assurance/quality information control. AQPS data are used for sensing and to assist in meeting/not meeting standards Designate and evaluate the national implementation plan non-arrival areas, modelling for permitting the analysis of review and other air quality prediction management. Productivity management system This information is also used for report preparation for Congress because "Clean Air Act" regulations.

4.4 Air Quality Standards

The Office of Air Quality Plans and Standards (OAQPS) is in charge of overseeing EPA programmes aimed at improving air quality in the following regions. The current qualities are not acceptable, & prevent contamination in areas where air is polluting more. To achieve this, OAQPS established NAAQS Every standard pollutant. There are two types Standard-Elementary and Middle School.

- Main criteria: they can prevent bad health effect;
- Secondary criteria: prevent welfare Impacts, such as injures the vegetables, injures to crops, injures the building for preventing welfare impacts.

Since other contamination has other effects, the standards of NAAQS are not same. Some particles have some standards differently for longer-term average time & shorter-term average time. The standards for shorter-term are developed to prevent shorter-term effects of health, while longer-term Established standards for the prevention of

chronic health effect. Since some other pollutants will have effects differently, the standards for NAAQS are different as well. These standards or regulations designed to prevent shorter-term impact on health while establishing longer-term regulations or standards are to protect from health effects such as chronic.

4.5 Air Quality Analysis

Today, air quality solutions or designs are gaining attention and getting more interest. With help of "data mining", we can design the air system with strong dynamics and strong spatial scalability. It is heterogeneous in behaviour. These models get data from various sources, such as satellites, public Institutions, sensors, etc. Advancement in satellite sensors provide new dataset for monitoring and shrivelling urban zones air pollution scales. Articles published in accordance with Chicago policy Comment [5], in contrast with traditional data dependent on data sets Sample or rough summary, "big data" is huge large size, fast speed, and various types. Since the year 2000, the amount of data exploded because of the speed at which technologies is being developed and implemented, Networking, information processing, & storage od data is also part of the digital infrastructure. Can generate big data from targeted, and voluntary resources. The difference between availability the required resolution. Key for doing big data the feasible method is to use, standardize and integrate massive data. Figure 1 we are discussing some important research Using data-driven research around the world the method of predicting air quality in the following paragraphs.
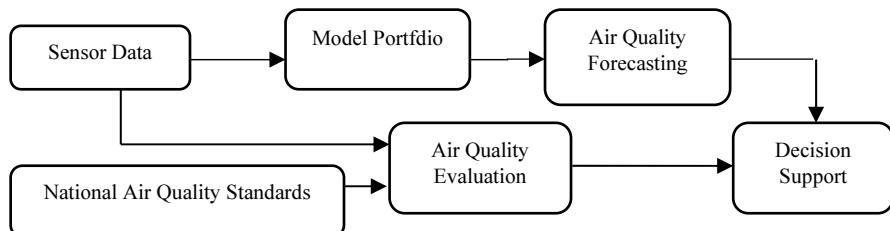
Figure 1. Block Diagram

The following are the steps to follow. Here the data will get imported first, after importing the data we have to train the model. Then accuracy functions have to be constructed. After that the model should be graphically represented. Evaluate the model which is previously represented graphically for getting the results. Input- Training the Data (D) to the model, Method Used-Here is Classification method and produce the result as a Decision tree. Initially, we extract the training data set (D). Let us assume I as input variable to model&T as the targeted variable. Figure 2 initially, the root nodes which have data of whole training set or may contain the data is constructed.
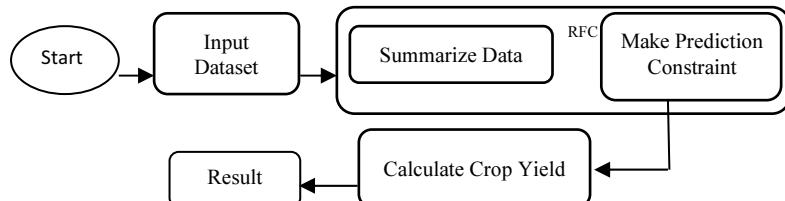
Figure 2. Schematic Diagram

Computing the Gini ratio for all the inputs I by the help of formula. Let us assume, the Gini ratio for total data set as $G(d)$ and the attribute for Gini ratio as $GA(d)$. Now the formula is:

$$G(d) = 1 + \sum_{i=0}^{m} pi^2$$

Now the Gini variable is calculated as:

$GA(d) = (d1/d)G(d1) + (d1/d)G(d2)$, where pi is referred as the relative frequency of class d. Let us assume X as splitting variable with min. index $GA(d)$ among I. Splitting the d variable w.r.t. X. Now create a child node for current node for every subset and we have to pass these instances to the node. We have to repeat everything until the $GA(d)$ is minimum. All the instances or values now belong to same class if the leaf or end node is less.

5. Conclusion

In this paper we find a proposed method of solution to standard air quality prediction techniques using big data &ML algorithms. we can monitor air quality and evaluate the future prediction. We can get real time quality of air and evaluate it with aid of traditional air prediction techniques, big data technology, and machine learning techniques. This paper reports on our latest literature review, which discusses and compares work quality assessment using big data analytics, ML approaches, models, and technology. Finally, some observations on future research issues, needs and challenges.

References

[1] Jerry ZeyuGao, GangXie, Sen Chiao, Gaganjot Kaur Kang, ShengqiangLu. Air Quality Prediction: Big Data and Machine Learning Approaches, 1 January 2018.

[2] J. Ditsela and T. Chiwewe, "Machine learning based estimation of Ozone using spatio-temporal data from air quality monitoring stations", IEEE, 2016.

[3] Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.

[4] P. S. Rao and V. M. Niharika, "A survey on air quality forecasting techniques", International Journal of Computer Science and Information Technologies, vol. 5, 2014.

[5] G. Vijay Kumar, M. Sreedevi, NVS Pavan Kumar "Mining Regular Patterns in Transactional Databases using Vertical Format" International Journal of Advanced Research in Computer Science, Volume (2), Issue (5) 2011.

[6] G. Vijay Kumar, V. Valli Kumari, "Sliding Window Technique to Mine Regular Frequent Patterns in Data Streams using Vertical Format", 2012 IEEE International Conference on Computational Intelligence and Computing Research, 2012.

Text Summarizing Using NLP

G. Vijay Kumar [1], Arvind Yadav, B. Vishnupriya, M. Naga Lahari, J. Smriti,
D. Samved Reddy
*Department of Computer Science &Engineering, Koneru Lakshmaiah Education
Foundation, Greenfields, Vaddeswaram, Guntur-522502, India*

Abstract. In this era everything is digitalized we can find a large amount of digital data for different purposes on the internet and relatively it's very hard to summarize this data manually. Automatic Text Summarization (ATS) is the subsequent big one that could simply summarize the source data and give us a short version that could preserve the content and the overall meaning. While the concept of ATS is started long back in 1950's, this field is still struggling to give the best and efficient summaries. ATS proceeds towards 2 methods, Extractive and Abstractive Summarization. The Extractive and Abstractive methods had a process to improve text summarization technique. Text Summarization is implemented with NLP due to packages and methods in Python. Different approaches are present for summarizing the text and having few algorithms with which we can implement it. Text Rank is what to extractive text summarization and it is an unsupervised learning. Text Rank algorithm also uses undirected graphs, weighted graphs. keyword extraction, sentence extraction. So, in this paper, a model is made to get better result in text summarization with Genism library in NLP. This method improves the overall meaning of the phrase and the person reading it can understand in a better way.

1. Introduction

The whole idea of automatic text summarization is to collect the necessary and crisp points from a large amount of data. There is a lot of information that is available on the internet and it also keeps growing every day and having to collect the main data from it becomes hard since it takes a lot of time. The use of automatic text summarization makes it easier for the users to collect the important data from huge information. Some of the graph base ranking algorithms are Text Rank [1], Hyperlinked Induced Topic Search [2], Positional Power Function [3] and so on. In this paper we are going to implement Text Rank algorithm. Noting down the important points manually from large amount of data can be a very stressful job. So, automatic text summarization takes out the crucial words and sends them back in a way that the readers find it easy. This, automatic text summarization is a small piece of NLP which cut downs the information and sends to the readers. It also arranges the information and sends back the sentences that are useful to create a crisp summary.

---

[1] G Vijay Kumar,Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation
Email: gvijay_73@kluniversity.in.

The words that occur the most no of times are considered the most worth. The top most words are also arranged and then a summary is created. The Extractive methodology chooses the principal significant lines from the information text and utilizes them to think of the outline. The abstractive methodology addresses the information text in a type then generates the outline with the desired output of words and sentences that disagree from the first text sentences. Extractive systems extract vital text units (such as sentences, paragraphs etc.) within the input document. The theoretic approach is practically identical to the way that human summarizers 1st perceives the most ideas of a document, so generate new sentences that aren't seen within the original document. The general design of an associate ATS system has the subsequent tasks: Pre-processing, Processing, Post-Processing. Text summarization is in this field as a conclusion that monitors are needed to grasp what humans have composed and generated human-readable outputs. human language technology also can be viewed as a study of computer science (AI). Therefore, several existing AI algorithms and strategies, as well as neural network models, are used for finding human language technology connected issues. With the present analysis, researchers typically believe 2 kinds of approaches for text summarization as shown in Figure 1 extractive summarization and abstractive summarization.
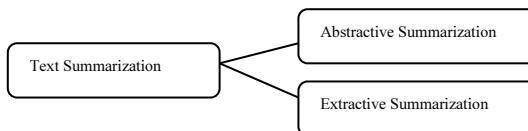
Figure 1. Types of Summarizations

2. Related work

Pratibha Devi Hosur et al [4] suggest the system by implementing unsupervised learning during automatic text summarization. This paper expresses the overall depiction of summarization of text using NLP which includes input text document, pre-processing, lesk algorithm and finally generating the summary. The lesk algorithm outputs, calculations, conclusion and proposed system. Narendra Andhale et al [5] This paper discussed about abstractive and extractive summarizations. They described how long texts are summarized in less time. and they focused on the extractive summarization methods. Deepali K Gaikwad et al [6] This paper expresses regarding how the necessary information is extracted from the long text document and forms the summary. Regular patterns [7, 8] be useful in Text summarization to extract useful keywords. They mainly discussed about abstractive method and extractive method and their approaches. Neelima Bhatia et al [9], in this study paper researched the famous and considerable effort done in the area of unit and numerous archive outlines. The creators examined the technique-based methodologies for summarization of text. These technique-based methodologies incorporate term-based recurrence strategy, diagram-based technique, time-based strategy, division and combining based strategy, semantic reliance strategy, theme-based methodologies, talk based methodologies, Latent Semantic based methodologies, approaches dependent on lexical chain, approaches dependent on fluffy rationale.

N. Moratanch et al [10] clarified about the strategy of summarization of text is that removed data is gotten as synopsis report and introduced as a small outline to the client. In the work creators examined about word level highlights, sentence level highlights and different extractive content synopsis techniques. The creators in this paper proposed a portion of the assessment measurements like human assessment, Rogue score, Recall, Precision, F-measure, Compression proportion. Shohreh Rad Rahimi et al [11] claimed that NLP explores are with more interest in summing up literary data. In this paper, creators characterized text synopsis as interaction of naturally making and lessening type of given report and holding its data content source into more limited variant with right importance. In this paper the creators likewise clarified about the connection between text mining and text outline. At last, this paper examines about different ways to deal with text outline, for example, Statistical methodologies, Lexical Chain based methodologies, Cluster based methodologies, Fuzzy rationale-based methodologies.

3. Problem Definition

In our busy schedule, it is very difficult for us to go through the entire article or document. So, we prefer to read summary. In this paper we are going to summarize the large text in to a short summary which reduces reading time for users.

4. Methodology

NLP is a part of Artificial Intelligence reasoning that manages analyzing, understanding, finding and producing the dialects that people use in a characteristic manner to make interface with PCs in both composed and spoken settings utilizing common human dialects rather than codes.

4.1. TextRank Algorithm

Text rank algorithm is a diagram-based positioning model for text processing which can be used in order to find the most applicable sentences in text and also to find keywords. Text rank algorithm is similar to page rank algorithm. Page rank algorithm is used to mark Webpages in web search conclusions and in web usage mining. In text rank algorithm, in position of Webpages sentences are taken.

1. Identify content units that best characterize the current task, and add them as vertices in the diagram.
2. Identify the relations that append the content units, and in the chart utilizing these relations draw edges between vertices. Edges can be un-weighted or weighted and undirected or coordinated.
3. And at that point loop the diagram-based positioning algorithm until union.
4. Based on their last score mastermind the vertices. For positioning and determination choices Use the qualities appended to every vertex.
5. At last, the highest-level sentences will shape a synopsis.

5. Proposed System Results

As shown in Figure 2, Source document is the input text given. Preprocessing: Tokenization is the technique used to split the text in to tokens (words or paragraphs or sentences). Stop words is used to reduce the size of text, we have a dictionary in pre-processing which is made up of stop words. It compares the words in given text and then remove the matched words. Hence removal of stop words will increase the performance. feature extraction: word frequency means most common word that occur
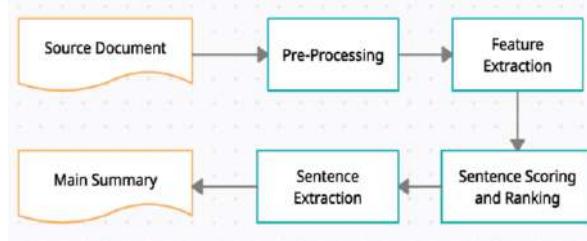
Figure 2. Flow Chart

in text are measure of information. It is determined as the quantity of occurrence of word by complete number of words in the archive. Too long or too short sentences are eliminated utilizing length of sentences. It is determined as number of words in the sentence by number of words in longest sentence. Sentence scoring and ranking: it calculates the score for each sentence and rank them. Sentence extraction: The main target of this is to identify best in the text. The target of this is to rank complete sentences. Main summary: place the sentences in order and generate the resultant summary.

5.1. TextRank Model

Graph based algorithms is the most required strategy of determining the powerful of a vertex in the actual graph, elicited from over all information gathered from the entire graph. The fundamental idea we have implemented here is voting and recommendation. Based on the votes casted, the score is related with vertex. We implement "random surfer model" as the probability that skip from one vertex to some other vertex. The score of a graph, starting from arbitrary values and the computation iterates. The score of a vertex is based on the importance of vertex and the last qualities are not affected by beginning qualities. Figure 3, 4, 5, 6 shows the results for TextRank algorithm and for single document we use textrank and for multi-document we use lexrank.

5.2. UnDirected graphs
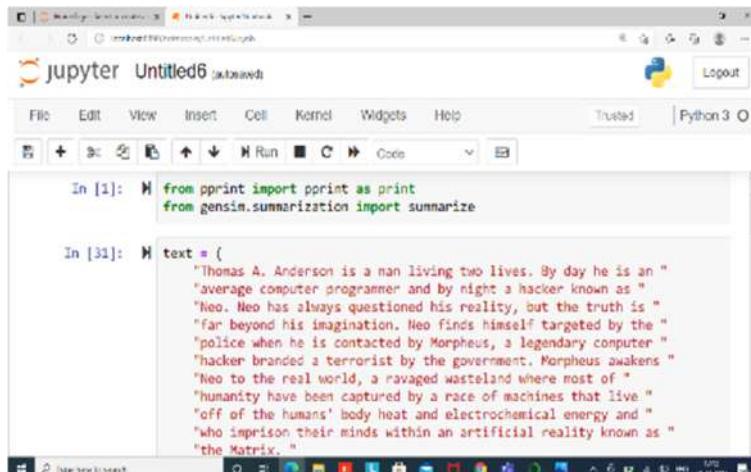
Basically, we apply recursive graph-based ranking algorithm on directed graphs, as the out-degree and in-degree is equal it is also enforced for un directed graph In convergence curve as the connectivity of the graph increases then fewer iterations take place and the convergence curve for in-directed or directed graph practically overlap.

5.3. Weighted graphs

As the main definition of PageRank for graph based learning algorithm is we have to assume un weighted graph and as the graphs are constructed from nlp, textrank may include multiple or partial link between units. Based on the weight of edge textrank incorporate the power of connectivity which we can see in Figure 7.

$$WS(V_i) = (1 - d) + d * + \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{Vk \in Out(V_j=0)} w_{jk}} WS(V_j)$$

When we compute the score related with the vertex in a graph then the latest anon takes in to account edge weights. The final vertex ranking and scores differ as compared to original measure and the number of iterations is nearly same for unweighted or weighted graphs.

Figure 3. Text Paragraph

Figure 4. Summarizing Text Paragraph
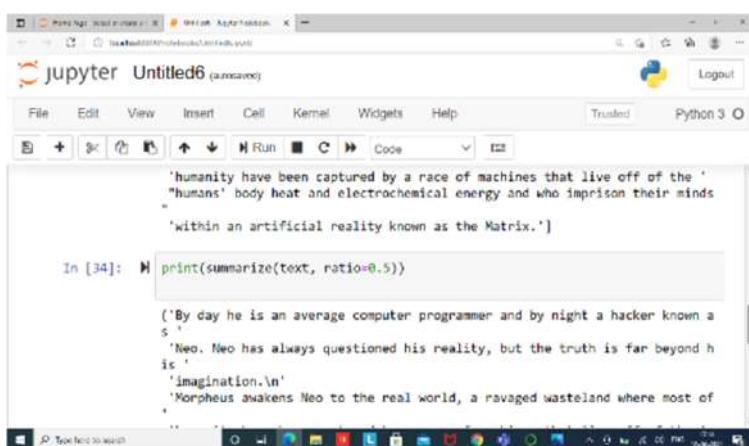
Figure 5. Summarizing Text with word count = 50

Figure 6. Summarizing Text with ratio = 0.5
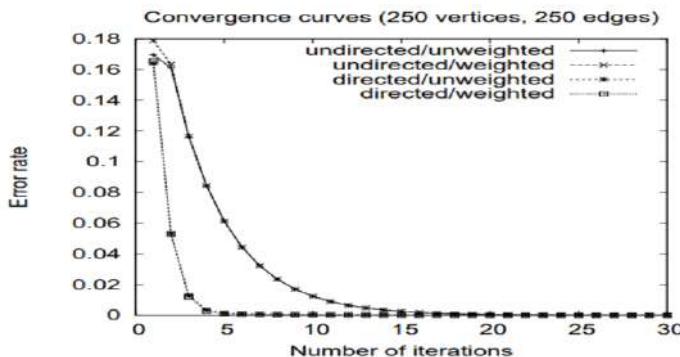
Figure 7. Convergence Curves for weighted graphs

In this, the module automatically summarizes the given input text and it finally it picks up the important sentences. It can also extract keywords as shown in execution.

6. Conclusion

The paper demonstrates that we use advanced techniques to apply on the document for text summarization using extractive summarization method called TextRank algorithm. At first, we loaded necessary libraries and related function in python and then code was implemented to summarize the text. Afterwards, a model is proposed with slight expansions to improve by showing the outline text. The techniques that are presented in this paper to get better result in text summarization with Genism library in NLP. With this the overall meaning of the document can be understand easily.

References

[1]   Hans Peter Luhn, "The automatic creation of literature abstracts", IBM Journal.
[2]   Kleinberg J. M., "Authoritative sources in a hyperlinked environment". Journal of the ACM, Volume 46 issue 5, pp.604–632, Sep 1999.
[3]   Herings, G. van der Laan, and D. Talman, "Measuring the power of nodes in digraphs", Technical report, Tinbergen Institute, 2001.
[4]   Pratibha Devihosur, Naseer R. "Automatic Text Summarization Using Natural Language Processing" International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 08, Aug-2017
[5]   Deepali K. Gaikwad, C. Namrata Mahender, "A Review Paper on Text Summarization", "International Journal of Advanced Research in Computer and Communication Engineering". Vol.5, Issue 3, March 2016.
[6]   G. Vijay Kumar, M. Sreedevi, NVS Pavan Kumar, "Mining Regular Patterns in Transactional Databases using Vertical Format"."International Journal of Advanced Research in Computer Science", Volume 2, Issue 5, 2011.
[7]   G. Vijay Kumar and V. Valli Kumari, "Sliding Window Technique to Mine Regular Frequent Patterns in Data Streams using Vertical Format", IEEE International Conference on Computational Intelligence and Computing Research, 2012.
[8]   Neelima Bhatia and Arunima Jaiswal, "Automatic Text Summarization and its Methods-AReview", 6th International Conference. Cloud System and Big Data Engineering, 2016.
[9]   Potharaju, S. P., & Sreedevi, M. (2017). A Novel Clustering Based Candidate Feature Selection Framework Using Correlation Coefficient for Improving Classification Performance. Journal of Engineering Science & Technology Review, 10(6).

[10] Shohreh Rad Rahimi, Ali Toofan zadeh Mozhdehi and Mohamad Abdolahi, "An Overview on Extractive Text Summarization"."IEEE 4th International Conference on Knowledge Based Engineering and Innovation" (KBEI) Dec. 22nd, 2017, Iran University of Science and Technology – Tehran, Iran.

Hybrid Parallel Feature Subset Selection for High Dimensional Datasets

Archana Shivdas Sumant [a, 1], Dr. Dipak V. Patil [b]

[a] *MET's Institue of Engineering Bhujbal Knowledge City, Nashik, India*
[b] *Gokhale Education Society's R.H. Sapat College of Engineering Management Studies and Research Nashik 422005, India*

Abstract. High dimensional data analytics is emerging research field in this digital world. The gene expression microarray data, remote sensor data, medical data, image, video data are some of the examples of high dimensional data. Feature subset selection is challenging task for such data. To achieve diversity and accuracy with high dimensional data is important aspect of this research. To reduce time complexity parallel stepwise feature subset selection approach is adopted for feature subset selection in this paper. Our aim is to reduce time complexity and enhancing the classification accuracy with minimum number of selected feature subset. With this approach 88.18% average accuracy is achieved.

Keywords. High dimensional data, parallel feature subset selection, stepwise selection, symmetric uncertainty, chi-squared.

1. Introduction

Data analytics on high dimensional data is a challenging task. As high dimensional data contains big number of features compared to number of samples in the datasets. If number of features as p and number of samples as n, then p>>n is the high dimensional data. All features are not equally important for extracting meaningful information from such dataset. It increases the time and space complexity as these data contains many redundant and irrelevant features. To avoid this problem ranking methods are used before applying algorithm. But ranking methods have disadvantage that does not considers feature dependency. Search methods plays important role here which selects optimal feature subset by considering feature dependencies.

The way toward distinguishing and evacuating unessential and excess features is known as feature subset selection (FSS). FSS boosts the algorithm to operate much quicker and accurate by reducing the dimensionality of data. FSS in other words known as variable selection, attribute selection or variable subset selection. Feature selection gives many advantages as it enhances expectation execution, understand-ability, versatility, and speculation capacity of the classifiers. It additionally diminishes computational complexity and storage, provides faster and more commercial model.

High Dimensional Data (HDD) poses different challenges on predictive algorithms. Let's say we have *n* samples and *p* features.

---

[1] Archana Shivdas Sumant, MET's Institue of Engineering Bhujbal Knowledge City, Nashik
Email:archana.s.vaidya@gmail.com.

Here features are attributes, independent variables, explanatory variables. High dimensional data is data having n<<p and p are usually high in thousands or ten thousand. So dealing with these numbers of dimensions with high predictive accuracy is the challenge in these coming data era with lots of data generated is of high dimension. Two solutions are their one is dimensionality reduction and the second one is selecting a subset of features. There are several search strategies of different types, but no best algorithm for feature selection is found in general. Prior art [1] compare FS algorithms and conclude that there is not a single approach that outperforms all the others for all datasets. Therefore, it is necessary to continue providing the community with new feature selection alternatives as well as strategies to enhance the performance of the current ones.

Search Strategy in FSS contribute to reduce the time complexity and also to increase the accuracy. In this paper proposed hybrid parallel approach for high dimensional data is implemented.

2. Overview of Feature Subset Selection methods

Feature subset selection problem is stated as given the input data as N samples and M features. The objective of feature selection is to find a subspace of features from the M-dimensional observation space to reduced feature space X, that could be optimally separated the c classes.
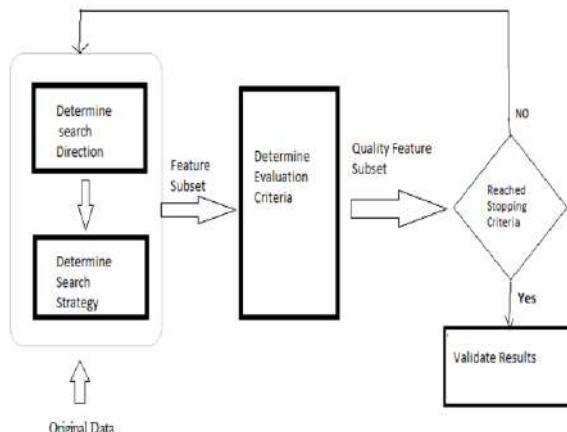
Figure 1. General process of feature subset selection

Figure 1 shows the process of feature subset selection includes search direction, search strategy and evaluation criteria. Feature selection aims to select a feature subset from the original set of variables from its relevance and redundancy.

The system in [2] classify the features into four categories: (i) completely inapt and noisy features, (ii) poorly relevant and surplus features, (iii) weakly relevant and non-redundant features, and (iv) powerfully relevant features. The best subset primarily contains every one of the features from the group (iii) and (iv).

Several main approaches for feature selection are distinguished in [3] as filter, wrapper, Hybrid and embedded methods. In recent years, new techniques are emerging, i.e., ensemble feature selection [4] and deep learning-based feature selection [5][6].

3. Proposed Parallel FSS Approach

Search strategies searches in feature space and selects relevant features by evaluating their performance. The proposed parallel approach is explained in algorithm parallel FSS using stepwise search for high dimensional data. In High dimensional dataset number of features are very large so in the first step features are ranked based on symmetric uncertainty and Chi- square as per equation 1 and 2. Features whose score is zero for these two measures are eliminated and rest of the features are selected for further processing. Symmetric uncertainty (SU) is normalized value measure of Mutual Information (MI) [7] calculated as in eq. (1)

$$SU(X,Y) = \frac{2*MI(X,Y)}{H(x)+H(y)} \qquad (1)$$

Value 1 indicates strong relevance between X and Y, while 0 indicates X and Y are independent. SU measure is symmetric in nature therefore SU(X, Y) is same as that of SU(Y, X).

Chi-squared attribute evaluation evaluates the worth of a feature by computing the value of the chi-squared statistic with respect to the class. The initial hypothesis H_0 is the assumption that the two features are unrelated, and it is tested by chi squared formula:

$$\chi 2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij} \right)^2}{E_{ij}} \qquad (2)$$

Eq. (2) gives chi-square score for each feature. O_{ij} is observed value E_{ij} is expected value. It determines significant relationship between two nominal feature vectors. In this feature to predictive class relationship is tested and score is calculated. A Stepwise selection [8] is a combination of forward and backward selections. It starts with no predictors, and then sequentially adds the most contributive predictors like forward selection. After adding each new variable, remove any variables that no longer provide an improvement in the model fit like backward selection.

4. Results Discussion and Implementation Details

The proposed approach is implemented with R 3.6.0 and validation with classifiers was done with Python 3.7.2. System configuration used here was an Intel i7 processor with 8 GB RAM. R language is used for feature selection and Python is used to validate results. Table 1 gives dataset details used in experiment. All datasets are downloaded from [9]. Here n is number of instances, p is number of features and C_k is number of classes. The KalR [10] package from R is used for implementation. Accuracy is used to measure system performance. To calculate accuracy of classifier cross fold validation technique is used. It divides dataset into train and test dataset. Here tenfold cross validation is used. The working of the system is explained with an algorithm 1.

$$\%Accuracy = \left(\frac{\text{Total number of correclty classified instances}}{\text{Total instances}} \right) * 100 \qquad (3)$$

Accuracy is a measure of correctly classified instances in total number of instances. Equation (3) gives accuracy in percentage.

Algorithm 1: Parallel FSS using stepwise search for high dimensional data

Input: Dataset with (number of samples n and number of features p)

Output: Validated results with RF, SVM, KNN classifiers.

Start:

1 :   **for** features i   1 to p **do**

2 :   [scores]   Calculate(Symmetric Uncertainty) with Equation (1)
        [scores]   Calculate(Chi-Squared score) with Equation (2)

3 :   **end for**

4 :   q ⟵ be the number of features with positive score from earlier stage

5 :   Input select ranked subset with top scoring q features

6 :   **for** features i   1 to q **do**

7 :   for subset j 1 to m

8 :   selecting feature i and putting it in subset j

9 :   end for

10   **end for**

11   **do** parallel for each subset stepwise feature selection

12   Top selected features from each subset are combined to form final feature subset

13   **for** each classifier calculate classification Accuracy based on final selected feature subset

14   Random Forest(RF)

15   Support Vector Machine (SVM)

16   K- Nearest Neighbor (KNN)

17   **end for**

18   Return Accuracy for feature subset

19 **End**

Table 1. Dataset details used in experiment

DN	Dataset Name	n	p	Ck
1	COLON	62	2000	2
2	Lung-cancer	203	12600	5
3	Ovarian	254	15154	2
4	CNS	60	7130	2
5	Leukemia	72	7129	2
6	Prostate	102	12600	2
7	DLBCL	47	4026	2

The description of the classifiers used is as follows.
Random Forest (RF) is an ensemble approach based on decision trees. Support Vector Machine (SVM) generates hyperplanes to separate samples belonging to different classes. For multi-class problems it converts problem as one versus rest i.e. dividing problem into multiple binary class problems. Here linear kernel is used for generating hyperplanes. K Nearest Neighbor (KNN) works based on proximity of test samples with neighbors instances. It is a supervised learning algorithm and works with neighbor samples. Here neighbor count is set to 5 and brute-force search algorithm is used.

Table 2. Proposed parallel stepwise search accuracy measured on high dimensional datasets

DN	1	2	3	4	5	6	7	Classifier wise Average
Dataset Name	COLON	Lung-cancer	Ovarian	CNS	Leukemia	PROSTATE	DLBCL	
SVM	94	88	99	74	90	88	90	89
RF	93	99	99	43	92	90	75	84.42
KNN	94	99	99	74	96	87	89	91.14
Average of All three Classifiers	93	95.33	99	63.66	92.66	88.33	84.66	**88.18**

Table 2 states accuracy measured with proposed method with parallel stepwise search. The highest performance has been observed on ovarian dataset as 99%. And the lowest performance has been observed for CNS dataset. The KNN classifier achieves the highest average accuracy as 91.14% while RF classifier achieves lowest average accuracy as 84.42 %. The average accuracy achieved is 88.18%. The time complexity analysis shows parallel approach is three times faster than sequential approach.

5. Conclusion

Feature subset selection is NP -hard problem and not a single solution generalizes the system for classification. So, there is need of solution to improve performance of existing system. Feature subset selection plays an important role in case of high dimensional dataset. The proposed parallel stepwise feature selection achieves average accuracy as 88.18%. In future ensemble approach can be adopted to increase system accuracy.

References

[1]. Guyon, A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.* **3**, 1157–1182 (2003)

[2]. Geem Z.W., " State-of-the-Art in the Structure of Harmony Search Algorithm," In: Geem Z.W. (eds) Recent Advances In Harmony Search Algorithm. Studies in Computational Intelligence, vol 270 2010, https://doi.org/10.1007/978-3-642-04317-8_1.

[3]. K. S. Lee and Z. W. Geem, A new meta-heuristic algorithm for continuous engineering optimization: Harmony search theory and practice, *Comput. Meth. Appl. Mech. Eng.*, vol. 194, nos. 36–38, pp. 3902–3933, Sep. 2005

[4]. M. B. Dowlatshahi, V. Derhami, and H. Nezamabadi-Pour, "Ensemble of filter-based rankers to guide an epsilon-greedy swarm optimizer for high-dimensional feature subset selection," Inf., vol. 8, no. 4, 2017, doi: 10.3390/info8040152

[5]. Yawen Xiao a , Jun Wu b , Zongli Lin c , Xiaodong Zhao b "A deep learning-based multi-model ensemble method for cancer prediction ",dx.doi.org/10.1016/j.cmpb.2017.09.005

[6]. Van-Sang Ha , Ha-Nam Nguyen, "Credit scoring with a feature selection approach based deep learning," DOI:10.1051 2016.

[7]. Hoque, N., Singh, M., & Bhattacharyya, D. K. (2017). EFS-MI: an ensemble feature selection method for classification. *Complex & Intelligent Systems*. https://doi.org/10.1007/s40747-017-0060-x

[8]. Geem Z.W. (2010) State-of-the-Art in the Structure of Harmony Search Algorithm. In: Geem Z.W. (eds) Recent Advances In Harmony Search Algorithm. Studies in Computational Intelligence, vol 270. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04317-8_1

[9]. https://archive.ics.uci.edu/ml/index.php

[10]. https://rdrr.io/rforge/klaR/man/stepclass.html

A Compact Disc Shaped Microstrip Patch Antenna Using Inset Fed at 5GHz for Satellite Communications

A. Pon Bharathi [a,1], Allan J Wilson [a], S. Arun [a] and V. Ramanathan [a]

[a]*Dept of ECE, Amrita College of Engineering and Technology, Nagercoil, India*

Abstract. This examination work is focused around planning and simulating another kind of inset feed Disc Shaped Microstrip Patch Antenna (DSMPA) with Inset feed and Defected ground plane (DGP). By presenting a round space at the focal point of the ground plane, improved attributes of Microstrip patch antenna can be accomplished. The proposed Disc Shaped Microstrip patch antenna is reverberating at 5 GHz. Simulation has been finished by utilizing reenactment programming HFSS version15. From recreation results, it discovers that our examined Disc Shaped Microstrip patch antenna yields better return loss of - 25.1 dB & VSWR estimation of 0.96 dB. The examined DSMPA is yielding a higher radiation efficiency of 77.20 %. The minimized size and higher radiation efficiency contrasted with rectangular Microstrip patch antenna makes it all the more generally helpful for satellite communications.

1. Introduction

Microstrip patch antenna apparatus has significant applications particularly in the field of versatile deices, military, medical, business applications and remote interchanges. Their usage has become different in view of their minimal size and light weight. An Antenna is normally a metallic segment utilized for transmission or accepting electromagnetic waves. As remote applications require increasingly more transmission capacity, the interest for wideband antennas working at higher frequencies gets inescapable. Quick and practical manufacture is particularly significant with regard to the prototyping of radio wires for their exhibition assessment. Naturally Microstrip patch antennas have slender transmission capacity and low productivity and their presentation significantly relies upon the substrate boundaries. The main aim of the proposal is designing a Disc Shaped Microstrip Patch antenna because we need to find only one parameter (radius) for designing the Disc Shaped Microstrip Patch antenna.

---

[1] A. Pon Bharathi, Assistant Professor, Dept of ECE, Amrita College of Engineering and Technology, Nagercoil, India
Email: bharathpon@gmail.com.

[3] For high velocity remote neighborhood and other remote correspondence frameworks covering 5.15–5.825 GHz recurrence band, wideband E-molded microstrip patch antenna has been planned. To irritate the surface current way two equal openings are joined, neighborhood inductive impact presented is answerable for excitation of another frequency. To tune about recurrence of second resounding mode without influencing the crucial full rage, length of middle arm is managed. Reflection Coefficient estimated underneath −10 dB about 5.05 GHz to 5.88 GHz. Presentation excess than satisfying interest transfer speed determination to cover the high recurrence. [4] In which reception antenna utilizes corporate feed strategy using inset feed input reverberating at 5.216 GHZ. FR−4 material is utilized as a dielectric substrate for the proposed cluster structure. The proposed reception antenna has been intended for the reach 5.15 − 5.35 GHz. According to IEEE 802.11 standard 5.15GHz-5.35 GHz reach can be considered as one of the reaches for 5 GHz band remote neighborhood (WLAN). The most extreme output yield and direction of radiation about the proposed exhibit are 9.019 dB & 12.81 dB separately and the cross polarization is low. The introduced cluster array is appropriate for 5 GHz. [5] An inset microstrip patch antenna is planned, recreated and fabricated at recurrence 5 GHz. To plan the proposed microstrip patch antenna, substrate GML 1000 (lossy) is taken with dielectric constant value of 3.2 and tangent loss 0.002. The components of the patch antenna are 20.07 mm (width) and 16.56 mm (length). In this fabricated on Microstrip patch antenna the impedance is 50 Ω's. Ground plane soldered with external conductor using co-pivotal and afterward the middle conductors are checked not to have cut off the ground plane. [6] Creator proposed triband minimized printed microstrip patch antenna that can be used for WiMAX& Wireless LAN applications. Microstrip patch antenna design comprises of collapsed open stub, asymmetric trapezoid ground structure plane, long and short L-formed strips. The construction planned is operable at 3 different frequency ranges (2.4 GHz, 5 GHz Wireless LAN and 3.5/5.5 GHz Wireless MAX groups). The imploded open stub, long and short L-shaped strips recognize impedance organizing at 2.4, 3.5 and 5.2/5.8 GHz independently, and the lopsided trapezoid ground plane changes impedance planning at 5.2, 5.5 and 5.8 GHz. [7] Plan of Microstrip patch antenna as model of variety cluster as 4X4 is intended for 5 GHz band Industrial, Scientific & Medical (ISM Bands) and Wireless LAN application. Single component strip is intended for desired particular & later 16 component strip is intended for frame exhibit and a force divider to fabricate the reception apparatus antenna with the referenced determinations. In this paper, various procedures used to upgrade acquire have been audited. Investigation of size decrease strategies is all around considered. By utilizing a 16-component patch cluster exhibit at 5GHz band, addition is expanded trade off in beam width. The acreage of reception antenna considerably diminished utilizing TLC 30 overlays with compact substrate misfortunes at 5GHz. [8] A couple of rearranged T-formed cuts and an inset feed care of circular Shaped molded microstrip patch antenna for remote correspondences is introduced. The proposed reception antenna is reverberating at four unique frequencies (2.33 GHz, 4.18 GHz, 5.04 GHz and 6.49 GHz) accompanied by impedance data transfer capacities of 2.7%, 2%, 2.1% and 2.6%. This reception microstrip patch antenna enfolds Wireless LAN (2.4 - 2.48/4.97 - 5.12 GHz), X-band lower satellite correspondence (6.42 - 6.59 GHz) usage and shows a main side radiation trademark at wanted recurrence bands. [11] Applications of DGS structure and its effects in ground plane.

2. Design Methodology of the proposed DSMPA

Figure 1 shows the top perspective on the proposed Disc Shaped Microstrip patch antenna with intercalate feeding technique. Figure 2 shows about base perspective on the examined Disc Shaped Microstrip patch antenna with DGP structure.
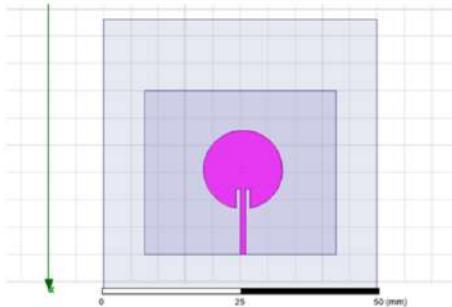
Figure 1. Top view of the proposed DSMPA
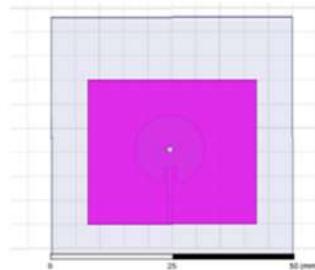
Figure 2. modelBottom view of the presented DSMPA model

Table1. Parameters of the Presented Disc Shaped Microstrip Patch Antenna

Substrate Length	30 millimetre
Substrate Width	35 millimetre
Substrate Height	3.6 millimetre
Patch Radius	7.2 millimetre
Feed Length	12.2 millimetre
Feed Width	1 millimetre
Substrate Dielectric Constant (Ɛr)	4.4

Design Equations with respect to Presented Disc Shaped Microstrip Patch Antenna:

$$p = \frac{F}{\left\{1+\frac{2h}{\pi \varepsilon_r F}\left[ln\left(\frac{\pi F}{2h}\right)+1.7726\right]\right\}^{1/2}} \text{ ------------------------------------- (1)}$$

$$F = \frac{8.791\times10^9}{f_r\sqrt{\varepsilon_r}} \text{ ------------------------------------- (2)}$$

Where, p = Radius of the proposed disc shaped microstrip patch antenna

Ɛr = Substrate Dielectric Constant

h = Height of the substrate material (cm)

f_r = Resonance Frequency

F = Frequency of operation

3. Simulation Results & Discussion

Figure 3 addresses simulated Return Loss of the proposed CMPA model and it is resonating at 5 GHz. Return Loss concerning the proposed Disc Shaped Microstrip Patch Antenna is - 25.1 dB. Figure 4 addresses Voltage Standing Wave Ratio (VSWR) with respect to the presented DSMPA and its worth is 0.96 dB. Figure 5 addresses radiation pattern concerning the presented Disc Shaped Microstrip Patch Antenna has high attainment estimation of 3.76 dB. Figure 6 description about 3D Polar Plot of examined Disc Shaped Microstrip Patch Antenna. Table 1 shows Parameters of the Presented Disc Shaped Microstrip Patch Antenna.
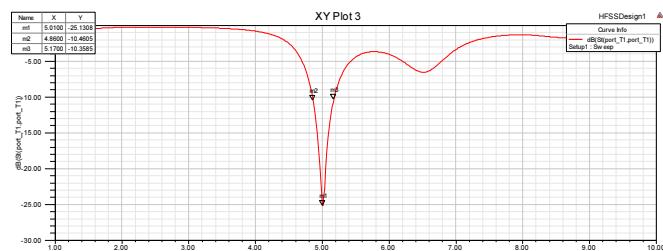
Figure 3. Proposed DSMPA model - Return Loss
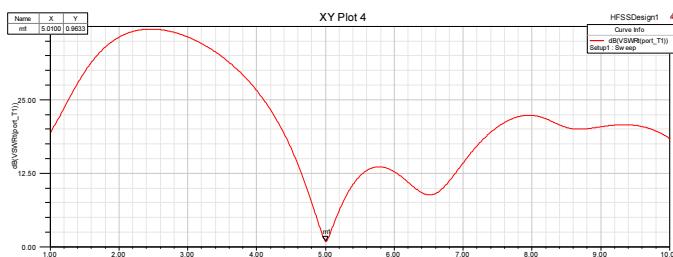
Figure 4. Proposed DSMPA model - VSWR
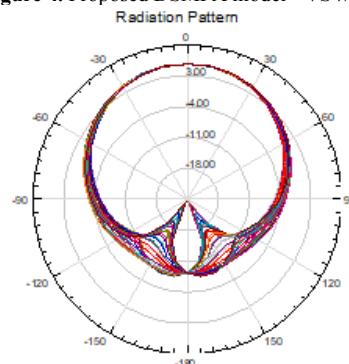
Figure 5. Radiation Pattern of the Proposed CMPA model

Figure 6. 3D Polar plot of the Proposed CMPA model

The Figure 7 shows the simulated results like Front to Back Ratio, Peak gain, Peak directivity, Radiated power & Radiation efficiency with respect to the proposed Disc Shaped Microstrip Patch Antenna.

Antenna Parameters:

Quantity	Freq	Value
Max U	5GHz	0.0029846 W/sr
Peak Directivity		4.8787
Peak Gain		3.7668
Peak Realized Gain		3.7507
Radiated Power		0.007688 W
Accepted Power		0.0099573 W
Incident Power		0.01 W
Radiation Efficiency		0.77209
Front to Back Ratio		25.332
Decay Factor		0

Figure 7. Proposed DSMPA model - Simulated characteristics

4. Conclusion

A high radiation efficiency Disc Shaped Microstrip Patch Antenna has been planned simulated & yield qualities are estimated. Radiation efficiency about Disc Shaped Microstrip Patch Antenna is improving make usage of DGS hole. Antenna designed is simulated and the outcomes are introduced. The introduced Disc Shaped Microstrip patch antenna radiates at recurrence range around 5 GHz. Greatest increase gain & most extreme directivity of the proposed Disc Shaped Microstrip patch antenna are 3.76 dB and 4.87 dB separately. The proposed Disc Shaped Microstrip Patch Antenna is reasonable around 5 GHz satellite applications.

References

[1]. Ayyappan, Manoj. B, Jagadish Chandran, Design and Analysis of Disc Shaped Microstrip Antenna at 5.8 GHz with Fr-4 Substrate. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. March 2016; Vol. 5, Special Issue 4.

[2]. Zahid Hasan, Ashiq Zaman, Dr. Anis Ahmed, Design and Fabrication of a Disc Shaped Microstrip Patch Antenna for GPS Application. IJECT. July - Sept 2017; Vol. 8, Issue 3, ISSN : 2230-7109 (Online) | ISSN : 2230-9543 (Print).

[3]. B.-K. Ang and B.-K. Chung A, Wideband E-Shaped Microstrip Patch Antenna for 5–6 GHz Wireless Communications. Progress in Electromagnetics Research, PIER 75, 397–407, 2007.

[4]. B. W. Ngobese, P. Kumar, A High Gain Microstrip Patch Array for 5 GHz WLAN Applications. Advanced Electromagnetics, August 2018; Vol. 7, No. 3.

[5]. Priyanka Mishra, Design Simulation and Fabrication of Inset Microstrip Patch Antenna At 5 GHz. JETIR, July 2018; Volume 5, Issue 7, www.jetir.org (ISSN-2349-5162).

[6]. Pracha Osklang, Chuwong Phongcharoenpanich, and Prayoot Akkaraekthalin, Compact Printed Antenna for 2.4/3.5/5 GHz WLAN/WiMAX Applications. International Journal of Antennas and Propagation.

[7]. Mr. Chandrahasa, R.Salian, Mr. Santhosh, Mr. Sandeep Vedagarbham, A Review on Design and Development of Low Cost 5GHz Microstrip Patch Array Antenna. International Conference on Innovative Mechanisms for Industry Applications. IEEE (ICIMIA 2017); 978-1-5090-5960-7/17.

[8]. Ashwini C. Tengli1 and P. M. Hadalgi, Design of Inset-fed Quad-band Disc Shaped Microstrip Antenna.for 2.4/5 GHz WLAN, WiMaX and X-band Satellite Applications. Emerging Technologies: Micro to Nano (ETMN-2017) AIP Conf. Proc. 1989; 030021-1–030021-5; Published by AIP Publishing. 978-0-7354-1705-2.

[9]. C. E. Balanis. Antenna Theory: Analysis and Design, 3rd Edition Constantine A. Balanis, Book. 2005; https://www.miwv.com/5g-radio-frequency.

[10]. http://www.antennatheory.com/antennas/patches/antenna.php.

[11]. A. Pon Bharathi, Dr. P. Kannan, S. Maheswari, Dr. S. Veluchamy, A Compact Microstrip Patch Antenna using DGS for 5G Applications. International Journal of Emerging Trends in Engineering Research, April 2021; Vol 9. No.4. ISSN 2347-3983.

An Automated Music Recommendation System Based on Listener Preferences

Mukkamala. S.N.V. Jitendra [a, 1], Y. Radhika [a]

[a]*Department of Computer Science and Engineering, GIT, GITAM (Deemed-to-be University), Visakhapatnam-530045, AP, India*

Abstract. Recommender systems play a vital role in e-commerce. It is a big source of a market that brings people from all over the world to a single place. It has become easy to access and reach the market while sitting anywhere. Recommender systems do a major role in the commerce mobility go smoothly easily as it is a software tool that helps in showing or recommending items based on user's preferences by analyzing their taste. In this paper, we make a recommender system that would be specifically for music applications. Different people listen to different types of music, so we make note of their taste in music and suggest to them the next song based on their previous choice. This is achieved by using a popularity algorithm, classification, and collaborative filtering. Finally, we make a comparison of the built system for its effectiveness with different evaluation metrics.

Keywords. E-commerce, recommender system, collaborative filtering, user-based recommendation, item-based recommendation, music information retrieval.

1. Introduction

On the web, where the measure of decisions is overpowering, there's a prerequisite to filter, organize and proficiently convey significant data to lighten the matter of data overload, which has made a potential issue for a few Internet users. Recommendation settles the issue of giving the best outcomes by giving the results of top information which is dynamically generated using the required methods and data.

Moreover, of late, interpersonal organizations turned out to be generally utilized and well-known modes for information dissemination likewise on account of the facilitators of social interaction. User commitment and exercises give important understanding into singular conduct, encounters, assessments, and interests. Considering the metadata together with the client information gives more approaches to build the exhibition of strategies like shared separating. Recommendation systems have been introduced into a variety of areas with challenges. These include Government recommendation systems, Government service recommendation, E-business recommendation systems, A Telecom recommendation system, Commerce/shopping recommendation systems, Library recommendation systems, Learning recommendation systems, Tourism recommendation systems, Service recommendation systems [1].

---

[1] Mukkamala.S.N.V. Jitendra, Department of Computer Science and Engineering, GITAM, India

Email: jitendra.mukkamala@gmail.com.

Various models used models for music recommendation can including music recommendation models such as content-based, emotional, which is used for people; collaborative, which uses information in addition to the content; and additional content; and metadata-based, which includes information about music and genres [2]. The collaborative recommendation has served well but has weaknesses such as being prone to popularity bias, the limitations of human actions, etc. Concerning the other disadvantages, the hybrid system has not yet been studied extensively, but on the other hand, it does offer greater performance concerning the breadth of opinion. Both the emotional and social information are considered models of the recommendation that have a great impact on the overall quality of the recommendations This topic is just beginning to be researched and will continue to expand in the coming years.

Today almost every e-commerce website or online streaming site uses the Recommender System to make it easier for the users to access and profitable on both ends. The sites like Netflix, Amazon, Flipkart, etc. generally use these to maintain their users by providing suggestions that will be mostly liked by them. The Recommenders use the user's search history and other statistics to recommend the products. A famous example of such a recommendation is the one with amazon. It suggests the products similar to the product being viewed and also the products which the other customers bought and viewed the same product. Similarly, Netflix does the recommendations based on the type of movie watched, language, cast, theme, or genre. Some online music streaming websites offer recommendations to meet the user's demand, for example, Spotify, Pandora, iTunes, etc [3].

2. Literature Survey

The Internet is a source of infinite information with rapid growth. There are millions of e-commerce websites on the internet and so are the products available, which leads customers not to make the right decisions. The different components of recommendation systems are items, users and user-item matching algorithms, various approaches of recommendation systems.

The introduction of recommendation systems has proved itself to be priceless and is appreciated because of their ability to make customers make the right choices in time. Recommender Systems are software tools that use techniques for providing product suggestions to a user. The hints identify with different decision-making processes, for example, what things to purchase, what music to tune in to, or what online news to peruse. The recommendation system carries out its process by analyzing the user's visit to its website and remembering their choice. This analysis is further used to suggest the user [4].

Recommendations are classified into two types based on the number of users the system suggests to. When the interest of a single user is noted to provide him the suggestion, it is known as a personal recommender system. Since a personal taste cannot be the same for different users, this type of recommenders is used based on the user's taste or preference. Another type of recommender system is called public recommendations or impersonal recommendations. When a system grabs the interests of users on a large scale, basically to make recommendations based on the popularity of the item [5].

By the definition of the recommendation system, it is clear that there would be something that captures the user's behavior, and analyzing that behavior would result in future predictions. The thing that provides this capability to a system is said to be "user modeling" or "user profile". User profile or user modeling is the basic unit of every kind of recommendation system. The framework stores data about the user's behavior into the user's profile. The data is about the user's most continuous visits, top quests, and so forth. A famous e-commerce website, Amazon also uses a recommender system to provide suggestions to its customers. It uses the active search recommender which suggests the buyer items which are similar to his previous searches and similarity-wise [6]. This form of recommender uses the technique called user to use collaborative filtering as the algorithm does the searching of items with users having similar patterns of purchasing. There are millions of choices of music available online. A proper method of filtering and prioritizing among them is needed for different users. To decipher such problems, recommender systems play a vital role by providing the results of top data which is dynamically generated using the required methods and data [7].

3. Proposing system

To achieve the goal of the paper, the first process is to do enough background study, so the study will be conducted. The whole paper is based on a big amount of music data so that we choose the quantitative research method which is shown in figure 1. For philosophical assumption, positivism is selected because the paper is experimental and testing character. An approach is a deductive approach as the improvement of our research will be tested by deducing and testing a theory. Ex post facto research is our research strategy, the music data is already collected and we don't change the independent variables. We use experiments to collect music data. Computational mathematics is used in data analysis because the result is based on the improvement of algorithms. For quality assurance, we have a detailed explanation of algorithms to ensure test validity. Similar results will be generated when we run the same data multiple times, which is for reliability [8]. We ensure the same data leading to the same result by different algorithms.
- Popularity algorithm
- Collaborative filtering recommendation
1. User-based collaborative filtering
2. Item-based collaborative filtering

FLOWCHART
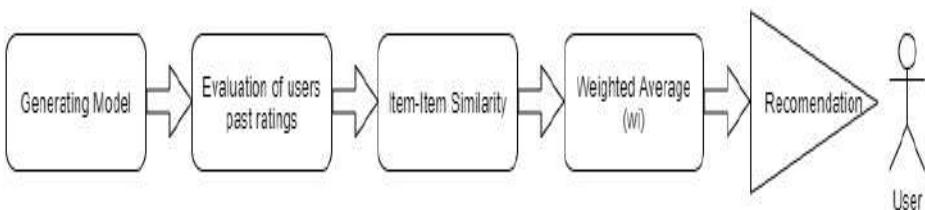
Figure 1. Control flow chart for Proposed model.

We model the dataset and using Hadoopsten's model of item-based collective filtering (using Hadoi's model). We focus our predictions on users' previous ratings when we use this CF. Another approach to determine item similarities is to build a unit-item matrix by utilizing the data already in the system. To sum up, more wood was needed to finish the staircase than to make the tree and its surroundings as spectacular as it was anticipated, yet more scenery was needed to make the surroundings look acceptable as it did." The similarity coefficient (competition coefficient) is used for the estimation. It is the ratio of the intersecting set to the union set of objects or variables taken in the relationship as given in equation 1. When any user rates an item, the top n similar items are derived from the similarity matrix which is recommended.

$$T(x, y) = \frac{Nz}{(Nx + Ny - Nz)} \qquad \text{E.q (1)}$$

Where N_z is the No. of Attributes in the intersection set and N_x and N_y is No. of Attributes in object X and Y respectively.

4. Approaches in Recommendation System

4.1. Popularity model:

So far, the audience or crowd are reliable with their knowledge contributes to the majority of new business expansion strategies' advice and feedback, which are strong overall. Leveraging content is the core concept of a recommender framework to rank and sort long-tail users' particular preferences, in a straightforward way is rather than serving the objects with simple precision [9]. One limitation of the effective recommendation filtering for active filters is "popularity". This problem originated with the Long Tail phenomenon, which states that a large number of users use very few but popular items while a small number of users consume less popular items. Since collaborative filtering is based on the preferences of people to produce recommendations, it leads to poor variations of recommendations (since most people prefer to use only popular items). E.g. Celma has shown that the music industry follows a long tail [10]. This algorithm isn't personalized; it simply recommends the most popular items to a user. As the popularity is based upon the people count hence it provides better results. The final motive of a system is to provide the best recommendation based on the available features that are both user data and song data.

4.2. Collaborative filtering recommendation:

It is a very common method not only for music recommendations but also for other types of recommendation systems. This method relies on a given user content (ratings or full response), and based on the "word of mouth" process of recommended content is recommended by the user if liked by the same users. Because of this, filtering systems do not need to deal with content, which means they do not support the decision to recommend something or not in the description, or the physical properties of an item

[11]. In the case of music recommendation, it allows avoiding the task of analyzing and categorizing music content. This is an important advantage, given the complexity of music signal analysis and music metadata. And this can be done in two ways collaborative and content-based filtering which shown in figure 2.[12].
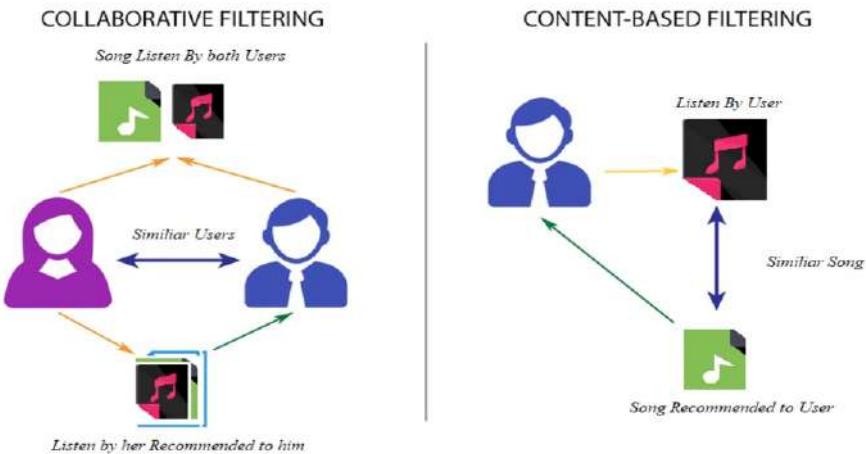
Figure 2. Shows the difference between CF and C-BF.

4.3. User-Based Collaborative Filtering:

User-based filtering is different, it takes into consideration of user-liked items, and depending upon that it predicts better results. The first step of item-based filtering is to calculate items that are liked by the users so they can be recommended to other users. The core point of item-based filtering is to calculate the matching of items. In collaborative filtering, the users are considered similar when they like similar items.

In collaborative filtering, the users are considered similar when they like similar items in Equations 2 and 3 and Table 1.

$$S_{U,V} = |N(u) \cap N(v)| \, |N(u) \cup N(v)| \quad \text{E.q(2)}$$

There are a lot of similarity algorithms, formulas

$$P_{ui} = \sum_{v \in S(u,k) \cap N(i)} S_{u,v} \, R_{v,i} \qquad \text{E.q(3)}$$

Table 1. an example of a User-based CF recommendation. According to the interesting history of User A, only User C can be the neighbor him, so Item D will be recommended to User A.

User/Item	Item-1	Item-2	Item-3	Item-4
User-A	Yes		Yes	Recommend
User-B		Yes		
User-C	Yes		Yes	Yes

4.4. Item-Based collaborative filtering:

Item-based filtering is different, assuming users will like the same things that user preferences do. So, the first step in item-based filtering is to find things like what the user liked before. The main point of item-based filtering is to calculate the similarity of two items. The CF items consider that users' preferences are the same. Item CF considers that items that are liked by more same users, the more similar they are. Assume N(i) and N(j) are user sets who like i and j respectively. Hence the similarity of i and j can be defined as in equations 3 and 4 and Table 2.

$$S_{i,j} = |N(i) \cap N(j)| \, |N(i) \cup N(j)| \qquad \text{E.q (3).}$$

$$P_{ui} = \sum_{j \in S(i,k) \cap N(u)} S_{i,j} \, R_{u,j} \qquad \text{E.q (4).}$$

While User C prefers Item 1, so we can find that maybe User C loves Item C too.

Table 2. Item-based CF

User/Item	Item-1	Item-2	Item-3
User-A	Yes		Yes
User-B	Yes	Yes	Yes
User-C	Yes		Recommend

5. Results and Discussion

The best algorithm is given by the line which occupies much space in the graph when the two algorithms collaborative and popularity are taken into consideration. From the results of both popularity and collaborative algorithms, we can observe that collaboration gives high efficiency when compared to popularity. Hence the collaborative filtering algorithm recommends well. The popularity curve is represented by blue color and the collaborative curve is represented by an orange curve, high the curve good is the performance. As the orange curve is more the performance of collaboration is efficient when compared to popularity algorithms. Which was shown in Table 3.

Table 3. Comparison between popularity algorithm and collaborative algorithm

Model	Precision	Recall
Popularity algorithm	0.87	0.83
Item- Based Collaborative filtering algorithm	0.90	0.86
User-Based Collaborative filtering algorithm	0.92	0.88

6. Conclusion

Understand how the recommendation works in the case of songs and provides the users with the best results. The final result made us understand how to use the data to get a good recommendation output. Furthermore, the results made us understand that good precision and recall from our work improved the overall result. Finally, few methods have been developed that work for providing the users with the best results using details like listen, count, user id, song – artist. Recommendation programs have proven to be the best solution to solve the problem of information overload. Decisions can be made more quickly and easily by the use of time and resources saved Research on music, human behavior, and how it is linked to the impact of music has risen over the last few years, and in particular, to the expansion of the temporal lobe has been active over the last decade Since music is so vital to our daily lives, now that we have greater advancements in technology, we can connect with people anywhere people, more easily. It is very difficult to satisfy the requirements of one's interests and maintain the service in the long term while serving clients are so diverse. Therefore, prospective recommendation systems can enable decisions that are made more intuitive to the consumer, so that they can make the best decisions possible. And of course, it will provide automatic music suggestions to bootleg copies of the song results, which will please the consumer as well. In this paper, we have described the elements of the music recommendation system and the various models that can be used for recommendation such as popularity, collaborative item filtering, user-based filtering. The cooperative complementary model has achieved great success but has issues such as willpower, human effort, etc. Future work will focus on expanding the existing methods and algorithms used so that the forecast recommendation system and the quality of the recommendations can be improved.

Reference

[1]. Liang, D., Zhan, M., & Ellis, D. P. (2015, October). Content-Aware Collaborative Music Recommendation Using Pre-trained Neural Networks. In ISMIR pp. 295-301.

[2]. Thota, J.R., Kothuru, M., Shanmuk Srinivas, A., Jitendra M, S.N.V. (2020) " Monitoring diabetes occurrence probability using classification technique with a UI" International Journal of Scientific and Technology Research, 9 (4): 38-41.

[3]. Amiripalli, S.S., Venkatarao, R., Jitendra, M.S.N.V., Mycherla, N.M.J. (2020) "Detecting emotions of student and assessing the performance by using deep learning" International Journal of Advanced Trends in Computer Science and Engineering, 9 (2): 1641-1645.

[4]. Potharaju, S. P., & Sreedevi, M. (2017). A Novel Clustering Based Candidate Feature Selection Framework Using Correlation Coefficient for Improving Classification Performance. Journal of Engineering Science & Technology Review, 10(6).

[5]. Jitendra, M.S.N.V., Radhika, Y. (2020). "A review: Music feature extraction from an audio signal" International Journal of Advanced Trends in Computer Science and Engineering, 9 (2): 973-980.

[6]. Xing, Z., Wang, X., & Wang, Y. (2014, October). Enhancing Collaborative Filtering Music Recommendation by Balancing Exploration and Exploitation. In Ismir : 445-450

[7]. Wang, H., Wang, N., & Yeung, D. Y. (2015, August). Collaborative deep learning for recommender systems. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining : 1235-1244).

[8]. Amiripalli, S. S., & Bobba, V. (2020). An Optimal Graph based ZigBee Mesh for Smart Homes. Journal of Scientific & Industrial Research Vol. 79(4): 318–322.

[9]. Kavinkumar, V., Reddy, R. R., Balasubramanian, R., Sridhar, M., Sridharan, K., & Venkataraman, D. (2015, August). A hybrid approach for recommendation system with added feedback component. In 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI) : 745-752. IEEE.

[10]. Sunitha, M., & Adilakshmi, T. (2018). Music Recommendation System with User-Based and Item-Based Collaborative Filtering Technique. In Networking Communication and Data Knowledge Engineering (pp. 267-278). Springer, Singapore.

[11]. Mukkamala S.N.V. Jitendra and Y. Radhika, "Singer Gender Classification using Feature-based and Spectrograms with Deep Convolutional Neural Network" International Journal of Advanced Computer Science and Applications (IJACSA), 12(2), 2021. http://dx.doi.org/10.14569/IJACSA.2021.0120218.

[12]. Potharaju, S. P., Sreedevi, M., Ande, V. K., & Tirandasu, R. K. (2019). Data mining approach for accelerating the classification accuracy of cardiotocography. Clinical Epidemiology and Global Health, 7(2), 160-164.

#Vaccine: Using Hashtags from Indian Tweets to Capture and Analyse the Sentiments of People on Vaccination for Covid'19 Pandemic

K. Anuratha [a,1], S.Sujeetha [a], J.M.Nandhini [a], B.Priya [a] and M.Paravthy [b]

[a] *Dept of Information Technology, Sri Sai Ram Institute of Technology, Chennai, Tamilnadu, India*
[b] *Department of CSE, Sethu Institute of Technology, Madurai, Tamilnadu, India*

Abstract. To prevent the public from pandemic Covid'19 the government of India has started the vaccination from mid of January 2021. The government has approved the two vaccines, Covishield from the university of Oxford and Covaxin from Bharat Biotech.The vaccination started with frontline workers and is further extended to common public prioritizing the elders of above 60 years and people aged 45 years above with co morbidities. Though many people have got benefitted from it there is still a group of people not convinced with the vaccination. We have carried out this work to analyze those Indian people sentiments on the vaccines through the hash tags of tweets. The results show that though majority of the community has a positive belief on the vaccines but some of them still express negative emotions.

Keywords. COVID, Vaccine, India, Covaxin, Covishield, Pandemic

1. Introduction

The covid-19 pandemic has infected about billions of people across the globe and deaths [1]. Major countries have imposed lockdown at different phases in 2020 but it's not the feasible solution considering the economic benefits of any nation. The only possible solution is to break the chain of the infection and is possible only through immunizing the public with vaccination. As India is one among the country with majority of population affected by the pandemic, government has announced the vaccination from January 2021.The nation has approved two vaccines covishield and covaxin from the University of Oxford and Bharat Biotech respectively.

[1] K. Anuratha, Assistant Professor, Department of IT, Sri Sai Ram Institute of Technology, Chennai
Email: anuvinaya2003@gmail.com.

The vaccination benefits for about 40 million of healthcare workers and frontline workers. The nation aims at benefitting for a about 300 million people on priority basis [2]. As the Government of India has approved the two vaccines as mentioned and also there is huge demand and difficulties in logistics it's not possible for the individuals to opt for the vaccine. They need to proceed with the plans of the government [3]. The people, who are vaccinated and waiting for their chance to get vaccinated, express their views in social platforms. In this work we made use of emotion-word hash tags used in twitter in analyzing the views of general public towards the government initiative to contain the pandemic in India. Twitter has become the popular medium for day to today information sharing on various fields. It has around 400 billion active users, daily. Not only for information sharing, public uses this platform for sharing their thoughts. Researchers make use of these opinions to analyze the sentiments of peoples on various issues like politics, sports, product reviews etc., The emotions expressed by the Twitter community during 2020 has been used to understand the people sentiments during a pandemic [7-12].

2. Literature Review

In [13] the authors have used the Twitter data with the covid-19 keyword to analyses the sentiments of Indonesian public towards the vaccine for covid. They have shown negative sentiments were more than that of positive sentiments. Deep learning algorithms have become popular in recent years for sentiment analysis [14].
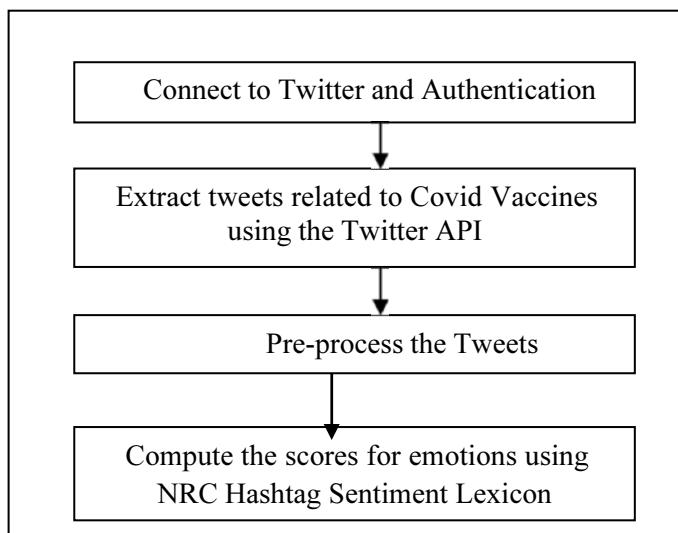
Figure 1. Steps in analysing the sentiments expressed in tweets

3. Proposed Methodology

The emotions **ex**pressed in the tweets extracted using the hashtags (#) are analyzed using deep learning algorithms as mentioned in the figure 1
As we have to analyse the Indian tweets associated to vaccine covishield and covaxin we have prepared two different data sets with the tweets collected during January and February 2021.   Once we are done with the data sets, we have pre-processed the data. Then we have adopted the NRC hash tag sentiment lexicon to categorize the tweets in to emotions anticipation, joy, anger, disgust, fear, trust and surprise.

4. Results and Discussion

The results of the research are discussed here. Figure 2 shows people have tweeted with 82% positive sentiments and 18% negative sentiments for the vaccine covishield. For Covaxin, the tweets with positive sentiments were with 68% and 32% were with negative sentiments.
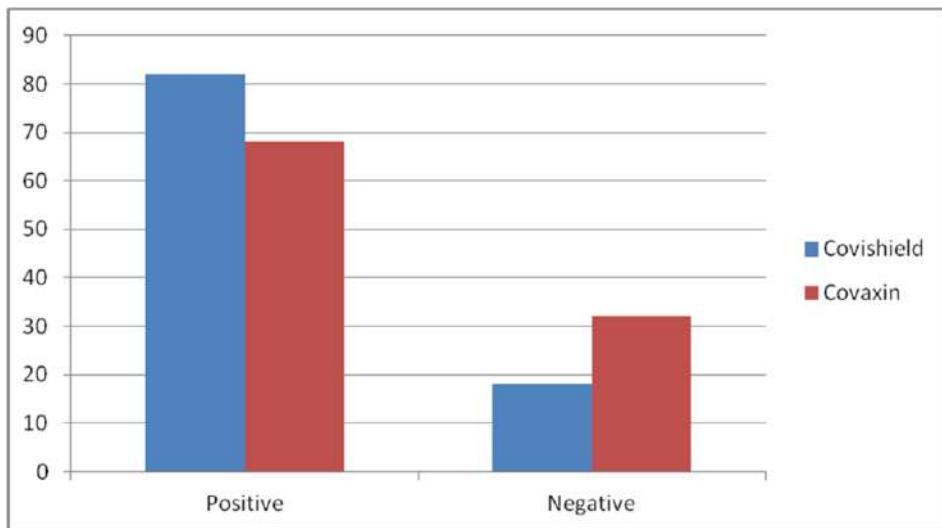
Figure 2. Sentiment Analysis of Indian Tweets on the Vaccines for Covid'19

Then, the emotions on the tweets were analyzed and the same is mentioned in Figure 3. It is observed that the emotions joy, trust, anticipation on Covishield is high than that of Covaxin. The emotions anger, fear and digest associated with Covaxin is high than Covishield trust and anticipation emotions associated with Covishield was more than the value of trust and anticipation associated with Covaxin. While the emotions surprise and sadness associated with both are equal.
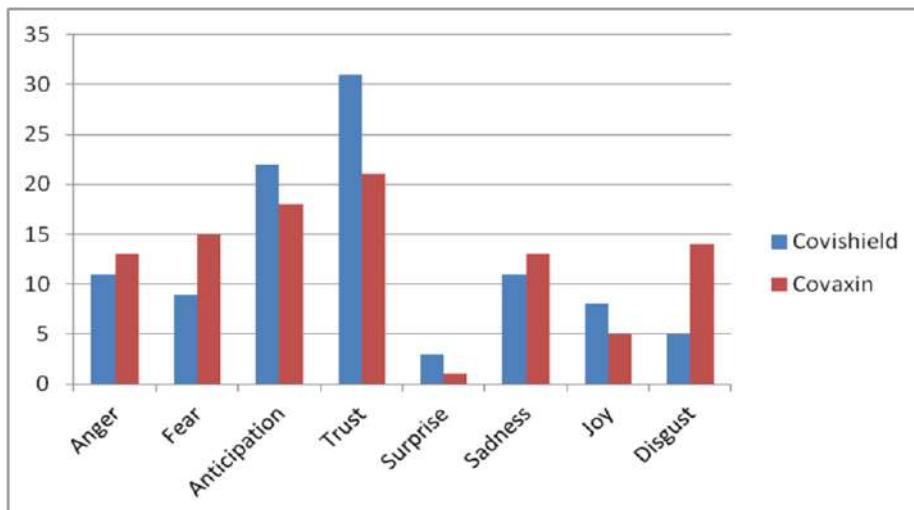
Figure 3. Emotion Analysis of Indian Tweets on the Vaccines for Covid'19

5. Conclusion

This work was carried out to analyze the impact of Vaccination in India through tweets of people. It was observed that the vaccine Covishield had positive sentiments higher than that of Covaxin. Also, the emotions like trust, anticipation. In the study, it was found that the tweets regarding Covishield had more positive sentiments as compared to Covaxin. Emotions like trust and anticipation were more in tweets associated to Covishield. The emotions anger and disgust were expressed more for tweets associated to Covaxin as compared to Covishield.

References

[1] Jagdale, Rajkumar S., Vishal S. Shirsat, and Sachin N. Deshmukh. "Sentiment analysis on product reviews using machine learning techniques." In Cognitive Informatics and Soft Computing, pp. 639- 647. Springer, Singapore, 2019.

[2] UMA RAMYA, V; THIRUPATHI RAO, K. Sentiment Analysis of Movie Review using Machine Learning Techniques. International Journal of Engineering & Technology, [S.l.], v. 7, n. 2.7, p. 676-681, mar. 2018.

[3] M A, Ravikumar P. Survey: Twitter data Analysis using Opinion Mining. International Journal of Computer Applications. 2015;128(5):34-36.

[4] UPI Systems in India Post Demonetization Using IBM Watson." In 2018 International Conference on Computer Communication and Informatics (ICCCI), pp. 1-6. IEEE, 2018.

[5] Wang Y, Taylor J. Coupling sentiment and human mobility in natural disasters: a Twitter-based study of the 2014 South Napa Earthquake. Natural Hazards. 2018;92(2):907-925.

[6] Dubey, A. D. (2020). Twitter sentiment analysis during COVID19 outbreak. Available at SSRN 3572023.

[7] Manguri, K. H., Ramadhan, R. N., & Amin, P. R. M. (2020). Twitter sentiment analysis on worldwide COVID-19 outbreaks. Kurdistan Journal of Applied Research, 54-65.

[8] Bhat, M., Qadri, M., Noor-ul-Asrar Beg, M. K., Ahanger, N., & Agarwal, B. (2020). Sentiment analysis of social media response on the covid19 outbreak. Brain, Behavior, and Immunity.

[9] Pastor, C. K. (2020). Sentiment analysis of Filipinos and effects of extreme community quarantine due to coronavirus (Covid-19) pandemic. Available at SSRN 3574385

[10] Pristiyono et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1088 012045

[11] Rev., Data Mining Knowl. Discovery, vol. 8, no. 4, Jul. 2018, Art. no. e1253, doi: 10.1002/widm.1253.

Help Farmers - Farm Era App

Mrs.D.Thamizhselvi [a,1], Ms.S.V.Supraja [a] and Ms.G.Priyadharshini [a]

[a] *Dept of IT, Sri Sairam Engineering College, Chennai, Tamilnadu, India.*

Abstract: The income of the famers has decreased drastically over the past years as they do not have the proper channel for marketing their produce. This has also proved to be the factor that favors the landlords and money lenders to gain possession over their agricultural products at a very low cost and obtain a large profit from it. This also reflects the inability of farmers to obtain the righteous profit from their produce. The main aim of our project is to organizing and uniting the farmers under one umbrella, to reduce the unbalanced accumulation of the profit from perishable farm produces for the traders and the sellers and help maximise the income level of the farmers by self-marketing. This system has been implemented by considering the entire supply-demand eco system and it also helps avoid product wastage.

Keywords. Marketing, Direct Selling, Direct Marketing, E-business, E-commerce, Self-Marketing.

1.   Introduction

Food wastage is a major problem in India. It is estimated that around Rs 58,000 crores worth of food is wasted each year during production, processing, retailing and consumption. About 25-30% of the fruits and vegetables produced are wasted due to poor logistical support, lack of proper storage, and a weak marketing sector. India ranks 103 out of 119 countries on the Global Hunger Index (GHI). This wastage of food products has a huge impact on the lives of the general public (the consumers) and the producers. The solution to this problem is an android application with an interface that allows the producers to register themselves, post details of their product (i.e the quantity, contact details and the location). A separate interface for the transport service providers will be given where they provide the details like their service range, the quantity that they can transport and so on. Further, yet another interface for our app would be for the end consumer. Now our application consolidates all the details given and finds the best transport service provider to accommodate the products as per the order placed by the consumer. Our app aims to solve this issue by providing a way for farmers to identify the demand of each product in different regions.

---

[1] Mrs.D.Thamizhselvi,Dept of IT, Sri Sairam Engineering College, Chennai
Email: thamizhselvi.it@sairam.edu.in.

Instead of letting the food go to waste, the farmers can transport the food to various other marketplaces, rather than just the nearby/local marketplaces [1-4]. Our app helps carry this out by letting the farmers, marketers and consumers communicate with each other. This eliminates the need for brokers and other such middlemen that usually add to unnecessary expenses. The application can support farmers to gain maximum profit .The application can support customers to get a fair price for the products. Efficient access of data from firebase for better data handling. Can solve the problem of getting transport at the right time for a lower fare [5-7].

2.   Literature review

In 2019, IEEE published a paper 'Crop shop an application to maximize profit for farmers', but the drawback is the app did not give suitable solution for transport. In 2017, Agriculture Economic Research Association published a paper 'Challenge options and strategies for doubling farmers income in west Bengal', drawback the application did not provide a proper justification for supply demand. In 2017, SSDN publishers published 'Doubling farmers income ', the drawback is it didn't provide good marketing solutions. In 2015, IEEE has published 'Virtual fruits market' and the con of the project is it did not give solution for pricing and this project was based only on fruits [8].

3.   Proposed Module

The major problem prevalent among farmers is the denial of the right profit reaching the farmers. This problem happens mainly due to the reasons such as lack of proper channels for marketing the product and the fear of crops being wasted if they don't find the right buyer at the right time. This problem also leads to the condition where the farmers sell the goods for the lowest price to a person who acts as an intermediate between the actual buyers and the farmers in figure 1. They tend to retail these goods for a much higher price to yield higher profit which actually belongs to the farmers. This also reflects in the higher selling price of those products in the market for the consumers. If the above. The situation is to be stopped. The farmer may not find the proper buyers at the right time and the products may be wasted [9].

4.   Producer Module

This module is constituted by the interface provided to the farmers to upload and view the product details. The producers can make use of the application to post the details of their agricultural produce prior to the cultivation date in figure 2. The details being posted consist of the actual description of the product, the time of availability, available quantity and the price quoted by the farmers for their products. The payment interface is used to view and verify the payments made to the farmers through the application. The producers can also avail transport facilities to deliver the products to the consumer.
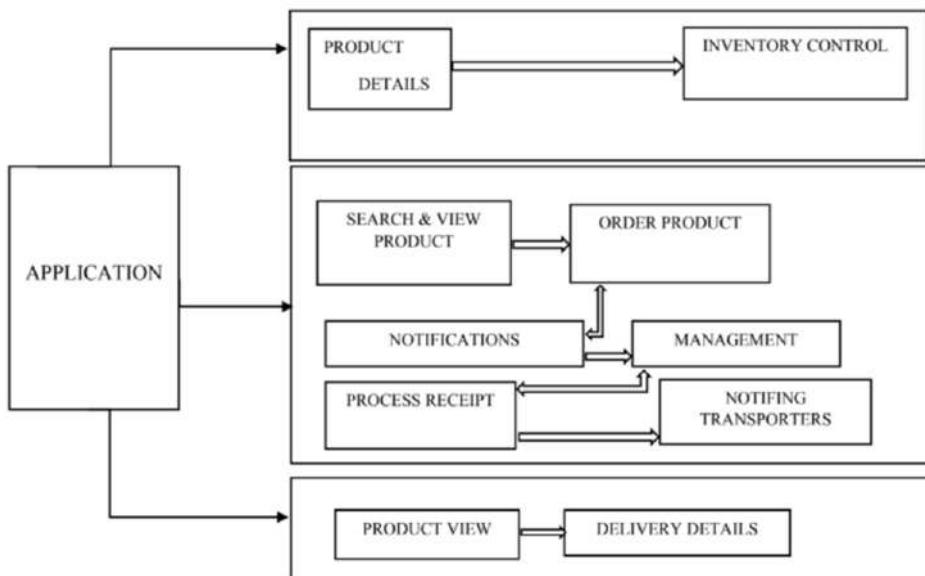
Figure 1. System architecture of the application
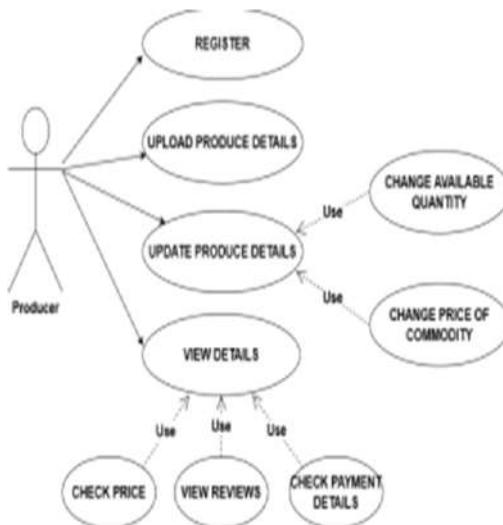
Figure 2. Use Case diagram of Producer module

5.   Customer Module

The consumer module helps the consumers to buy the agricultural products online directly from the farmers. The consumers can view various products posted in the application and also find the required products for their consumption. Consumers can also view the location details of the product The products can also be viewed based on

the user requirements. The payment interface in the consumer module helps to make payment directly to the producers to ensure the profit being directly gained by the producers. Figure 3 Consumers are also capable of availing transport through the application that helps them to deliver the products to the required location.

Figure 3. Use case diagram of consumer module

6.   Result and Discussion

Android Studio, Java programming language, Firebase are the software requirements The mobile application has been developed to bring out the smart farming technologies that enable the farmers to increase profit. First the cultivators should login into the application by giving the basic information. The harvesters/cultivators who are going to harvest their crops in the near future, such as 2 or 3 days in advance, can provide their product details. They can enter the details such as the approximate amount of products that can be produced, estimated cost of those products and the location of those products. Based on their location the users, cultivators, marketing dealers can choose the logistics services. Farm era currently does not have price prediction and proper price exchange, while strategies are still being explored; we are currently focusing on the user base. And by further analysis and data collected from the app can improve its facilities.

7.   Conclusion

The goal of our project is to sell the farmed products easily and give some sustainable profit to the farmers and to give the good product satisfaction to the consumers. We are in the early stages of the research and we focused on a particular group of people. And the research will continue with a quantitative study .By this we can see better marketing and also understand and predict post consumption behaviours. In summary, we need to expand the scope for this research in order to test the reliability of the results.

References

[1]   Gunasekara, A; Patel, C. and Tirtiroglou, E. (2001) 'Performance Measures and metrics in Supply Chain Environment'

[2]   Haan, J.D.; Groot, D.G.; Loo, E. and Ypenberg M. (2003), 'Flows of Goods or Supply Chains: Lesson from Natural Rubber Industry from Kerala, India'

[3]   Jha, S. and Srinivasan, P.W. (1999), 'Grain Price Stabilisation in India: Evaluation of Policy Alternatives'

[4]   Joshi, P.K., Gulati A, and Cummings Jr R. (2007). 'Agricultural Diversification and Smallholders in South Asia'

[5]   Khan, K.B. and Mentzer, J.T. (1998). 'Marketing's integration with other departments'

[6]   Kumar P. (2002). 'Agricultural Performance and Productivity. Indian Agricultural Policy at the crossroads'

[7]   Bertolini, R. 'Making Information and Communication Technologies Work for Food Security in Africa- 2020'

IoT Based Security Alerts for the Safety of Industrial Area

K.G.Kharade[a, 1], S.K.Kharade [a], V.R.Sonawane [a], S.S.Bhamre [a],
S.V.Katkar [b], R.K.Kamat [a]

[a] *Department of Computer Science, Shivaji University, Kolhapur, MH*
[b] *Department of Information Technology, MVP's K.B.T. College of Engineering, Nashik, MH*

Abstract. The security of any business plays a vital role. All enterprises expect high security because of the increase in robbery. It is challenging to manage security with traditional ways of protection. This paper emphasizes the sensor-based security system to protect against any unwanted entry in the business area. This system is developed using IoT-based sensors, and electronic materials develop the security system. The present scenario ensures protection and security have become inevitably necessary. There is regressive progress in the protection sector as the influence of new technology is hitting its height. It's well-known as a modern home when there is a current home with minimal human effort. This technology aims to automate industrial area security and partially replace the security individual, enabling us to monitor unsuspecting activities and be warned during critical situations. Since wireless and emerging technology is taking place, an automated intelligent protection system is being introduced.

Keywords. Home automation, Home appliances, IoT, Motion Sensors

1. Introduction

Internet of Things It is the concept of connecting real-world objects to the internet and monitoring their performance remotely through the internet. It envisions the interconnected and testable items connected via the internet [3]. IoT is used to consolidate real, virtual, and computerized conditions to make a savvy domain that simplifies life [20]. It is another period of synchronization and communication between machines and objects or situations command-and-and-and-control methodology can benefit our lives. Recently, technology has been making a recovery with a slant or rising, or running, as well as something that tilts, such as the Internet of Things [7]. Innovation is required to govern the world within a couple of years. Items can communicate with data and can make vital choices at whatever point is required. When items can speak with one another using the web, we have to take a favorable remote access position [1]. The ultimate goal of IoT is to make artificial life frameworks use new robotics frameworks rather than using new frameworks to automate human-based ones. The sensor-based business has many benefits; it is flexible, simple to use, easily applied, and inexpensive [2]. Almost all current computing devices can be connected to WiFi or be associated with a smartphone and home computers. Apart from that, they

---

[1] K.G.Kharade ,Department of Computer Science, Shivaji University, Kolhapur, MH, India
Email: kabirkharade@gmail.com

don't require the use of one another, these tools to work, nor can they exchange information with one another, these devices [22]. Current home security and surveillance systems are trying to overcome the problems of conventional security systems [4]. Using only motion sensors, you can secure your home from anywhere in the world; you can gain remote access to the system to know the current state of the IoT. People enrich their lives through the provision of a variety of service options [5]. It seeks to meet people's and enable their simple needs [18]. The most innovative and most cost-effective way to handle all things regarding your house, in general, is to replace all the things you have with the items you want [21]. Advances in automation have eased life in all areas of our lives. The adage "automation" is replacing "manuality" these days [23]. There has been a recent explosion in the number of internet usage, and emerging internet technologies, such as the Internet of Things, are part and parcel of daily life now-day activities [19]. While you're working, the Internet of Things network includes machines and consumer goods made to communicate and get the job done when you aren't .
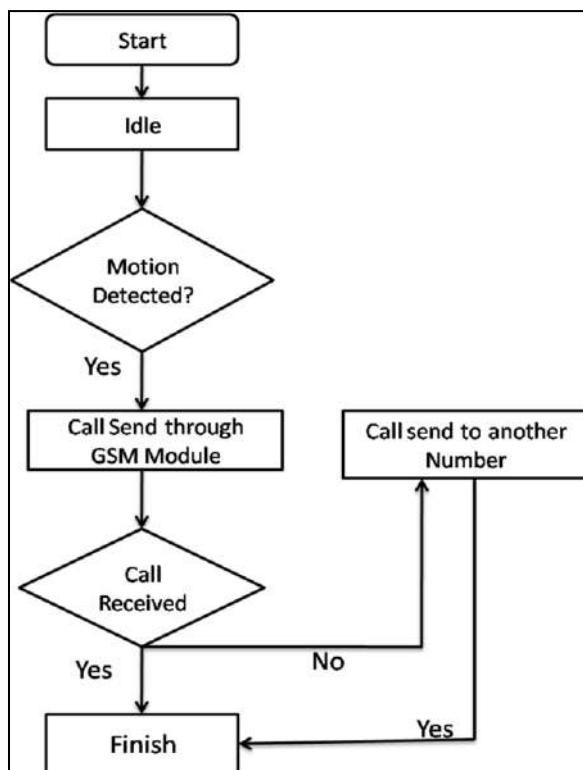
Figure 1. Flowchart of the system

2. Existing System

[26] Observed that, in this rapid implementation of networked digital technology is found in the home. It connects new and existing devices to a higher level of automation

and provides additional options for existing ones. the remote control has started to develop. [27] Stated, In IoT, internet connectivity is available for everything in the physical world, including things like smart meters, intelligent door locks, intelligent appliances, and other objects. Virtualization allows various processes to be done without synchronizing between the hardware devices. Without restricting the number of locations, the Internet of Things can run multiple tasks on unlimited resources and distant locations using intelligent devices and fast networks. [28] described Smart houses would gain in desirability due to the benefits of automation. The main problem with safety and security in the home is that they have been neglected for too long. The safety and security of people in the province of modern technology focus on Home Automation features, including a camera functionality that supports it. If movement is detected at the house's entrance, a warning is sent that includes a real-time picture of the entrance. The house owner will receive this message from the internet, allowing the app to send a notification. [29] reported, Smart homes require complexity control in their various gadgets, which are fundamentally electronic appliances. Advanced mobile phones are now highlighted consummate and can be made to conveyor interface with different devices in an ad-hoc network with available Bluetooth and WiFi options. With the appearance of portable phones, Mobile application improvement has seen a significant episode. Figure 1 Flowchart of the system to utilize this chance for a smart home, we select the cell phone ordinarily because it is found in an ordinary family that can be participated in a brief organization inside a home with electronic equipment. Android, by Google Inc., gives the stage to advance the portable applications for the Android gadgets.

3. The Need for PIR Sensor

PIR is a short form of Passive InfraRed motion detectors. The logic behind the working of PIR. is simple. Both humans and animals dissipate heat; the fluctuation in the surroundings' emitted energy can be measured. When a human enters the surroundings, the sensor detects the radiated energy's change within the surroundings, concerning the previously noted energy levels [5].

4. Proposed System

The system is designed to give alerts when motion detects in prohibited areas. For this, once the space and time of surveillance are selected, adding the list of mobile numbers over which you want to receive alerts. The system will notify the registered mobile numbers with an alert when the motion is detected in the designated area of surveillance. The Motion detection sensor is connected to Arduino to see all the activities in the prescribed zone. It will trigger action through the GSM module to the specified numbers. The fail-safe mechanism is available if the first registered number is unavailable; the system redirects the call to the successive number registered. Once notified, the user will be able to take necessary action depending on the need.
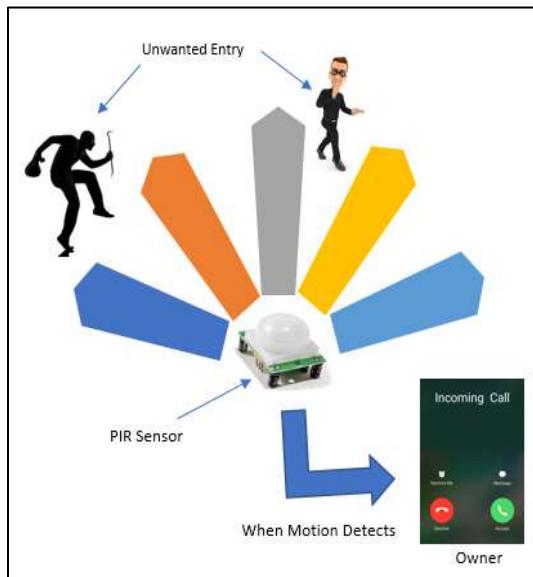
Figure 2. A working model of the system

5. Hardware Requirement

Following hardware list required to develop the system [13] [16]
  a) **GSM Module SIM 900A:** This module is commonly used in mobile phones and PDAs. Also useful for IoT (Internet of Things) and Embedded Applications development, it can also be used as a module. This SIM900A radio engine works on frequencies that use the terms EGSM 900MHz and DCS 1800MHz.
  b) **PIR Sensor:** The sensors used to detect human motion work well with PIR motion sensors.
  c) **5V power supply:** One of the most popular power supplies today is a 5V model. When the output of a regulated 5VDC power supply is being regulated, a dissipative regulating circuit is used.
  d) **Node MCU:** The Node MCU project offers free hardware schematics and board designs for open source design prototyping.
  e) **Data cable:** Cables used to transfer electronic data from a source to a destination are referred to as data cables. type of data cabling is either copper or fibre optics.
  f) **SIM Card:** The SIM cards are small, removable, and transferable chips that contain chips that can be placed in other phones.

6. Advantages of the System

This system has the following advantages;
a) Low-cost system and that can take care of industries in the absence.

b) No need to invest in a separate smartphone.

c) The system is platform-independent

d) The same device can be deployed for many industries where security is a significant concern in our absence [14].

7. Specialized area where this system can be implemented

This system can be implemented in the areas mentioned below, but not limited to [15].

a) Demilitarized zone
b) Home/Shop security
c) Hospitals
d) Hotels
e) Laboratories
f) Penitentiary
g) Banks
h) Institutes

8. Result and Discussion

As discussed in Figure 2, it is observed that once you install this system in your enterprise, it will keep checking all the entries on their premises. Once it found any unwanted access, first it will play an alarm, and at the same time, it will call the number specified. If the person didn't receive the call, the call would be transferred to the following number available in the system. In this way, security can be maintained of your enterprise in your absence. The developed method is very cost-effective but valuable. There is a provision of deciding the range of your system with the help of potentiometer available in the system.

9. Conclusion

This system has been developed at significantly less cost but having huge benefits. Once you implement this system into your business, the consumer does not need to think about the security issues as this system will take its care. Android-based automation offers a simple and attractive interface and makes the device more stable and versatile. To connect with the home automation system, we use mobile devices. To power the computers, WiFi is used to communicate between the Arduino and the android program. The overall device is constructed using the SIM 900A Specification GSM module, PIR sensor, power supply, MCU node, data cable, and SIM card.

References

[1]   Dongale, T. D., Katkar, S. V., Khot, K. V., More, K. V., Delekar, S. D., Bhosale, P. N., & Kamat, R. K. Simulation of randomly textured tandem silicon solar cells using quadratic complex rational function approach along with artificial neural network. Journal of Nanoengineering and Nanomanufacturing, VI(2), 2016,103-108.

[2]   Dongale, T. D., Kharade, K. G., Naik, G. M., & Kamat, R. K. Artificial Neural Network Modeling of NixMnxOx based Thermistor for Predicative Synthesis and Characterization. Journal of Nano And Electronic Physics,2017, 03042(1)-03042(4).

[3]   Indrajit Patil, S. J. A Survey on IoT Based Security System. International Journal of Advanced Research in Computer and Communication Engineering, 2016,29-32.

[4]   Katkar, S. V., Kamat, R. K., Kharade, K. G., & Kharade, S. K. Inclusion of .NET framework for Calculating electrical parameters of solar cell. International Journal of Research in Electronics and Computer Engineering, 2019,1167-1169.

[5]   Katkar, S. V., Kamat, R. K., Kharade, K. G., Kharade, S. K., & Kamath, R. S. Simulation of Cd(SSe) Solar Cell Using Artificial Neural Network. International Journal of Advanced Science and Technology, XXIX(2), 2020,2583-2591.

[6]   Katkar, S. V., Kharade, K. G., Kharade, S. K., & Kamat, R. K. An Intelligent Way of Modeling and Simulation of WO3 for Supercapacitor. In N. Thapa, Recent Studies in Mathematics and Computer Science ,Hooghly, West Bengal, India: Book Publisher International,2020,109-117

[7]   Kevin Otto, C. W. (2014). A comparison of the popular home automation technologies. EEE PES Conference on Innovative Smart Grid Technologies. Kuala Lumpur, Malaysia: IEEE Computer Society,2014,600-605.

[8]   Kharade, K. G., Kamat, R. K., & Kharade, S. K. Online Library Package to Boost the Functionality and Usability of the Existing Libraries. International Journal on Future Revolution in Computer Science & Communication Engineering, 2019,5-7.

[9]   Kharade, K. G., Kamat, R. K., Kharade, S. K., & Katkar, S. V. Automation of Paper Setting Process to Improve Effectiveness of The Examination System of The University. Journal of Emerging Technologies and Innovative Research, 2019,490-493.

[10]   Kharade, K. G., Mudholakr, R. R., Kamat, R. K., & Kharade, S. K. Perovskite Solar Cell Simulation using Artificial Neural Network. International Journal of Emerging Technologies and Innovative Research, 2019,336-340.

[11]   Kharade, K. G., Mudholkar, R. R., Kamat, R. K., & Kharade, S. K. Artificial Neural Network Modeling of CZTS Solar Cell For Predicative Synthesis And Characterization. International Journal of Research And Analytical Reviews, 2019,997-1006.

[12]   Kharade, S. K., Kamat, R. K., & Kharade, K. G. Simulation of Dye Synthesized Solar Cell using Artificial Neural Network. International Journal of Engineering & Advanced Technology (IJEAT), 2019,1316-1322.

[13]   Kharade, S. K., Kamat, R. K., Kharade, K. G., Kamath, R. S., & Shinde, S. A. Artificial Neural Network Based Efficiency Prediction and Its Impact on Dye Synthesized Solar Cell. International Journal of Recent Technology and Engineering, 2020,2696-2699.

[14]   Kharade, S. K., Kharade, K. G., & Kumbhar, V. S. Impact of Digital India on Various Sectors. Indian Journal of Innovation in Management And Excellence in Research, 2018,37-40.

[15]   Kharade, S. K., Kharade, K. G., Kamat, R. K., & Kumbhar, V. S. Setting Barrier to Removable Drive through Password Protection for Data Security. Our Heritage, 2020,19-23.

[16]   Md. Wahidur Rahman, M. H.-A.-R. Embodiment of IOT based Smart Home Security System. International Journal for Research in Applied Science & Engineering Technology,2018.

[17]   N. Sriskanthan, F. T. Bluetooth based home automation system. Microprocess. Microsystems, 2002,281-289.

[18]   R. K. Kodali, V. J.IoT based smart security and home automation system. 2016 International Conference on Computing, Communication and Automation (ICCCA), (pp. 1286-1289). 2016,Noida.

[19]   R.P.Pandav, S. G. Security System And Home Appliances Control Using IoT. International Journal of Advance Research and Innovative Ideas in Education, 2018,4593-4597.

[20]   Rao, G., Kumari, K., Shankar, D., & Kharade, K. G. A comparative study of augmented reality-based head-worn display devices. Materials Today: Proceedings.,2021, 1-9.

[21]   S.Hrushikesava Raju, D. R. IoT Based Home Automation System with Cloud Organizing. International Journal of Engineering & Technology, 2018,412-415.

[22]   Shaik Anwar, D. K. (IOT based Smart Home Security System with Alert and Door Access Control using Smart Phone. International Journal of Engineering Research & Technology , 2016,504-509.

[23]   Kumar, V. A., Kumar, V. A., Malathi, S., Vengatesan, K., & Ramakrishnan, M. (2018). Facial recognition system for suspect identification using a surveillance camera. Pattern Recognition and Image Analysis, 28(3), 410-420

[24]   Tanaya, K. V. Home Security System using IoT. International Journal of Pure and Applied Mathematics,2018, 1863-1867.

Study on SDN with Security Issues Using Mininet

Ms. Florance G [a,1] and Dr R.J Anandhi [b]

[a] Research scholar, Department of CSE, TOCE, Bangalore, Affiliated to VTU, Belgaum
[b] Professor & HOD, Department of ISE, NHCE, Bangalore, Affiliated to VTU, Belgaum

Abstract. The internet is faced with many problems daily, one of them is decrement in network bandwidth because of Distributed Denial of Service (DDoS) attack on host server, which deplete host resources. Researchers has been invented many protection mechanisms such as detection, trace back, prevention, reaction, and characterization are in case of DDoS attacks, which will control the number of malicious packets received by the victim. But it does not provide efficient detection technique with high rate in real time network infrastructure. Thus, modern technologies are prepared on Mininet network simulators, which give more impact to simulate the real network. The architecture of Software Defined Networks (SDN) and OpenFlow architecture is used to demonstrate a programmable network model and centralized management of real network. In this research work, we provide design of software defined network (SDN) using mininet simulator and security issues related to the Software Defined Network.

Keywords. DDoS attacks, Software Defined Network, OpenFlow, and Mininet simulator.

1. Introduction

Distributed denial-of-service (DDoS) attack makes any organization fail, thus resulting to stop providing services to legitimate users and exhaust the victim's resources. DDoS attacks can be classified as resource and bandwidth consumption. Attacker will perform DoS attack on more than one network to destroy the victim's resources, so that victim is unable to provide regular network services [2].

The primary cause of a Distributed denial of service attack is that attackers are typically accomplished by flooding the victim network from different sources. Nowadays high-rate DDoS (HDDoS) attacks certified with the persistent detection techniques quickly [2]. Presently, sending more packets to different network by using DDoS attacks which will flood traffic. Such attacks have been increased because the attacker wants to disturb the entire network to stop the legitimate packet to reach the destination. Due to presence of weakness in Internet Protocols, it becomes very easy for the attacker to find and exploit different loop holes in different applications.

The data plane and control plane are combined very strictly on the device in traditional networks. Consequently, creating the new set of programs to perform the task and changes in devices which is already present is a very monotonous task [10]. Software Defined Networks (SDN) surmounts these difficulties by separating the data plane and control plane, simplifying network management.

---

[1] Ms. Florance, Research scholar, Department of CSE, TOCE, Bangalore, Affiliated to VTU, Belgaum, Email: vijiflorance59@gmail.com.

The important role of SDN is to increase the capabilities of traditional networking system, where the network control is changing dynamically, data transfer, managing, adjustable, and extracted from SDN devices [1].

Software defined networks (SDNs) are one among the most frequently used software-based network model having loosely coupled control and data planes. By supporting the centralized control techniques and application programming interface, SDNs have enlarged the changeability of network management and its functions [1]. The network's control logic plane has controller is disconnected from the data plane by the Software-Defined Networking (SDN). SDN solves the problem of dynamic nature, scalable computing and storage needs of more computing environments that the consistent architecture of conventional network does not support [3].

2. Background and motivation

This portion expresses the biography as well as the necessary of defence mechanism for DDoS attacks. In the beginning, DDoS attacks were first started in August 1999 in case of different consortiums and keep attacking in different organization [1]. In the year of 2009, a DDoS attack was started that disturb the organization assistance of most preferred websites like Live Journal, Facebook, Amazon, and Twitter. In the year of 2010 to 2011, more than 80,000 computer systems in 2500 consortiums and 4 million computers in 100 countries were disturbed by DDoS attacks correspondingly. For each day, attacker started sending more than 8000 DDoS attacks. But recently DDoS attack has happened in short duration [2].

The Internet hosts are used to emanate a DDoS attack, because of having the frail protocols, insufficient attestation plans, insecure computer systems and operating systems. Weak systems require stringent protocols for implementing a higher security standard which give the better remedy to the secure systems. Attack detection techniques do the experiment of analysing on the attack packet and disclose the attacks then remove attack packet from normal traffic. One of the most important technique is Traceback technique which is used to trace source of attack and even in the case of spoofed IP addresses before the origination of attack [2]. During the processing time attack reaction strategy try to reduce the loss, which is created by DDoS attack. The Reaction stage mitigates the impact of attack and increments the quality of services delivered to the authenticate users under attack.

3. Software Defined Network architecture

Software Defined Network (SDN) is a new networking innovation, which permits centralized, programmable control logic planes and data plane conceptual that can overcome the drawbacks of present network infrastructures [3]. SDN is designed to make a network flexible and is logically centralized through SDN controllers. SDN controller provides a centralized vision of the whole network [9]. The reason behind SDN is to keep the data forwarding plane will be separated from the control plane so that network operators and service providers can directly control and manage their own virtualized resources and networks beyond using hardware mechanism. It is an application used to control packet forwarding all the devices in the network and manages flow control for improved network management [1]. Manual configuration is reduced, for individual network devices because of the forwarding policies. In SDN, the control and data planes are distinguished which in turn allows control to be

programmable and manageable such that control remains centralized and data plane to be shorten and conceptual [10].

The motivation that SDN is necessary for network operator and service providers are as follows [3],

- Network operator and service provider need to use SDN technologies to easily and efficiently control and manage the network.
- High complexity of operations and management in network can be reduced with software defined network rather than configured networks.
- Network operator and service providers should provide an interaction method between the infrastructure layer and network layer so that service should be securely isolated from existing traffic.
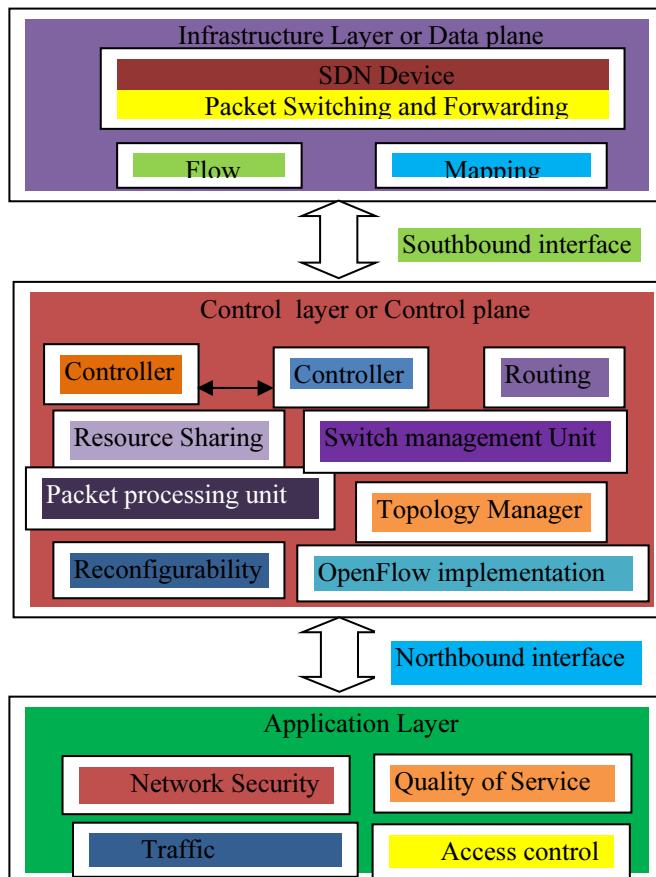
Figure 1. Architecture of SDN

- The controller aware of the necessities of the applications and resources available in the network infrastructure by using its north bound and south bound interface [3].

- Application program interface (API) open to grow a benefit from network infrastructure in terms of services in application layer, devices in infrastructure layer and programmability in control layer.

Figure 1 shows the typical architecture of SDN. It consists of three open interfaces [10]:

1. Southbound interface provide interface between control plane and data plane. It is used for creation of virtual network, dynamic reconfiguration of network, accessing resources and provide secure virtual network [3].
2. Northbound interface provide interface between control plane and application layer. It is used to provide routing related information, management related information and policy related information.
3. East west interface provide interface between controllers. It is used for intra domain; inter domain, scalability and interoperability.

3.1. SDN consists of three main layers:

3.1.1. Infrastructure Layer or data plane

The infrastructure layer responsible for transferring data between SDN devices (both physical devices and virtual devices) used to perform packet switching and forwarding. Flow table is used to store the rule populated by the controller for controlling and directing the packets [3]. Mapping table is used to store the data transfer between the different network and SDN devices communicate with control plane through south bound interface.

3.1.2. Control Layer or control plane

Control layer is placed between the infrastructure layer and application layer of the SDN network where network services are specified [10]. This layer manages with overall view of the network. Control plane is directly programmable in a centralized manner to provide hardware abstractions to SDN applications. It contains, set of controllers that interact with network services and SDN devices through north bound interface and south bound interface. Software that manages all the resources in network infrastructure is called Controller. Network infrastructure consist of packet switching and forwarding, mapping table and flow table provide a abstract view of the overall network and the controllers interact with each other through their east west bound interface to provide a stable view of the whole network infrastructure. Important functions of controller as follow [10],

- **Packet processing unit:** With respect to the protocol, processing of packet header and its payloads in the network.
- **Switch management unit:** Modification in switches and message arrival can be informed by controller.
- **Topology manager:** Maintain up to date information about network topology and changes in topology can be identified.
- **Routing:** By using forwarding table routing manager find out the routes to reach the destination address based on protocol.
- **Openflow implementation:** Controller performs the function related to openflow protocol such as action, table entry, flow rules, and message queues.

- **Management interface:** It provides access to functions that the controller provides.

SDN controller usually remains aware of all available network routes and can send packets based on different network characteristics are the important benefit.

3.1.3. Application Layer

Basically, it contains end-user applications that perform the SDN communications and network services [3]. SDN services such as access control management, network security, traffic engineering, load balancing, quality of service and other network function virtualization services. Flow of SDN devices in data plane is affected by communicating their necessities over northbound interface.

3.1.4. Advantages of SDN [4]:

- It allows a quicker response to modifying traffic (group of spoofed packets) conditions.
- It also supports more opportunity for dynamic provisioning, load balancing, monitoring specific traffic engineering, increase the utilization of network resource, and improve better occasion to implement many different types of solution.

4. Open Flow architecture

SDN is implemented using OpenFlow communications protocol which will access data from the data plane of a switch or router to control plane through the network. Communication between control plane and data plane is provided by this protocol. Packet can be moving with centralized decisions by using OpenFlow protocol [6]. Switch operations remains in control of OpenFlow controller. The action may be either Reactive or Proactive.

Reactive approach signifies that a switch will remain unaware of actions when a packet arrives. So, the packet is sent to SDN Controller [5]. By using OpenFlow protocol SDN controller responsible for inserting a flow entry into the flow table of switch. Switch totally dependent on the SDN controller is considered a major drawback.

Proactive approach overcome the drawback of reactive approach. Each entry in the flow table is pre populated by flow entries of each switch in case of Proactive approach. It does not disrupt traffic, even though if the switch loses the connectivity with control plane.

In OpenFlow architecture, we have set of OpenFlow instructions or commands that are transmitted from openflow control plane to open flow switch [7]. OpenFlow switch perform the following operations such as 1) Depends on the packet header fields identify and categorize packet from an incoming port 2) Packets can be processed in various ways by changing the header field. 3) Drop or push the packets to a respective egress port (outgoing port) or to the Openflow control plane.

Figure 2 describes the typical architecture of OpenFlow switch. OpenFlow switch consists of number of flow tables (organizing flows in table), group table (collection of action to be performed) and secure channel [7]. Flow table consist of flow entries which are forwarded, each entry match with its correspondent flow and packets, then provide the functions that are to be performed on the packets then sent. These flow entries have some set of parameters; 1) match fields used to perform for matching the

incoming packets and it uses the information there in the packet header, ingress port (incoming port), and metadata; 2) counters, used to make up the statistical data for each flow will be the count of received packets, amount and time limit for a particular flow; and certain group of commands, which apply when there is a match in table; they signify how to manage matching packets [6].
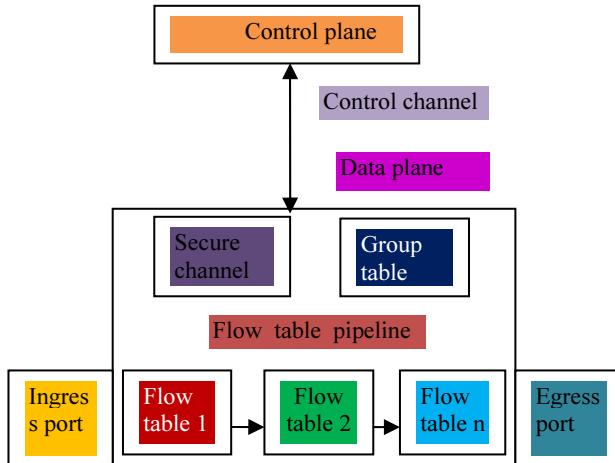
Figure 2. Architecture of OpenFlow switch

Group table consists of set of group entries. Each entry in group has certain specific semantics dependent in group type which consists of collection of action buckets. Action bucket state the action, which are performed in a group [7]. The group will select one or more bucket for each packet. So SDN controller will manage all communications by using openflow.

5. Mininet Simulator

Large number of network devices implemented through a network is very costly and difficult to implement. To reduce these problems the strategies made to for purpose of structuring and mitigating these kind of network technologies is MININET [9]. It's a free open-source software that simulates software defined network which consists of devices and controllers. Mininet was created in Python language and provide its API for user interaction and capable of emulating different network elements. SDN network can be easily virtualized and tested by using mininet [11]. It utilizes virtualization for the purpose of simulating real network by decoupling of data forwarding plane form control plane in Mininet VM.

6. Characteristics of Mininet [4]:

- **Flexibility:** Software is capable of managing modern features and newly introduced topologies, using various programming languages and variety of system software.
- **Applicability:** Without changing source codes, implementations of Network conducted in real networks.

- **Interactivity:** Simulated network should be manageable and runnable in real time network.
- **Scalability:** The prototyping environment can be scaled to larger network on only a computer.
- **Realistic:** The simulation behaviors represent real network behavior with a high degree of assurance on application of the network, so that application remains usable without any modification.

Mininet simulator is possible to create software defined network by using network programmer in a simple manner.

7. Experimental Result.

In MININET, a single command is used to create the network using linear topology with three host such as h1, h2, and h3.

$ sudo mn --topo linear, 3

8. Basic commands with Mininet [9]

Only a single console is required to control and manage entire virtual network. The basic commands such as ping, pingall, pingallfull, dump, net and iperf.

i) Ping: It uses to check the connectivity between different hosts by using ICMP protocol.

ii) Pingall: Connectivity between all hosts and tells which hosts are connected to each other

iii) Pingallfull: It gives more detail about how the hosts are connected. And tells minimum, maximum and average time between two hosts in millisecond.

iv) Dump: It is used to display the IP address and process identification of the host.

v) Net: It is used to list out links available in network between interface, host and switch.

vi) Iperf: It is used for TCP connection and to test bandwidth between hosts.

Figure 3. Connectivity between h2 and h3

Figure 4. Min, Max and Average time between two hosts in ms.

Figure 5. Display IP address and their process.

Figure 6. Links between interface, host and switch.

Figure 7. Testing bandwidth between hosts.

Experimental result of basic commands is shown. In Figure.3 how the connectivity between host h2 and host h3 by transferring two packets is shown. Figure 4 shows the connectivity between all host such as h1, h2, and h3 and also display the minimum, maximum and average time between all hosts in millisecond. Figure 5 tells the IP address and process id of each host and OVSswitch, controller dumps the details of all the nodes such as host list, switch list and controllers in the topology. Figure 6 display the link exists between the interface (eth0, eth1, eth2, and eth3), switch (s1, s2, s3) and host (h1, h2, h3). Figure 7 checking bandwidth between h1 and h3 using TCP and shows the result as measure of speed of the network bandwidth and helps us to check the bandwidth speed from one host to another host.

9. Security issues of SDN

Virtual network in SDN is provided to improve confidentiality, integrity and availability with security [6]. Due to attacks and vulnerabilities network performance and efficiency can be reduced which in turn affects the security properties such as confidentiality, integrity and availability. Modifying the information and inserting unnecessary codes can take place just because of an unauthorized user access.
There are various threads in SDN with respect to the different layers [8].

- **Rule's insertion:** New rules are created and implemented in different domain which causes various conflicts.

- **Malicious code:** Insertion of various malicious code leads to damage of information and corruption in data.
- **Distributed Denial of Service Attack:** Network traffics will be increased by attacks at channel, controller and switches.
- **Attacks from application:** Illegal access to the protected data about the network.
- **Man in the middle attack:** Data will be transferred to host without using any intermediate devices such as switches and router, etc. So, anyone in the middle with a connection enabled device can intercept the protected information without privacy.

Some security features of Distributed Denial of Service (DDoS) and Intrusion prevention systems are loop elimination and storm attack identification can be developed in SDN to take care of the security [8].

A spanning tree is created which automatically reconfigures the security function in loop elimination.

Storm attacks can be identified in the network by using unnecessary transmission of broadcasting spoofed packet. One of the famous detection techniques used in intelligent networks for DDoS attacks are anomaly detection, which are a subset of intrusion detection systems.

10. Conclusion

Manageability of network devices by SDN has grown beyond the expectation. This paper discusses about the overall concept of SDN such as SDN architecture and OpenFlow architecture. In addition to that analysis of SDN and detailed requirement on each standard interface has been described. During this study, design of SDN can be done using Mininet simulator. Mininet simulator acts as a useful alternative to run SDN problem cases on emulated network. Virtual machines provide an easier way to provide configuration and topology change but with real machine it is difficult. Because of decoupling data plane from control plane SDN is vulnerable to more attacks. Because of the vulnerability, performance of SDN could be rigorously affected.

References

[1] K. Benzekki, A. El Fergougui, and A. ElbelrhitiElalaoui, "Software-defined networking (SDN): A survey," *Security and Communication Networks*, (2017), DOI: 10.1002/sec.1737

[2] Parneet Kaur, Manish Kumar &Abhinav Bhandari, "A review of detection approaches for distributed denial of service attacks", *Systems Science & Control Engineering: An Open Access Journal*, (2017), VOL. 5, 301–320

[3] D. B. Hoang and M. Pham, "On software-defined networking and the design of SDN controllers",*6th International Conference on the Network of the Future (NOF), Montreal, QC, Canada*, (2015), pp. 1-3, doi: 10.1109/NOF.2015.7333307.

[4] F. Keti and S. Askar, "Emulation of Software Defined Networks Using Mininet in Different Simulation Environments",*6th International Conference on Intelligent Systems, Modelling and Simulation, Kuala Lumpur, Malaysia*, (2015), pp. 205-210, doi: 10.1109/ISMS.2015.46.

[5] Karamjeet Kaur, Japinder Singh and Navtej Singh Ghumman "Mininet as Software Defined Networking Testing Platform", *International Conference on Communication, Computing & Systems* (ICCCS-2014)

[6] D. Dobrev and D. Avresky, "Comparison of SDN Controllers for Constructing Security Functions," *2019 IEEE 18th International Symposium on Network Computing and Applications (NCA)*, Cambridge, MA, USA, (2019), pp. 1-5, doi: 10.1109/NCA.2019.8935053.

[7] Sm Shamim, Shahadat Shisir, Ahosanul Hasan, Mehedi Hasan & Arafat Hossain "Performance Analysis pf Different OpenFlow based Controller Over Software Defined Networking", *Global Journal of Computer Science and Technology: C, Software & Data Engineering*, (2018),Volume 18 Issue 1 Version 1.0.

[8] Maham Iqbal, Farwa Iqbal, Fatima Mohsin, Dr.MuhammadRizwan, Dr.Fahad Ahmad "Security Issues in Software Defined Networking (SDN): Risks, Challenges and Potential Solutions", *(IJACSA) International Journal of Advanced Computer Science and Applications*, (2019),Vol. 10, No. 10.

[9] CasimerDeCusatis, Aparicio Carranza, and Jean Delgado-Caceres "Modeling Software Defined Networks using Mininet", *Proceedings of the 2nd International Conference on Computer and Information Science and Technology (CIST'16) Ottawa, Canada,* (2016) Paper No. 133.

[10] M. Shin, K. Nam and H. Kim, "Software-defined networking (SDN): A reference architecture and open APIs," *2012 International Conference on ICT Convergence (ICTC)*, Jeju, Korea (South), (2012), pp. 360-361, doi: 10.1109/ICTC.2012.6386859.

[11] Ligia Rodrigues Prete, A. A. Shinoda, C. M. Schweitzer and R. L. S. de Oliveira, "Simulation in an SDN network scenario using the POX Controller," *2014 IEEE Colombian Conference on Communications and Computing (COLCOM), Bogota, Colombia*, (2014), pp. 1-6, doi: 10.1109/ColComCon.2014.6860403.

Analysis and Design of Advance Scalable QoS Based Resource Provisioning Framework

Pallavi Shelke [a,1], Dr. Rekha Shahapurkar [b]
aOriental University, Indore,Madhya Pradesh, India
bSt Vincent Pallotti College of Engineering & Technology, Nagpur, Maharashtra, India

Abstract. In today's growing cloud world, where users are continuously demanding a large number of services or resources at the same time, cloud providers aim to meet their needs while maintaining service quality, an ideal QoS-based resource provisioning is required. In the consideration of the quality-of-service parameters, it is essential to place a greater emphasis on the scalability attribute, which aids in the design of complex resource provisioning frameworks. This study aims to determine how much work is done in light of scalability as the most important QoS attribute. We first conducted a detailed survey on similar QoS-based resource provisioning proposed frameworks/techniques in this article, which discusses QoS parameters with increasingly growing cloud usage expectations. Second, this paper focuses on scalability as the main QOS characteristic, with types, issues, review questions and research gaps discussed in detail, revealing that less work has been performed thus far. We will try to address scalability and resource provisioning problems with our proposed advance scalable QoS-based resource provisioning framework by integrating new modules resource scheduler, load balancer, resource tracker, and cloud user budget tracker in the resource provisioning process. Cloud providers can easily achieve scalability of resources while performing resource provisioning by integrating the working specialty of these sub modules.

Keywords. Quality of Service, Scalability, Resource Provisioning, QoS parameters, Cloud Computing

1. Introduction

Now a day's Cloud computing is a technology evolving in developments and it is a very important part of today's life. Cloud provides us mobility, virtualization, it is easy to maintain from home, reliability, and service provisioning as per our demand most organization and businesses switch to the cloud. The cloud environment is made up of collective resources to provide services to its user over the internet. [1] Three basic layer structures are used in a cloud environment, named Software as Service (SaaS), Platform as Service (PasS), and Infrastructure as a service (IaaS). It will easily possible on SaaS environment, to access type of software like educational and business software which is present in cloud and user can access it through the internet on SaaS Platform. As cloud users, we can design various software-based programs and services through hardware, storage, also willing to do certain computation as well as hosting in that case also IaaS platform help us.

[1]Pallavi Shelke PG scholar , Department of CSE, India
Email: pwankhede@stvincentngp.edu.in.

Generally, most of the cloud user demands about hardware parts as its to buy it personally is very expensive. IaaS platform provides virtualized part of underlying cloud hardware. In the cloud Environment 70% of people make use of the IaaS platform and 30% people used PasS and SaaS [2]. In the cloud environment, everyday cloud infrastructure providers and cloud users face big problems in terms of resource management. Generally, management of various types of resources is done through Resource provisioning and Resource scheduling mechanism. Resource Provisioning Mechanism helps cloud providers to find out the best resources within the required time to their user or client, whereas resource scheduling helps to schedule the resources, map them to the workload then perform real execution. Resource scheduling is always performed after the resource provisioning. State of art of network provisioning strategies fails to reconcile cloud user and cloud service vendor benefits. It is the very important automatic identification of suitable resources as per customer 's request because it directly affects service response time and cost. To overcome this issue, successful design of resource provisioning framework is very important in cloud computing environment [2] This paperwork mainly contributes towards Quality of service (QOS) in the resource provisioning process, work completed by researchers in same field discussed in detail. In this paper mainly we target scalability as the main QoS parameter.

The paper is structured as follows: In Section 2, background and related work highlighting the importance of the Resource Provisioning process in cloud environment, analysis existing resource provisioning proposed models along with scalability as the main QoS parameter, its motivation, importance in resource provisioning is discussed. Literature survey on QoS-based resource provisioning presented in Section 3. In section 4 under discussion, we have identified shortcomings and challenges while reviewing earlier research paperwork. In Section 5, the design of the proposed model is presented. Challenges and future work are discussed in section 6. Lastly, section 7 is based on the conclusion.

2. Background and Related Work

Quality of Service (QoS) plays a crucial role in the cloud computing world, and researchers are working hard to design their proposed model from the perspective of both cloud users and cloud providers to reach deadlines, perfect execution times, and budget constraints.
User Perspective: cloud user wants a variety of services and resources as per their time with different expectations [3]
Provider Perspective: cloud provider wants to host many clients as per at a time to used resources available with them [3].
QoS Perspective: service often plays an important role between user and provider, where the assessment of the overall performance of service matters. Adequate QoS cannot be provided to cloud users until provisioning is made key functionality of resource offered. Therefore, to provide effective resources QoS based resource provisioning strategy is required [4].

2.1. Emphasizing the Importance of The Resource Provisioning Process in Cloud Environment

Resource management is an umbrella practice that covers all cloud resource characteristics and utilization.[4] So under resource management, we have a resource provisioning process as the first step provision multiple resources by checking availability type resources available in the resource pool. In the cloud Resource Provisioning cloud user, cloud resource provider, Workload Resource manager play a major important role. In a cloud environment, the workload resource manager maintains Resource Description, QoS Metric, SLA Measure Under cloud resource manager resource provisioned work. The resource manager's main responsibility is to mapped resources to workload based on the Quality-of-service requirement of the user. At the same time, many users demand different types of resources and services. To manage the bulk of workload firstly workload queue is maintained in the cloud environment, to serve user requests as per priority. One's a user request for resource his /her request is shifted to the workload queue. As multiple type resource requests are there from cloud users, it shifted to workload analyzer where workload cluster is maintained where the type of the same request is the club to gather in one block. Then cloud user request checks for QoS measure and SLA measure. once this process is completed resource provider shifts the user request to the resource scheduler. Then Resource scheduler identified and detects the type of resource available in Resource pool. Once this step is complete required resource is shifted back Cloud resource provisioner. then there will rechecking of QoS and SLA measure before sending the required resource to cloud user by Cloud workload manger [4]. With a basic understanding of resource provisioning process, we have noted the following key point

- Identification of enough resource available in the Resource pool
- Provision Resource as per cloud user fluctuating demand
- Classification and clustering of different type of Resources and workload
- Nourishment of Quality of Service and required resources at a very high level
- Maintain a high-quality agreement at the service level
- Reduce Waiting time of cloud user requests.
- Serve multiple requests at the same time during peak demand by cloud users.

These key points will help in designing an advanced Resource Provisioning Framework

3. Literature Review

The following table shows a comparative analysis of the work done from 2016 to 2020 and QOS parameters targeted till year 2020.

Table 1: Comparative Analysis of work addressed by various researchers in field of QOS based Resource Provisioning from year 2016 to 2020.

Author Name, publication Year	Techniques /framework /algorithm used	Distinguished work completed	Findings	QOS parameters target
Hala Haasan et al /year- 2020 [5]	QOS based Trust Model	i). Trust value dynamically updated at reach transaction ii) fake users will be identified on basis of the covariance mathematical Technique.	type of environment suitable for model not clearly mentioned.	Performance, cost, security, capacity, network
Aishwariya Chakraborty et al /year -2020 [6]	SensOrch: QoS-aware resource orchestra- tion scheme	1. Maintain high QOS and portability of sensor as a service 2.with threshold value enhance the performance of sensor- cloud and provide higher network-life time 3. Fair revenue distribution among sensor owner is possible	sensor cloud interaction as well as cost aspect missing	Resource Utilization
Shivangi Dhariwal et al/ year-2020 [7]	Resource optimization, Profit maximization technique	Maximizing the profit of the cloud server with a limited amount of resources. Compute size and speed of the server	worked only on assumed value. real time value needs to be considered to improve server-side performance	cost
A. Meenakshi et al /Year -2019 [8]	k-means clustering, gray wolf optimization partitioning technique.GSO-AGA algorithm	Reduce the high load on the server Allocates resources with the least amount of waste and provides the maximum benefit. minimum memory storage and minimum time will be possible with this algorithm	not clearly monitions which QOS parameter are improving with proposed model	cost, execution time
Sukhpal Singh Gill et al /year- 2018 [9]	SCOOTER framework	1.Optimized QoS parameters 2 Managing the resources automatically	Framewor k work only for fixed requirement or old resource requirement.	cost, resource Utilization, execution time
PVinothiyalaksh mi et al /year - 2017 [10]	E-MCA Technique	Opens the way for the most efficient workload-resource pair to analyze workload clearly and distinctly by applying auction technique and	Need to check QoS attributes in cost and time by integrating proposed model in a real cloud	cost, Resource availability, response time (performance)

		QoS-based clustering.	environment, need to implement in a real-world cloud environment	
Xiaoyong Xu,et al/ year -2017 [11]	event-driven resource provisioning framework	1.It detects environmental changes in the form of Events, and agreements with events to reduce the expense of recruiting VMs 2. Reduce running cost of large computations	Real-time execution of proposed work required	Cost, scale up and scale down algorithm used
Tao Chen et al /year- 2017 [12]	online QoS model approach, used hybrid dual-learners technique	1. Produces better overall accuracy while having acceptable overhead 2. eliminates the need for heavy human intervention, which can be complex and error-prone.	To get the best result need to test & try the same proposed model on the new application	Response time
Himadri Shekhar Mondal et al/ year-2017 [13]	Fuzzy logic concepts	Improving QOS by Balancing the load with help if-then rule of fuzzy logic	This model can be improved by adding costing, reliability, more fruitful results Processing rate, etc. using the logic of Fuzzy.	Speed of processor, response time
Jolly Upadhyaya et al/year -2017 [14]	QOS Innovative model	Defined QOS parameters in terms of user oint of view help the cloud provider to maintain quality of service and also prepare their service. user and provider know in advance QOS variable and parameter	Proposed model only design for the education sector, QoS parameter or variable targeted by innovative model not mention.	Response time
Xianrong Zheng1, et al/ year- 2017 [15]	Spearman coefficient approach	Used to predict both QoS ratings and rankings for cloud services. To help the cloud Providers to improve their brand and consumers to the cloud Identify services that satisfy the requirements of their QoS	Provide less accurate rating.	QoS-based cloud service recommendati on, not targeted specific parameter

Tamal Adhikary et al /year -2016 [16]	QoS aware VM provisioning, VCA based resource provisioning method,	Decreases energy consumption, helps to increase resource usage across a range of multimedia social applications.	Type of SLA requirement violated is not mentioned in this paper	response time on average time as QOS parameter
Sukhpal Singh/- year 2016 [17]	QoS-based resource provisioning and scheduling framework (QRPS)	work in both environments, Homogenous and heterogeneous	Not focus on energy-saving parameters. provisioning must be done on maintaining the SLA as per user requirement	Energy, cost and execution time

According to our comparative study of resource-based QoS processes, it was found that a majority of work performed by researchers on QoS parameters is only targets cost, response time, execution time, and resource utilization. Also, after studying various proposed models, we found that scalability is the most important parameter to consider, as it will indirectly fulfill fundamental QOS parameters such as cost, response time, and resource utilization. Still, there is a lack of automating scalable real-time framework which suits all cloud computing platform. Few automated scalable frameworks available still facing a lot of issues which are discussed in the next paragraph.

4. Discussion

Since after working more on scalability, we discovered that majority of research work is often centered on the auto scalability concept and we have noted the following reason of less work on scalability and its issues.

4.1 Issues on Scalability Based Resource Provisioning Process

Researchers have come up with a novel concept of auto-scaling approach to improve resource provisioning work with scalability, where research paper [18], highlighted certain unique challenges like, a monitoring tool and an auto-scaling mechanism is required that meets quality of service standards and is also compatible with all deployment models at the SaaS and PaaS level also aunto scaling should have a high level of fault tolerance

4.2 Review Questions for Handling Scalability Issue

We have identified following questions based on above listed issues

- What will be the computation rate?

- How fast cloud user will get service benefits
- How the storage process will happen?
- How often do things /requirements change?
- What will be the success boundary with scalability?

4.3 Examples Based on Work Completed On Scalability Based Approaches /Frameworks with main findings

A lightweight approach is proposed in paper [19], that gives the fine-grained scaling at resource level itself which help to meet QoS requirements which will work efficiently over traditional VM-level cloud service scaling that overuse resources while still rising cloud provider operating costs. Finding: However, in this paper we found that more research is needed to establish how resources might be planned among applications with varying QoS requirements. A New auto-scaling mechanism was described in the paper [20], where auto-scaling method efficiently completes all jobs inside the user-defined deadlines .it helps to minimize costs for different workload patterns. The main findings of this paper are that workload forecasts will produce better results. This technique assumes a cloud customer with an unlimited budget, but it's better to consider it from the perspective of a user with a limited budget. Author presented an automatic resource provisioning approach in paper [21], for auto-scaling resources based on reinforcement learning with Marko decision process ((MDP). Here author tried to reduce the SLA violation increase stability but they do not focus on the type of services only work on fixed services but the status of the workspace will be altered based on its use not provided by the proposed approach, with the prediction of incoming load it can be possible with reinforcement learning but that not completed in the proposed approach. To address the issue of vendor lock, the author [22] described a model-driven approach for connecting a cloud platform-independent service model with cloud-specific operations. To provide auto-scaling deployment across clouds, the author makes use of cloud management tools and demonstrated data of different applications on multiple clouds. As a future work author would like to use machine-learning techniques to expand the model's predictive and constructive auto-scaling methods.

In reference to the above-mentioned research papers study, we have identified the following major challenges, which we will attempt to address through the proposed framework.

1. Still there is a lack of automating scalable real-time framework which suits all cloud computing platform.
2. Cost will impact scalability as per the on-time demand of cloud users.
3. To scale up and scale down resources as per resource availability in the resource pool automated resource pool tracker is required before start provisioning and scheduling the resources which save cloud user waiting time and cloud user provider's resource allocation time
4. There should be no delay in servicing resource requests as there is a heavy demand from cloud users.
5. There is a need for resource scheduling with various QoS requirements that must be met in one place

6. The framework should still take into account users with limited budgets also class of cloud users.
7. In the design context, the incoming user request should always be forecast

5. Proposed Model

Based on an existing research study on the above-mentioned resource provisioning proposed model, comparative analysis of QoS parameters, scalability issues, and auto scalability based completed research work, we proposed a scalable QoS based advanced resource provisioning framework, as shown in Figure 1, through which we will aim to overcome the challenges listed above.
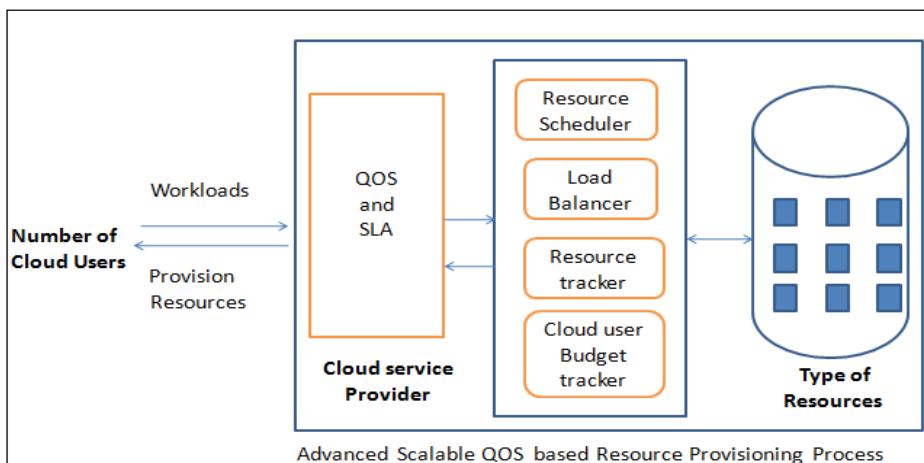
Figure 1. Advanced Scalable QoS based Resource Provisioning Framework

The following sub-module will play a key role in the resource provisioning phase, assisting in the reduction of the difficulties and limitations addressed in the previous sections.

- **Resource Scheduler**: can constantly coordinate workload or application requests coming from cloud users. It can also aid in coordinating demands depending on the availability of resources in the resource pool.
- **Load Balancer**: A load balancer can assist the cloud provider in balancing customer workload according to priority which will reduce massive request traffic
- **Resource Tracker**: A resource tracker can assist with keeping track of resource waste if any, as well as searching for over-provisioning and under-provisioning of resources regularly.
- **Cloud User Budget Tracker**: The budgetary requirements will be taken into consideration by Budget Tracker.

We would aim to satisfy the needs of cloud users and cloud resource providers with this proposed framework by integrating different QoS criteria as well as SLAs (service Level agreement) in one place

6. Challenges and Future Work

The demand for global computing services is rising day by day in today's world. The cloud service provider must get ready with sufficient resources with QoS requirements to handle several cloud user requests also will get benefits when they start increased resources. But what if the resources are limited? How to deal with limited resources. Even though there was a considerable amount of research completed to develop a dynamic and self-managed cloud system, still there is a lack of adequate QoS-based resource provisioning approach required. In this paperwork, we have studied research work completed by researchers which mainly focused on scalable QoS parameters in resource provisioning. Guaranteeing QoS requirements is a need of today's resource provisioning approach. In future work, we will implement the above scalable QoS based resource provisioning framework which will overcome the following challenges and in our next research paper we will present the implementation and results of the above-proposed model

- As cloud user resource requirement or service requirement or workload changes
- on time therefore an automatic provision framework is required at the cloud provider side so that he can self-configure the available resource.
- Resource Provisioning System needed at the user side also just to fulfil their QoS expansion needs less pay.
- Flexible approach is required handling workloads that are rising repeatedly
- Scale-up and scale-out strategies are required to handle massive user request traffic.

7. Conclusion

In this Paperwork, we have discussed the importance of QoS-based resource provisioning and classified existing approaches proposed by researchers who worked on the same subject. Existing research work papers help us identify resource provisioning with many QOS-based parameters. Many researchers target one or more important QoS parameters, but after surveyed we have found that very few of them focus on scalability as QOS's main parameter/attribute. We have also included information on scalability types, the need for scalability, and analysis questions for dealing with scalability issues. Our paper concentrated on the scalability parameter, which is critical for both cloud providers and cloud users in terms of resource management and utilization. As future work in our research paper we will the implement proposed work which is based advanced scalable QoS-based resource provisioning framework that assists in overcoming the above difficulties while also attempting to close the study gaps.

References

[1] Ahuja S. P, Deval, N. , "On the performance evaluation of IAAS cloud services with system-level benchmarks", *International Journal of Cloud Applications and Computing*, pp80–96, 2018.

[2] Durga S, Mohan S, Dinesh Peter J,Martina Rebecca Nittala. "Towards Benefiting Both Cloud Users and Service Providers   Through Resource Provisioning ",International Journal of Information Technologies and Systems Approach ,volume 12 ,2019.

[3] M.H. Ghahraman, Mengchu Zhou,Chi Tin Hon, "Towards Cloud Computing Qos Architecture :Analysis of Cloud System And Cloud Services", IEEE/Caa Journal Of Automatica Sinica, Vol. 4, No. 1,2018.

[4] Sukhpal Singh,Inderveer Chana,(2016).Cloud resource provisioning: survey, status and future research directions. Springer- Verlag London ,2016.

[5] Hala Hassan, Ali Ibrahim El-Desouky, Abdelhameed Ibrahim,El-Sayed M. El-Kenawy."Enhanced QoS-Based Model for Trust Assessment in Cloud Computing Environment ",IEEE Access ,2020

[6] Aishwariya Chakraborty, Sudip Misra, Ayan Mondal, and Mohammad S. Obaidat.,"SensOrch: QoS-Aware Resource Orchestration for ProvisioningProvisioning Sensors-as-a Service",978-1-7281-5089-5/20/$31.00 ©IEEE ,2020 .

[7] Shivangi Dhariwal, Dr. Hemant Makwana, "Maximize the Cloud Profit to Improved QoS inCloud Computing: Design and Analysis", International Conference on Inventive Systems and Control ,IEEE Xplore Part Number: CFP19J06-ART; ISBN: 978-1-5386-3950-4,2019

[8] A.Meenakshi, H. Sirmathi, J. Anitha Ruth1, "Cloud computing-based resource provisioning using k- meansclustering and GWO prioritization. J. Soft Computing, Springer-Verlag GmbH Germany, part of Springer Nature,2019.

[9] Sukhpal Singh Gill , Rajkumar Buyya. "Resource Provisioning Based Scheduling Framework for Execution of Heterogeneous and Clustered Workloads in Clouds: from Fundamental to Autonomic Offering", J Grid Computing ,https://doi.org/10.1007/s10723-017-9424-0,spinger ,2018

[10] P.Vinothiyalakshmi, R Anitha, "Efficient Dynamic Resource Provisioning using Auction based Technique with Improved QoS in Cloud Environment" ,ICCIDS, 978-1-5090-5595- ©IEEE,2017

[11] Xiaoyong Xu, Maolin Tang and Yu-Chu Tian , "QoS-guaranteed Resource Provisioning for Cloud-base MapReduce in Dynamical Environments. Future Generation Computer Systems, 78(Part 1), pp. 18-30,2017

[12] Tao Chen, Rami Bahsoon, "Self-Adaptive and Online QoS Modeling for Cloud-Based Software Service. IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 43, NO. 5,2017

[13] Himadri Shekhar Mondal, Md. Tariq Hasan, Taposh Kumar Karmokar, Shamlendu Sarker, "Improving Quality of Service in cloud computing architecture using Fuzzy Logic", ICAEE, 28-30 September, Dhaka, Bangladesh, IEEE,2017

[14] Jolly Upadhyaya , Dr. Neelu Jyoti Ahuja, "*Quality of Service in Cloud Computing in Higher Education: A Critical Survey and Innovative Model",*(I-SMAC).IEEE,2017

[15] Xianrong Zheng,Li Da Xu1,Sheng Chai . "QoS Recommendation in Cloud Services. *Digital Object Identifier 10.1109/ACCESS.2017.2695657.IEEE Access,2017*

[16] Tamal Adhikary,Amit Kumar Das, Md. Abdur Razzaque, "Majed Alrubaian Mohammad Mehedi Hassan, Atif Alamri , "Quality of service aware cloud resource provisioning for social multimedia services and applications", Spinger ,2016 .

[17] Sukhpal Singh, Inderveer Chana,"Resource provisioning and scheduling in clouds:QoS perspective" , J Supercomputer , 72:926–960 DOI 10.1007/s11227-016-1626-x, Spinger,2016.

[18] Hanieh Alipour, Abdelwahab Hamou-Lhadj, Xiaoyan Liu, "Analyzing Auto-scaling Issues in Cloud Environments. CASCON 24th Annual International conference on computer Science and software Engineering,2014

[19] R. Han, L. Guo, M. M. Ghanem, and Y. Guo, "Lightweight resource scaling for cloud applications In *Cluster, Cloud and Grid Computing (CCGrid)",12th IEEE/ACM International Symposium .* pp. 644–651,2012

[20] Ming Mao, Marty Humphrey, "Auto-Scaling to Minimize Cost and Meet Application Deadlines in Cloud Workflows", SC11 November 12-18,Seattle, Washington, USA Copyright 2011 ACM 978-1-4503-0771-0/11/11,2011

[21] Bahar Asgari , Mostafa Ghobaei-Arani , Sam Jabbehdari,"An Efficient Approach for Resource Auto-Scaling in Cloud Environment",International Journal of Electrical and Computer Engineering (IJECE)**,** Vol. 6, No. 5, pp. 2415~2424 ISSN: 2088-8708,2016

[22] Hanieh Alipour ,Yan Liu,"Model Driven Deployment of Auto-scaling Services on Multiple Clouds. IEEE International Conference on Software Architecture Companion, 978-1-5386-6585-5/18 ,© IEEE,2018.

MISP: Model for IaaS Security and Privacy

INDRA KUMAR SAHU [a,1], MANISHA J NENE [a]

[a] Dept. of Computer Science & Engineering, Defence Institute of Advanced Technology, Pune, Maharashtra, India

Abstract. Paradigm shift towards cloud computing offers plethora of advantages both for cloud users and Cloud Service Provider (CSP). For cloud users, it offers saving of cost, scaling of resources, pay per use, elastic and on-demand services. On the other hand, it offers centralized resource management and provisioning of operations, safety and security for CSP. By holding multiple virtual IT resources (CPUs, storage servers, network components and software) over the internet, Infrastructure-as-a-Service (IaaS) serves as fundamental layer for all other delivery models. Along with benefits of IaaS, there exists several security and privacy issues and threats to confidentiality, integrity, authentication, access control and availability. In this paper, detailed study of IaaS components, associated security and privacy issues are explored and counter measures for the same are determined. Furthermore, as a result of the study, Model for IaaS Security and Privacy (MISP) is proposed. The model presents a cubical structure and adds more features than the existing models to enhance the security and privacy of data and operations and guide security assessment for safer adoption by enterprises.

Keywords. Cloud Computing, Cloud Security, Cloud Deployment Models, Service Level Agreement, Model for IaaS Security and Privacy (MISP), IaaS, Virtualization

1. Introduction

Since the inception in late 1960s, cloud computing became a ubiquitous technology with hardware, software, computational and operational IT resources and services delivered via Internet to the users [1]. Elasticity, scalability, on-demand resources, cheap operational expenses, location and device independence and pay per use business model are the merits for its prime attraction [2]. Cloud computing has provided huge opportunity to migrate from maintaining, securing and operating own standalone, on-premise resources like infrastructure and applications to cloud. Recently, it attracted very considerable attention of academicians, industry people and researchers. As highlighted by NIST [7], cloud computing has three service models and four deployment models.

1.1 Service Models

IaaS with resources like data storage servers, computing hardware and networking components provides infrastructures to users to facilitate management of OS and applications.

---

[1] Indra Kumar Sahu, MTech scholar, Department of CSE,
 Email: ikambition.s23@gmail.com.

Platform-as-a-Service (PaaS) where users are provided with an environment to develop, create and use their own tools and software applications.

Software-as-a-Service (SaaS) with readymade application software and tools are delivered to the users with licenses to use remotely without buying them completely.

1.2 Deployment Models

The deployment models define the way cloud may be used. The different models offer varying resources and the cloud users can adopt the one that suit them the best based on their needs and budgets. Four types of deployment models are as under.

Private cloud offers cloud resources and infrastructure to be used as stand-alone resources with greater control over security and data backup facility.

Public cloud offers shared resources at lower cost but the security and privacy of the data and storage lowered as compared to the private cloud.

Hybrid cloud shows the best of first two models in terms of resources, controls and the cost. The security and privacy are in between that of private and public cloud.

Community cloud offers shared resources amongst the same types of organizations like banks, hospitals etc.

1.3 IaaS Model

Cloud computing primarily depends upon IaaS delivery model that provides rudimentary operating systems, networking components, security infrastructure and data servers for designing and developing required applications, databases, development tools and services [8]. The Oracle and KPMG Cloud Threat Report 2020 [4] shows the recent adoption trends for cloud computing. As compared to 62% in 2018, in 2020 76% of on-premise business-critical applications migrated to IaaS through 'lift-and-shift' approach. Being the fastest growing sectors amongst all other service model, IaaS is expected to grow to $63 bn in 2021 from $ 23.6 bn in 2017 at a rate of 27.6 % according to Gartner [5]. It also predicted that by 2025, 80% of the enterprises will use IaaS as compared to 65% in 2017.

On demand services and scalable resources with advanced technical capabilities are provided to the users in IaaS model. Hardware comprising of storage servers, networking components and computing hardware (CPUs, RAM, graphic cards etc.) and software like cloud Application Program Interfaces (APIs), Utility Interfaces (UIs), hypervisors, software modules, security and control management modules are two types of components. Quality of Service (QoS) is an important factor and is made part of legal contract [5]. The IaaS model can also be viewed as shown in Figure 1 below [9].

From rigorous study on security and privacy issues of IaaS delivery, a Model for IaaS Security and Privacy (MISP) is proposed adding various IaaS components to mitigate the threats of the delivery model.

The rest of the paper is presented in four sections. Literature survey is in the second section. Third section is of preliminaries. The fourth section is for the proposed model MISP with details for enhancing the security and privacy in IaaS against existing vulnerabilities and threats. The fifth section concludes the paper along with future scope.
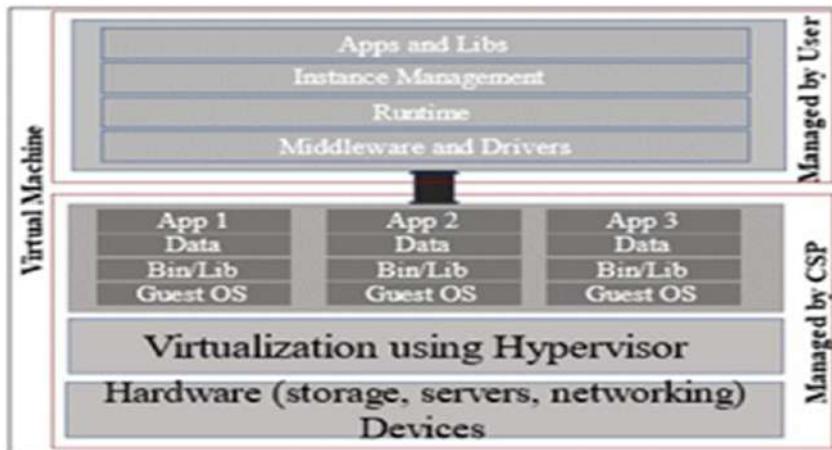
Figure 1. Virtual and Physical layer in IaaS [9]

2. Literature Survey

2.1 Related work

The security and privacy issues being faced in IaaS are related to Confidentiality, Integrity, Authentication, Availability and Access control (CIAAA) and a lot of research work has been done to mitigate these issues. Ravi et al. [10] carried out a sincere assessment of threats to security in IaaS along with responsibilities of cloud user and CSP. Their work mainly highlighted the issues in CIA triad and proposed possible solutions. The latest threats and focussed malicious approaches are not addressed.

Ahmed et al [9] presented brief of issues in IaaS components and analysed CSA top twelve threats in the model along with possible solutions for them. The threats mentioned in CSA report gets changed from time to time and hence are not very relevant at present time as per CSA report 2020 [4].

Cullum et al [35], in his paper presented host hypervisor security issues in public IaaS and their solutions. The detailed study on hypervisor gives out known attacks that exist in hypervisor shared environment. The solutions are focused mainly on virtualization related issues while other threats are not addressed.

Moutai et al [24], presented a secure architecture-based distributed testing to confirm CAA based on QoS. It is limited to information security. The parameters like security of storage, network and hardware are not tested.

Dawoud et al [8], presented IaaS security model with issues related to components, suggested secure policies along and restriction levels. The security model is limited to some issues only whereas, with the advancement in cyber spectrum, there is need of addition of latest issues and threats.

2.2.   Contributions

Our paper presents a comprehensive cubical MISP that comprises of components related issues along with associated threats to IaaS model; each in the common plane of cloud user and CSP. There are rules and policies to enhance the security and privacy of data and operations in second plane. The third plane of cubical presents levels of rules and policies for implementation varying from lenient level to strictest level. The model summarizes all threats and possible ways out to enhance the security and guides security assessment for safer adoption of IaaS delivery model.

3. Preliminaries

3.1 Service level agreement (SLA)

SLA is a legal document agreed and signed between CSP and a cloud user to describe the legal responsibilities, liabilities for both of them and define QoS offered by the CSP as part of the agreement [12]. It makes a mention of both the required and the expected level of services to be delivered maintaining availability and security and privacy with review or monitoring of the SLAs, riders and liquidation terms and time span of contract.

3.2 Virtualization of Platform

Virtualization is a process of abstracting and sharing a single hardware that facilitates aggregating multiple stand-alone computing resources like CPUs, memory, storage and network components [8]. A typical example is 'Server virtualization' in which several attributes of physical server is hidden and they are reproduced in a hypervisor in the form of virtual CPU (vCPU), Virtual RAM (vRAM), virtual NIC (vNIC) and virtual disks. It has two important characteristics namely, multi-tenancy and scalability. The virtual and physical layers in the model are illustrated in figure 2 below.

Two types of virtualizations namely OS based in which a software is installed in host OS and hardware based that refers the installation directly on the physical host hardware [14].
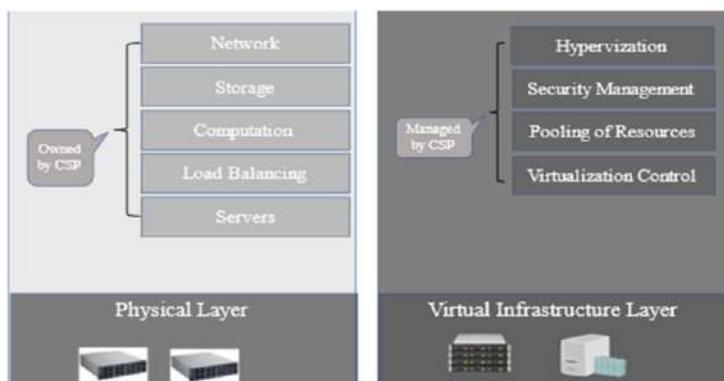
Figure 2. Virtual and Physical layers in IaaS

3.3 Utility Computing

Utility computing provides multiple resources on demand as per user's request. Various IT resources are packed for metered services and then provided to cloud users at low cost and as pay-per-usage basis with scalability support even if demand reaches to its peak [8].

3.4 Cloud Scalability

Cloud scalability being one of the basis of cloud computing, offers homogenous resources with infinite scalability at linear increase of performance; the answers to when, what and where to scale in multi-tier service-oriented applications in autonomic scaling [14].

4. Model for IaaS Security and Privacy (MISP)

Security of any service model in the cloud depends on the security of the infrastructure. Various components in IaaS are required to be looked into for user's satisfaction. Multiple agencies undertake works related to threat assessments on privacy and security on cloud computing. Distributed Management Task Force (DMTF), Open Cloud Consortium (OCC) and Cloud Security Alliance (CSA) are some of them that define standards, certifications and practices to ensure a secure cloud environment [18].

We propose a Model for IaaS Security and Privacy (MISP) in cubical form with three planes defined as shown in figure 3. The first plane gives out components of IaaS. The cloud user and CSP are common participants of the plane and they generally share responsibility in maintenance of security and privacy of the model.
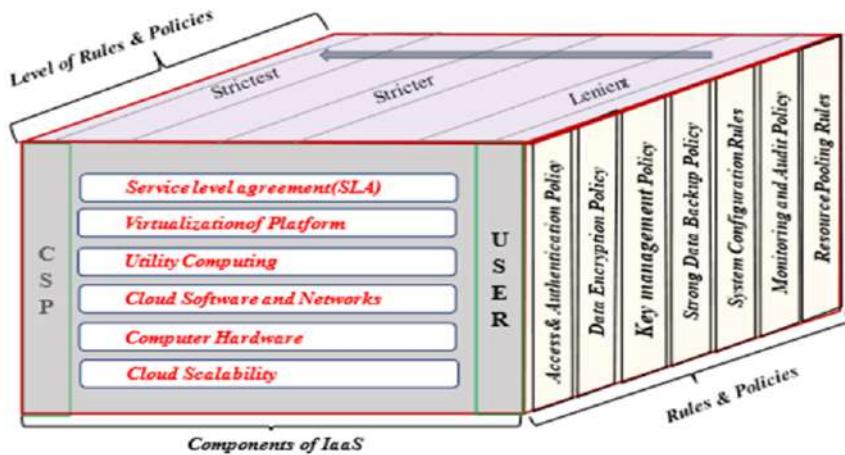
Figure 3. Model for IaaS Security and Privacy (MISP)

4.1 Threats Related to Components of IaaS

The plane consisting of components of IaaS in the proposed model is analyzed and threats associated with the solutions are described.

4.1.1 SLA Related Issues: Lack or non-existence of standardization in creating and performing the SLA between the involved parties creates big loopholes. The leading CSPs like Amazon (AWS), Google (GCP) and SalesForce hide numerous parameters regarding data safety and preservation in their proposed SLAs [6]. SLA may get exposed to vulnerabilities if any misunderstanding amongst the parties arises. So, it becomes imperative to detect user's concerns on priority [25]. The review and study of the environment displays several threats as per CSA classification.

Data breach and usage monitoring of data stored in the cloud is possible through human errors, application vulnerabilities, inadequate security practices or targeted attacks. Strong encryption techniques, prevention of leakage of secret data using neural networks [3], Data Leakage Prevention for Data in Transit using Artificial Intelligence and Encryption Techniques [34], strong backup and retention strategies and use of strong APIs [20] can mitigate this issue.

Insufficient due diligence while transferring responsibility of cloud control and cloud security to the CSP is a threat that is caused due to lack of transparency in security mechanism applied by CSPs [19]. Un-sanctioned application usage and sanctioned application misuse are the two key challenges in this threat. Strong key management [17], and use of SLA for cloud visibility are solutions for this issue.

Denial of Service (DoS) is a threat mainly due to external agents that can cause unavailability of resources to the cloud users in the form of network, application or bandwidth denial [18]. The threat can be mitigated with regular audits of log and monitoring of services with advanced methods like Software Defined Networks (SDN), EDoS and SEDoS [30].

Many of the researchers argue to propose Web Service Level Agreement (WSLA) that can manage SLAs in IaaS environment [36]. More conveniently, SLA monitoring and enforcement may be delegated to a third party to bridge the trust deficiency between the CSPs and the cloud users.

4.1.2. Virtualization Related Issues: Virtual-aware security is required to face the security issues in IaaS [15]. Three types of possible threats are determined here.

Threats from host Operating System*:* The host OS being privileged domain can monitor, configure, communicate and modify data or services and hence may cause threats to IaaS model. According to MacAfee Cloud Adoption and Risk Report [21], the average organization has 14 misconfigured IaaS instances at any given time making 2269 instances per month. 5.5% of AWS S3buckets in use are misconfigured. Strong data backup and retention techniques [22] and multi factor authentication can mitigate the threats.

Communication between host and the VM is through virtual network or shared virtual resources and hence vulnerable to threats. An attacker could exploit important features like Clipboard to monitor the activities between them [25]. In case of host being compromised, all the VMs get into risk of any kind of possible attacks. Trusted Virtual Domain (TVD) for infrastructure and security mechanism [29], Trusted Cloud Computing Platforms (TCCP) for confidentiality [31], VLAN for network virtualization and Identity Based Integrity Verification (IBIV) protocol for data integrity [13] are the solutions for such threats and issues.

Threats from VMs hosted on the same host: CSP provides API to carryout management functions such as provisioning, replication and decommissioning of resources on IaaS. But these insecure ill-designed, broken, exposed or hacked APIs and user interfaces (UIs) may lead to data breach or other security threats. Data Leakage

Prevention for Data in Transit using Artificial Intelligence and Encryption Techniques [34], Scarce Attack Datasets and Experimental Dataset Generation [27], multi-factor authentication and robust authentication mechanism [33] can mitigate these issues.

Other possible attack on virtualization platform is VM Escape in which isolation layer between host and VM is broken to get the access of hypervisor's root privileges. As the attacker gets control over the host OS, he can use the compromised OS to manipulate control as per his desires through covert channel for malicious code execution.

Network virtualization partitions or aggregates a collection of network resources and present them a unique and isolated physical view to the users. Communication between VMs is through network virtualization in a direct and efficient manner. To avoid attacks like sniffing, SQL injection and spoofing on virtual network, secure physical channels can be adopted.

4.1.3. Utility Computing Related Issues: The utility computing faces the challenge of complexity in cloud computing. A bigger CSP may lease the services to second level CSP who in turn provides metered service to users. For example, Amazon DevPay5 from Amazon is a second level CSP. In this, the second level CSP might use services and user may be charged for what he has not used. Strong multi-tier passwords and two-factor authentication mechanisms [23] maybe used to mitigate this issue.

4.1.4. Cloud Software and Network Related Issues: In IaaS model, CSP provides cloud software and networks. Open-source cloud software like Eucalyptus and commercially proprietary software are two options but security from vulnerabilities and bugs cannot be ensured in either of the two. Cloud providers either furnish APIs or web service protocols like XML Simple Object Access Protocol or simply SOAP to grant access to cloud users to orchestrate management functions.

4.1.5. Computing Hardware Related Issues: A pool of shared distributed physical resources is provided to cloud users through virtualization in IaaS. Threats and attacks in on-premises hardware scenario occurs internally as a study shows it to be 70% [16]. Threats can be categorized in various ways. Based on type of resources: threats to physical computing resources like CPU, monitor, other physical machines and threat to storage resources where attacker gets access of the data storage.

The other one is based on type of adversary: insider and outsider attackers. Insider attackers have access to the resources of the organization and can cause damage intentionally or otherwise [26]. The outsider may be any hacker or bot to damage the system. Policy Enforcement Points (PEPs) side caching [28], inclusion of human resource management are some of the mitigation techniques.

Management of various changes in internal, system practices and Identity and Access Management (IAM) affects identity, credentials, key and access management. Strong end to end encryption, multi-tier passwords and multi factor authentication, and LDPC decoders [11] are measures to mitigate it.

4.1.6. Cloud Scalability Related Issues: IaaS resources can be scaled as per the user requirements. While doing so, there is a threat of account hijacking and abuses to breach infrastructure through spam mails, social engineering, vishing and phishing. Strong encryption techniques, multi-factor authentication [23] for integrity and strict monitoring of unauthorized activities may help to tackle this issue.

4.2. Rules and Policies for IaaS

The rules and policies for security and privacy are presented in the second plane in a vertical axis that implicates their presence through all components of IaaS. They are as mentioned below.

1. *Access and Authentication Policy:* to restrict any unwanted and unwarranted users to get access and verify the authorized users of the IaaS delivery model.
2. *Data Encryption Policy*: to ensure confidentiality, integrity and authentication in IaaS model using strong encryption techniques.
3. *Key Management Policy:* to enforce no loss and misuse of keys used in the IaaS for various purposes.
4. *Strong Data Backup Policy:* to avoid loss, deletion, tampering or theft of data in event of any unprecedented natural disaster, data corruption or cyber-attack.
5. *System Configuration Rules:* to avoid system misconfiguration, system bugs and internal or external attacks through exploitation.
6. *Monitoring and Auditing Policy*: to prevent any intrusion, system failure, status of software, untoward event and possible security breaches.
7. *Resource Pooling Rules:* To utilize the resources available with CSP for users as per demand optimally and judiciously.

4.3 Levels of Rules and Policies

The third plane is level of rules and policies. The level of rules and policies implementation need to be based on judicious scrutiny of security of data and operation on IaaS infrastructure, expertise of the user and the environment. If the data and operation are of critical in nature, the strictest level to be followed. In case of normal or low value data and operations, lenient level may be implemented. Since the strictest level might be slow and time consuming, the levels may be decided accordingly.

The proposed model is an attempt to standardise the IaaS layers, various components present in the model that are threatened and rules and policies to mitigate the threats, issues and challenges. Level of rules and policies implementation suggest degree to enhance the privacy and security accurately traded off between operational time and required security.

5. Conclusion and Future Scope

IaaS delivery model provides the basis for all other models and faces the security issues across hardware and software. Virtualization is core of the IaaS model for isolation. The security and privacy issues arise due to numerous reasons like lack of adequate knowledge, complex policies, technical glitches, system errors, standardization, certification and violation of established policies and practices. In this paper, security issues associated with IaaS components are investigated. The security issues related to security of each IaaS components and proposed countermeasures are provided. The proposed MISP summarises all the issues and possible ways out to secure IaaS model to enhance the security and guide security assessment for safer adoption by enterprises. Cryptography and the best available techniques-based solutions are proposed to mitigate the threats to manage and secure the cloud in an optimal manner

Due to phenomenal rise in computing capabilities, the existing issues and challenges may get aggravated to unimaginable level of difficulties. New technologies like Network as a service (NaaS), Cloud of Things (CoT) etc. may pose different challenges. Timely review of the issues with the changes in policies and procedures will be warranted.

Another imminent threat is from quantum computing that possess extremely high computing capabilities. So, the security and privacy concerns of IaaS are required to be seen in the prism of quantum threats. The future work may be carried out to find quantum solutions in cloud computing for the post quantum era.

References

[1] Madni, S. H. H., Abd Latiff, M. S., & Coulibaly, Y. (2016). Resource scheduling for infrastructure as a service (IaaS) in cloud computing: Challenges and opportunities. Journal of Network and Computer Applications, 68, 173-200.

[2] Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. Journal of network and computer applications, 34(1), 1-11.

[3] Ghouse, M., & Nene, M. J. (2020, June). Graph Neural Networks for Prevention of Leakage of Secret Data. In 2020 5th International Conference on Communication and Electronics Systems (ICCES) (pp. 994-999). IEEE.

[4] Oracle and KPMG "Cloud Threat Report 2020", https://www.oracle.com/in/a/ocom/docs/cloud/oracle-cloud-threat-report-2020.pdf, assessed on 17 Dec 2020.

[5] Gartner, Infrastructure-as-a-Service Adoption and Risk Report. [Online] Available:https://www.gartner.com/en/newsroom/press-releases, accessed on 20 Mar 2021.

[6] Modi, C., Patel, D., Borisaniya, B., Patel, A., & Rajarajan, M. (2012). A survey on security issues and solutions at dif ferent layers of Cloud computing, 1-32.

[7] Mell, P., & Grance, T. (2011). The NIST definition of cloud computing.

[8] Dawoud, W., Takouna, I., & Meinel, C. (2010, March). Infrastructure as a service security: Challenges and solutions. In 2010 the 7th International Conference on Informatics and Systems (INFOS) (pp. 1-8). IEEE.

[9] Ahmed, A., & Zakariae, T. (2018). IaaS cloud model security issues on behalf cloud provider and user security behaviors. *Procedia computer science*, *134*, 328-333.

[10] Kumar, P. R., Raj, P. H., & Jelciana, P. (2018). Exploring data security issues and solutions in cloud computing. Procedia Computer Science, 125, 691-697.

[11] Rao, R. G., & Nene, M. J. (2019, March). CNR: A Technique for Data Replication Organization in BigData. In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) (pp. 1-6). IEEE.

[12] Remesh Babu, K. R., & Samuel, P. (2019). Service-level agreement–aware scheduling and load balancing of tasks in cloud. Software: Practice and Experience, 49(6), 995-1012.

[13] Sahu, I. K., & Nene, M. J. (2021, February). Identity-Based Integrity Verification (IBIV) Protocol for Cloud Data Storage. In 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) (pp. 1-6). IEEE.

[14] Lloyd, W., Pallickara, S., David, O., Arabi, M., & Rojas, K. (2014, March). Dynamic scaling for service oriented applications: implications of virtual machine placement on IaaS clouds. In 2014 IEEE International Conference on Cloud Engineering (pp. 271-276). IEEE.

[15] Amato, F., Moscato, F., Moscato, V., & Colace, F. (2018). Improving security in cloud by formal modeling of IaaS resources. Future Generation Computer Systems, 87, 754-764.

[16] Markatos, E. (2008). Large Scale Attacks on the Internet Lessons learned from the LOBSTER. Crete, Greece.[Online]. Available: http://www. ist-lobster. org/publications/presentations/markatosattacks. pdf.

[17] Seitz, L., Pierson, J. M., & Brunie, L. (2003, October). Key management for encrypted data storage in distributed systems. In Second IEEE International Security in Storage Workshop (pp. 20-20). IEEE.

[18] Sen, J. (2015). Security and privacy issues in cloud computing. In Cloud technology: concepts, methodologies, tools, and applications (pp. 1585-1630). IGI global.

[19] Alliance, C. S. (2020). Top Threats to Cloud Computing. [online] Available:https://cloudsecurityalliance.org/artifacts/top-threats-to-cloud-computing-egregious-eleven; accessed on Mar 16, 2021.

[20] Amara, N., Zhiqui, H., & Ali, A. (2017, October). Cloud computing security threats and attacks with their mitigation techniques. In 2017 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC) (pp. 244-251). IEEE.

[21] MacAfee, Cloud Adoption and Risk Report. [online] Available: https://www.mcafee.com/enterprise/en-us/assets/reports/restricted/rp-cloud-adoption-risk-report-iaas.pdf assessed on 4 Feb 2021.

[22] Singh, S., & Thokchom, S. (2018). Public integrity auditing for shared dynamic cloud data. Procedia Computer Science, 125, 698-708.

[23] Sharma, M. K., & Nene, M. J. (2020). Two-factor authentication using biometric based quantum operations. Security and Privacy, 3(3), e102.

[24] Moutai, F. Z., Hsaini, S., Azzouzi, S., & Charaf, M. E. H. (2019, March). Security Testing Approach for IaaS Infrastructure. In Proceedings of the 2nd International Conference on Networking, Information Systems & Security (pp. 1-5).

[25] Lawal, B. O., Ogude, C., & Abdullah, K. K. A. (2013). Security management of infrastructure as a service in cloud computing.

[26] Cho, Y., Qu, G., & Wu, Y. (2012, May). Insider threats against trust mechanism with watchdog and defending approaches in wireless sensor networks. In 2012 IEEE symposium on security and privacy workshops (pp. 134-141). IEEE.

[27] Devi, M. G., & Nene, M. J. (2018, March). Scarce Attack Datasets and Experimental Dataset Generation. In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1112-1116). IEEE.

[28] Yaseen, Q, Jararweh, Y., Panda, B., & Althebyan, Q. (2017). An insider threat aware access control for cloud relational databases. Cluster Computing, 20(3), 2669-2685.

[29] Shishir, K. C. DATA CENTER SECURITY & VIRTUALIZATION. Diss. Touro College, 2018.

[30] Rao, R. G., & Nene, M. J. (2017, March). SEDoS-7: a proactive mitigation approach against EDoS attacks in cloud computing. In 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET) (pp. 965-970). IEEE.

[31] Shaikh, A. H., & Meshram, B. B. (2021). Security issues in cloud computing. In Intelligent Computing and Networking (pp. 63-77). Springer, Singapore.

[32] Vaezi, M., & Zhang, Y. (2017). Cloud mobile networks (Vol. 5, No. 3, pp. 23-37). Springer.

[33] Dharmakeerthi, T. D. A Study on Cloud Security Concerns and Resolutions (September 2019).

[34] Ghouse, M., Nene, M. J., & Vembuselvi, C. (2019, December). Data Leakage Prevention for Data in Transit using Artificial Intelligence and Encryption Techniques. In 2019 International Conference on Advances in Computing, Communication and Control (ICAC3) (pp. 1-6). IEEE.

[35] Cullum, p. A survey of the host hypervisor security issues presented in public iaas environments and their solutions.

[36] Halboob, W., Abbas, H., Haouam, K., & Yaseen, A. (2014, August). Dynamically changing service level agreements (SLAs) management in cloud computing. In International Conference on Intelligent Computing (pp. 434-443). Springer, Cham.

Anti-Counterfeiting and Traceability Mechanism Based on Blockchain

Tan Ji [a, 1] and Dr S B Goyal [a]

[a] City University, Petaling Jaya, Malaysia

Abstract: The anti-counterfeiting traceability system based on blockchain technology can ensure the accuracy and consistency of the data stored by each participating node, protect the legitimacy of the data, ensure the product quality, improve the credibility of enterprises, and enhance consumers' confidence in products. However, due to the low system throughput, high energy consumption and poor data availability, the combination of blockchain and traditional anti-counterfeiting traceability mode has many challenges, such as low efficiency. This paper aims to find an improved consensus mechanism based on contribution proof to improve the mining efficiency of honest miners. And plan to introduce a credit system, give priority to the high credit value of the mining block to package, improve the overall packaging efficiency of the system, to solve the problem of low feedback efficiency of the blockchain anti-counterfeiting system.

Keywords: Anti-Counterfeiting, Traceability Mechanism, Blockchain, Network.

1. Introduction

Online shopping has brought significant improvement to people's quality of life. However, the huge number of transactions also gave birth to a large number of fake and inferior commodities. The anti-counterfeiting and traceability system based on blockchain technology can ensure the accuracy and consistency of the data stored by each participating node, protect the legitimacy of the data, ensure the quality of products, improve the credibility of enterprises, and enhance consumers' confidence in products. This chapter introduces the core technology of blockchain anti-counterfeiting traceability system.

1.1 Blockchain
The concept of Blockchain first appeared in a paper "Bitcoin: A peer-to-peer electronic cash system" published by Satoshi Nakamoto on the Bitcoin Forum in 2008, the article pointed out, Blockchain is the basic technology to build the Bitcoin system, and proposes an electronic cash system that is completely realized by peer-to-peer technology, and gives the method realization [1]. Blockchain has four characteristics: decentralization, openness and transparency, consensus mechanism, and anonymity [2-4].

---
[1] Dr S B Goyal,City University, Petaling Jaya, Malaysia
Email: drsbgoyal@gmail.com

Table 1. Blockchain concept

S.No.	Blockchain Concepts	Descriptions	Example/ Real-life Use case
1	Decentralization	Blockchain technology does not rely on centralized hardware or management institutions [5]. The authority of each node is equal, and the processes of data verification, storage, transmission, and maintenance are implemented on the distributed system, which is the most prominent and essential feature of blockchain [6].	Distributed billing Distributed propagation Distributed storage
2	Consensus mechanism	The core component of blockchain. Using the consensus mechanism, users in the blockchain do not need to consider each other's credit or trust each other. Blockchain uses a consensus algorithm based on mathematical principles to establish a trusted network between nodes [7].	PoW consensus algorithm PoS consensus algorithm Distributed storage Paxos consensus algorithm Raft consensus algorithm
3	Smart contract	The smart contract is a transaction agreement that is processed by a computer and can execute contract terms. Its overall goal is to meet the general contract conditions, such as a mortgage, confidentiality, payment, enforcement, etc., and minimize the possibility of malicious or unexpected events, as well as the function of trust intermediary [8].	Decentralized authority

1.2 Types of Blockchain

Blockchains are divided into three types: public blockchain, private blockchain and consortium blockchain [9-10].

1)Public blockchain: It can be understood as a public blockchain, which is completely open and transparent, that is, a blockchain that everyone can participate in. In other words, the behavior on the public chain is public, and it is not controlled by anyone, nor is it owned by anyone. It is considered a "fully decentralized" blockchain.

2)Consortium blockchain: is a semi-public blockchain, which refers to a blockchain that is managed by several institutions. Each institution runs one or more nodes, and

the data in it is only allowed by different institutions in the system. Read, write and send transactions, and record transaction data together.

3)Private blockchain: It is a completely closed blockchain whose write permission is controlled by an organization and institution. The qualifications of participating nodes will be strictly limited.

1.3 Cryptography in the blockchain

In order to ensure the security and integrity of data stored on the blockchain, a variety of modern cryptographic technologies are used in the definition and construction of blocks and blockchains, including public key encryption systems, hash functions, and Merkle trees, etc. [11-14] At the same time, a large number of related cryptographic techniques are also used in the design of a variety of different consensus algorithms.

Table 2. Cryptography in the blockchain

S.No.	Cryptography in blockchain	Descriptions
1	Public key encryption system	The public key cryptographic algorithm requires two keys: a public key and a private key.[15] The public key and the private key are a pair. If the public key is used to encrypt data, only the corresponding private key can be decrypted. On the contrary, if the private key is used to encrypt the data, only the corresponding public key can decrypt it.
2	Hash function	Cryptographic hash algorithm, also known as hash function, is a kind of math-emetically function that can create a small digital "fingerprint" from any da-ta.[16] That is, data of any length can be compressed into a fixed-length binary string within a limited and reasonable time, which is called a hash value.
3	Merkle tree	Another use of hash algorithm in blockchain is to build Merkle Tree, also known as hash tree.[17] It is a binary or multiple tree based on hash value. It consists of a root node, a set of intermediate nodes and a set of leaf nodes.

1.4 Consensus algorithm and classification

Proof-of-work (Po W), also known as workload Proof, is the core consensus algorithm of Bitcoin. Its core idea is to ensure data consistency and security of consensus by introducing computing power competition of distributed nodes, which is also known as "mining" [18]. This algorithm makes the blockchain system need to consume a large amount of computing power, so the consensus efficiency is low.

Table 3. Consensus algorithm and classification

S.No.	Classification	Descriptions
1	PoS consensus algorithm	Proof of Stake is a voting mechanism.[19] Compared with PoW consensus mechanism, calculation power resources are too much wasted, PoS consensus algorithm only needs a small amount of calculation to ensure the normal operation of block chain.
2	DPoS consensus algorithm	The Delegated Proof mechanism of Stake by DPoS (Delegated Proof of Stake) was first proposed and applied by Bitshares in August 2013[20]. On the basis of PoS, DPoS professionalizes the role of bookkeeper. First, bookkeepers are selected through equity. Then, bookkeepers take turns to keep accounts. The advantage is that the number of participating verification and billing nodes can be greatly reduced and the consensus verification can be achieved at the sec-ond level, but at the same time, the whole consensus still relies on token, which still does not solve the pain point of commercial use.
3	Paxos consensus algorithm	Paxos algorithm is the consensus algorithm proposed by LeslieLamport in her paper in 1990 [21]. The main purpose of Paxos algorithm is to gradually reach consensus among every participant participating in distributed processing through this consensus algo-rithm. In the concrete implementation, it is divided into three roles: Proposer is responsible for Proposer, accept receiver is re-sponsible for making decision on Proposer, Learner is responsible for learning the results of Proposer. A process may play multiple roles simultane-ously.
4	Raft consensus algorithm	Due to the complexity of Paxos algorithm, it was difficult to under-stand, so Diego proposed Raft algorithm [22]. As a simple implemen-tation of the Paxos algorithm, it ACTS like Paxos and is more efficient than Paxos, but it has a very different architecture, which makes Raft much easier to understand than Paxos.

1.5 Anti-counterfeiting and traceability

We are exploring the anti-counterfeiting to prevent counterfeiting and traceability to access the product moment in the system. We are doing literature survey for the same in this section.

1.5.1 Anti-counterfeiting

Anti-counterfeiting refers to a measure taken proactively to prevent counterfeiting as a means to imitate, copy or counterfeit and sell others' products without the permission of the trademark owner.[23]

Anti-counterfeiting technologies include laser anti-counterfeiting, query digital label anti-counterfeiting (one-dimensional code), textured anti-counterfeiting label, security thread anti-counterfeiting, unlimited anti-counterfeiting, etc.

At present, the anti-counterfeiting technology has been developed to the fifth genera-tion of products, the use of mobile phone Internet for anti-counterfeiting technology. By attaching anti-counterfeiting labels that automatically identify mobile phones to products and packaging, consumers can use their mobile phones to scan the QR code for authenticity identification. When combined with texture anti-counterfeiting and security line anti-counterfeiting, consumers will truly feel relieved and comfortable shopping.

From Figure 1, we can see that the industries affected by blockchain ecology include finance, health care, culture, social welfare, education, product supply chain, etc. At the same time, it can serve individuals and enterprises. Therefore, blockchain and anti-counterfeiting traceability will have a good combination.

1.5.2 Traceability Mechanism

The traceability is the earliest food safety management system established and perfected by the European Union in 1997 in response to the "mad cow disease" problem.[24] At present, traceability technology is usually combined with anti-counterfeiting technology. A single product is given a unique QR code as an anti-counterfeiting ID through professional machinery and equipment, so that "each product has its own identity code". Then data can be collected and tracked in various links such as product production, warehousing, distribution, logistics transportation, market inspection, sales terminals, and constitute a full life cycle management of product production, storage, sales, circulation and service.
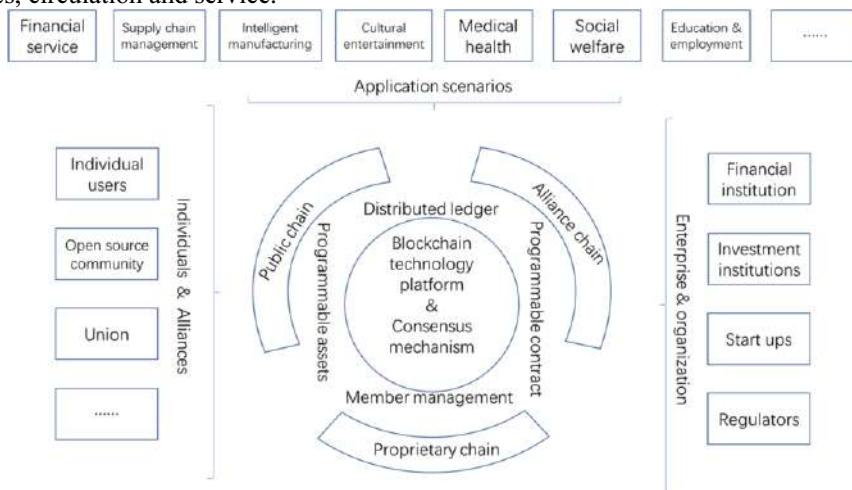
Figure 1. Blockchain ecology

It can be seen from Picture 2 that the existing anti-counterfeiting traceability system stores data in the traceability management system server through raw materials, logistics, storage, production, sales, and other processes. The anti-counterfeiting requirements of products can be queried through the traceability chain to find the corresponding information in each link of production. At the same time, government regulatory agencies, corporate regulatory agencies, and consumers can access the traceability management system server through the Internet and the Internet of Things to query the required information.

2. Literature Survey

At present, the application of blockchain in anti-counterfeiting is less in the world. The first application example is to input the information of containers and goods into the blockchain system, which arranges the transportation route and date of containers, so as

to realize the intelligent anti-counterfeiting and traceability of the goods in the container. Other applications mainly focus on luxury traceability, banks, insurance companies, etc.

China's domestic e-commerce companies introduced blockchain technology earlier. Both Alibaba and jd.com make use of blockchain, Internet of things, big data and cloud computing technology component blockchain anti-counterfeiting traceability platform to realize the whole process traceability of one code for one thing or one code for one batch across brands, retailers, consumers and channel providers.
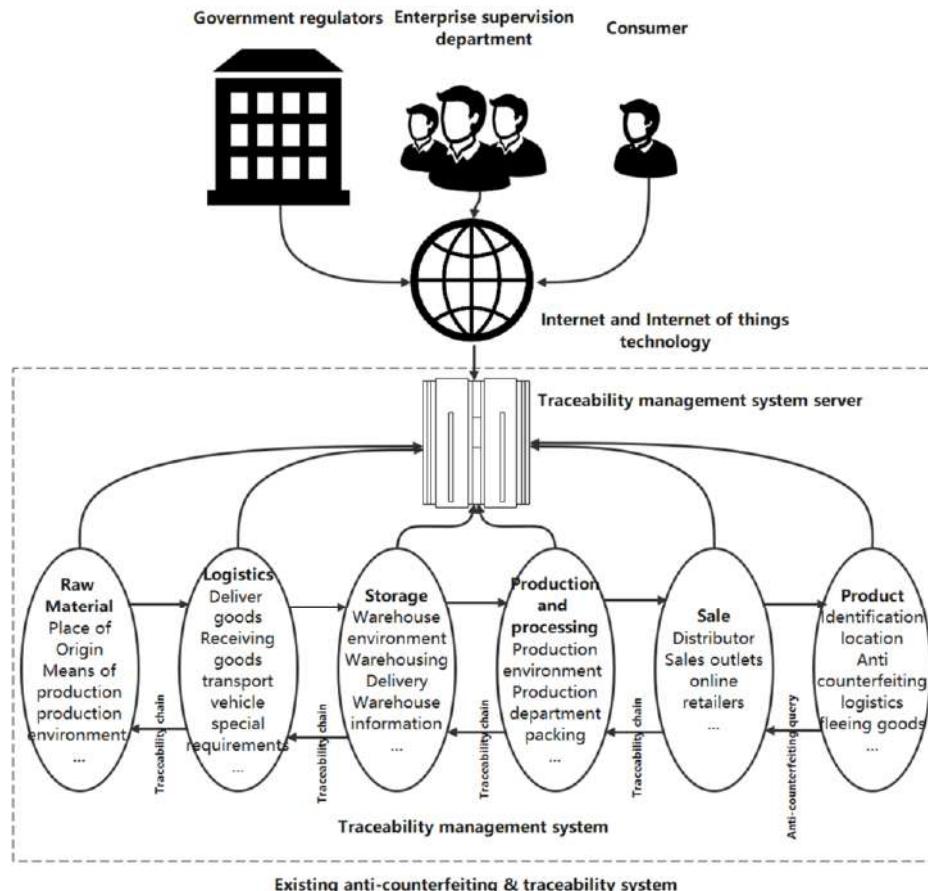
Figure 2. Existing anti-counterfeiting & traceability mechanism system

In terms of system efficiency, Jia Dayu et al. proposed an effective blockchain storage capacity scalable model query method Elastic QM, which stores data in the user layer, query layer, storage layer and data layer modules to improve system query [25]; Qiao Rui et al. designed a consensus mechanism for the security of dynamic data storage based on blockchain and gave a mathematical model for the security of dynamic data storage. Experiments have shown that this solution can effectively improve dynamic data. [26]; Based on the blockchain traceability platform, Liu Yadong proposed a rapid

block generation strategy and a dual-chain storage mechanism, which improved the storage performance and security of the platform [27].

3. Research Gaps

First, a method needs to be found to modify consensus mechanism based on contribution proof. Without modifying the existing data structure of bitcoin, we can reward the miners who successfully publish the block and punish the miners who publish the false block. Then, different mining difficulties are given to different addresses, so that honest miners have a higher probability of obtaining bookkeeping rights and rewards. However, it is necessary to control the probability curve so that the whole mining process is still friendly to the miners with only 0 or a few successful times.

Secondly, the credit system is introduced to pack the blocks issued by miners with high credit value, and improve the efficiency of package system, so as to solve the problem that enterprises and consumers can find the fake and inferior goods in the market for the first time.

Thirdly, we should decentralize the credit system to solve the problem that the transaction information is false when it is first released.

4. Objectives

1) Find a way to modify the consensus mechanism based on proof of contribution to improve the mining efficiency of honest miners.

2) Introduce a credit system, prioritize the packaging of blocks released by miners with higher credit values, and improve the overall packaging efficiency of the system.

3) Decentralize the management of the credit system to solve the situation that the transaction information is false when it is initially released.

5. Proposed Hypothesis

The research challenges of blockchain-related issues involved in this research:

1) How to design a contribution proof system to count the contribution of honest miners?

We are focusing on a contribution proof system using the concept of proof-of-stack.

The contribution system can reward honest miners and increase their probability of becoming bookkeeper.

2) How to join the contribution system into the consensus mechanism?

Joining the consensus mechanism is the way to apply the contribution system. The combination of the two algorithms can verify whether the contribution system can improve the probability of honest miners becoming bookkeeper.

3) How to design the contribution system so that it is still friendly to new miners?

In the same contribution system, if the honest miners' contribution is large enough, it will get a higher probability to become bookkeeper. But this is not friendly to new miners. It is necessary to make the probability curve tend to be flat in a certain probability.

4) Can credit system improve the overall discrimination efficiency of the system?
The introduction of credit system can make the block produced by honest miners with high credit degree pack in advance, rather than fixed time package, so as to improve the overall discrimination efficiency of the system.
5) How to decentralize the credit system?
Credit systems need to be managed as decentralized as other data. This can avoid the appearance of fraud.

6. Proposed Methodology

The research methods considered and selected in this study mainly include the following three types:
1) Literature research method: Extensively consult related literature materials related to the application of blockchain technology in anti-counterfeiting and traceability systems.
2) On-site investigation method: specifically investigate the needs of enterprises with anti-counterfeiting and traceability requirements, and conduct on-site visits to relevant government departments, logistics companies, commodity wholesale markets, supermarkets and other locations that may produce counterfeit and inferior products, and clarify the existing anti-counterfeiting and traceability systems Propose corresponding solutions to the problem.
3) Comparative analysis method: By studying relevant documents, comparing existing anti-counterfeiting and traceability systems, analyzing the advantages and disadvantages of each system, and providing technical and theoretical support for subsequent research.

7. Expected Result

1) The introduction of a contribution proof mechanism can improve the reward system of honest miners. The blocks distributed by honest miners can be packed first, which reduces the space occupied by blocks, reduces the block packing time, and effectively improves the efficiency of the system.
2) The decentralized management of credit systems can effectively avoid source data fraud.
3) Expected system architecture: Based on the traditional anti-counterfeiting traceability system, the contribution proof algorithm is added. The improved system architecture is shown in Figure 3.

8. Discussion

1) The introduction of a contribution proof mechanism can improve the reward system of honest miners. This kind of incentive mechanism makes more and more honest miners join the system, so as to realize the process that consumers help enterprises monitor and identify fake and shoddy products, and save a lot of costs. In addition, due to the decentralized and anonymous characteristics of blockchain, it can effectively avoid the occurrence of false information in the transaction process.
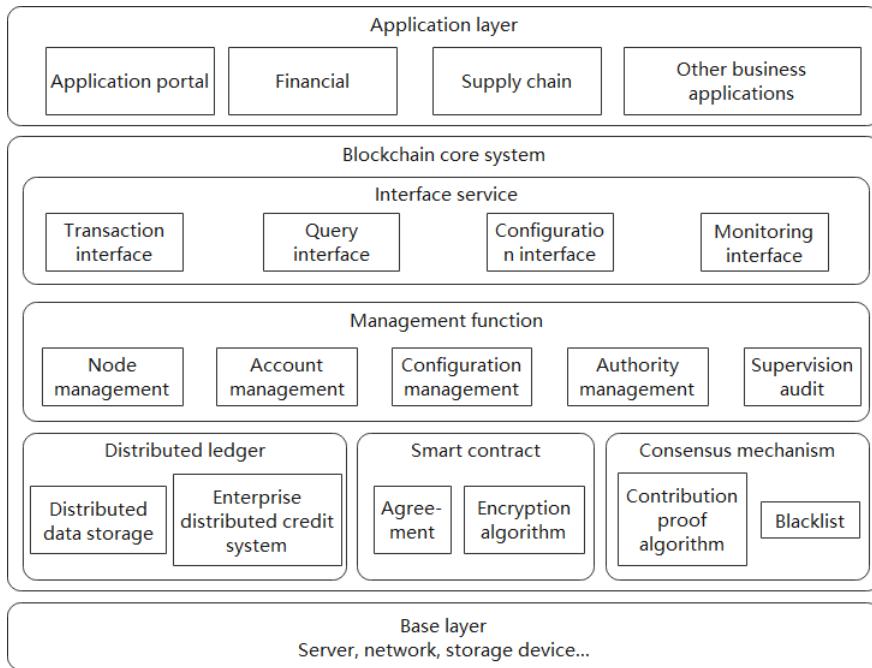
Figure 3. Optimized anti-counterfeiting traceability system architecture

2) The introduction of a credit system and the priority credit principle for enterprises and users with high reputations can effectively avoid the occurrence of source data fraud. At the same time, the distributed management of the credit system can improve the credibility and security of the credit system.

3) Figure 3 is the improved expected system architecture. As can be seen from Figure 3, the entire system architecture is divided into three parts, the application layer, the blockchain core system, and the base layer. The application layer contains the front-end architecture of the system, which is the user-accessible part. The core system of the blockchain is a key part of the entire architecture. Among them, the interface service provides a data link connecting the application layer and the data layer. The management function is the overall management part of the system. The distributed ledger stores all blocks generated by transactions. The smart contract contains part of the transaction agreement and encryption algorithm. The consensus mechanism includes an improved contribution proof algorithm and a blacklist. Finally, the basic layer includes hardware facilities such as servers, networks, and storage devices.

9. Conclusion

In this proposal, we have identified problems of counterfeiting and non-traceability concerns in traditional computing network system. Counterfeiting and non-traceability increases the inaccuracy and inconsistency in the system.

We are working on these problems to improve the accuracy and maintain the consistency of the data stored by each stakeholder and protect the acceptability of the data and credibility of the system; Effectiveness - The credit system is introduced into the consensus algorithm to improve the overall efficiency of the system. Decentralization - Management of the credit system can avoid the phenomenon that product information is false information at the beginning, and on the other hand can avoid the drawbacks of centralized management of the credit system and reduce the possibility of system attacks.

References

[1]   Nakamoto, S., "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008, pp. 1-9. https://bitcoin.org/bitcoin.pdf

[2]   Zhang N , Wang Y , Kang C , et al. Blockchain Technique in the Energy Internet: Preliminary Research Framework and Typical Applications[J]. ZhongguoDianjiGongchengXuebao/Proceedings of the Chinese Society of Electrical Engineering, 2016, 15(36):4011-4022.

[3]   Zhu X . Research on blockchain consensus mechanism and implementation[J]. IOP Conference Series: Materials ence and Engineering, 2019, 569:042058-.

[4]   Lan, Wang, Ranran, et al. A Study on the Influence of Blockchain Technology on Auditing[C]// 2018.

[5]   Chu S , Wang S . The Curses of Blockchain Decentralization[J]. 2018.

[6]   Li X , Mei Y , Gong J , et al. A Blockchain Privacy Protection Scheme Based on Ring Signature[J]. IEEE Access, 2020, 8:76765-76772.

[7]   Xu Y , Li Q , Min X , et al. E-commerce Blockchain Consensus Mechanism for Supporting High-Throughput and Real-Time Transaction[J]. 2016.

[8]   LinoyS ,Stakhanova N , Matyukhina A . Exploring Ethereum's Blockchain Anonymity Using Smart Contract Code Attribution[C]// 2019 15th International Conference on Network and Service Management (CNSM). IEEE, 2020.

[9]   Deng L , Chen H , Zeng J , et al. Research on Cross-Chain Technology Based on Sidechain and Hash-Locking[M]// Edge Computing – EDGE 2018. Springer, Cham, 2018.

[10]   Guillaume Vizierécole Vizier, GramoliV .ComChain: Bridging the Gap Between Public and Consortium Blockchains[C]// ComChain: Bridging the Gap Between Public and Consortium Blockchains. 2018.

[11]   Pinch R G E . On using Carmichael numbers for public key encryption systems[C]// Springer, Berlin, Heidelberg, 1997.

[12]   Chen L , Lee W K , Chang C C , et al. Blockchain based searchable encryption for electronic health record sharing[J]. Future Generation Computer Systems, 2019, 95(JUN.):420-429.

[13]   BelejO ,Staniec K , Wickowski T . The Need to Use a Hash Function to Build a Crypto Algorithm for Blockchain[C]// International Conference on Dependability and Complex Systems. 2020.

[14]   BaohongH ,Yihui Z , Sude Q . Overview of Blockchain Technology[J]. Computer Engineering, 2019.

[15]   Chen B , Wu L , Wang H , et al. A Blockchain-Based Searchable Public-Key Encryption With Forward and Backward Privacy for Cloud-Assisted Vehicular Social Networks[J]. IEEE Transactions on Vehicular Technology, 2020, 69(6):5813-5825.

[16]   BelejO ,Staniec K , Wickowski T . The Need to Use a Hash Function to Build a Crypto Algorithm for Blockchain[C]// International Conference on Dependability and Complex Systems. 2020.

[17]   Mohan A P , Mohamed A R , Gladston A . Merkle Tree and Blockchain-Based Cloud Data Auditing[J]. International Journal of Cloud Applications and Computing (IJCAC), 2020, 10.

[18] Mohan A P , Mohamed A R , Gladston A . Merkle Tree and Blockchain-Based Cloud Data Auditing[J]. International Journal of Cloud Applications and Computing (IJCAC), 2020, 10.

[19] Bach L M ,Mihaljevic B , Zagar M . Comparative analysis of blockchain consensus algorithms[C]// 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). 2018.

[20] Chen S Y , Liu X . Data mining from 1994 to 2004: an application-orientated review[M]. Inderscience Publishers, 2005.

[21] Yang F , Zhou W , Wu Q , et al. Delegated Proof of Stake with Downgrade: A Secure and Efficient Blockchain Consensus Algorithm with Downgrade Mechanism[J]. IEEE Access, 2019, PP(99):1-1.

[22] Zi-Can X U , Rong-Quan W U . Research on Paxos Algorithm Based on Messages Passing[J]. Computer Engineering, 2011, 37(21):287-290.

[23] TuylsP , Guajardo J , Batina L , et al. Anti-Counterfeiting[M]// Security with Noisy Data. Springer, 2007.

[24] Ajila S A ,Kaba A B . Using traceability mechanisms to support software product line evolution[C]// Information Reuse and Integration, 2004. IRI 2004. Proceedings of the 2004 IEEE International Conference on. IEEE, 2004.

[25] FazlaliM ,Eftekhar S M , Dehshibi M M , et al. Raft Consensus Algorithm: an Effective Substitute for Paxos in High Throughput P2P-based Systems[J]. 2019.

[26] Da-Yu J , Jun-Chang X , Zhi-Qiong W , et al. Efficient Query Model for Storage Capacity Scalable Blockchain System[J]. Journal of Software, 2019.

[27] Rui Q , Shi D , Qiang W , et al. Blockchain Based Secure Storage Scheme of Dynamic Data[J]. Computer ence, 2018.

Data Security and Privacy-Preserving Framework Using Machine Learning and Blockchain in Big-Data to Data Middle Platform in the Era of IR 4.0

CHEN CHUQIAO [a, b,1] , S. B. GOYAL [a]
[a] *City University, Petaling Jaya, Malaysia*
[b] *City College of Huizhou, China P.R.C*

Abstract. The modem data is collected by using IoT, stored in distributed cloud storage, and issued for data mining or training artificial intelligence. These new digital technologies integrate into the data middle platform have facilitated the progress of industry, promoted the fourth industrial revolution. And it also has caused challenges in security and privacy-preventing. The privacy data breach can happen in any phase of the Big-Data life cycle, and the Data Middle Platform also faces similar situations. How to make the privacy avoid leakage is exigency. The traditional privacy-preventing model is not enough, we need the help of Machine-Learning and the Blockchain. In this research, the researcher reviews the security and privacy-preventing in Big-Data, Machine Learning, Blockchain, and other related works at first. And then finding some gaps between the theory and the actual work. Based on these gaps, trying to create a suitable framework to guide the industry to protect their privacy when the organization contribute and operate their data middle platform. No only academicians, but also industry practitioners especially SMEs will get the benefit from this research.

Keywords. Data Security, Privacy-preserving, Machine Learning, Blockchain, Big-Data, Data Middle Platform, IR 4.0

1. Introduction

Modern data is nothing new to any enterprises as well as smaller and medium-sized firms due to multiple benefits like cost-cutting and increase efficiency and effectiveness in the data management system in the circumstance of Big-Data.

According to calculation, there are over 2.5 Eb of the data (2.5×10^{18} bytes) has been created in pre-day, and the number of data creating is still increasing [11]. Data is collected by using internet of things technology, transferred via the Internet, stored in the distributed cloud storage, released and using in the industry such as data mining, training artificial intelligence, or business decision making. To effectively manage and utilize these data, the Big-Data theory has been come up. In the engineering field, the Alibaba Group first create a new concept called the Data Middle Platform which is

---

[1] CHEN CHUQIAO, Ph.D. candidate in City University, Malaysia, working in City College of Huizhou
E-mail: james1989vip@163.com.

based on the Big-Data theory [26]. Sometimes Middle Platform also can be seen as a strategy to guide the Big-Data construction for enterprises. Because different organizations have different situations, the pathway of implement the Data Middle Platform is a few different [25]. The construction goal of the Data Middle Platform can be simply summarized as providing tools, processes, and methodology to realize the abstraction, reuse, and sharing of data capabilities, empower business departments and improve the efficiency of realizing data value. Alibaba creates this concept wants to address the problems of information island and repeated development and to highlight the concepts of data sharing and reuse. That is the main difference between the Data Middle Platform and the existing big data platforms.

New information technologies, that is represented by Artificial Intelligence, Blockchain, Cloud Computing, Internet of Things, and Big-Data, are integrated and have facilitated the progress of industry, promoted the fourth industrial revolution. But it also has caused threats and a challenge for data security and privacy protection. When private data gets in the wrong hands, it causes the interests of people or organizations are harmed. For example, the government`s privacy breach can put confidential information in the hands of the enemy power. A breach in an organization can put asset data in the hands of a competitor. Educational institutions` data breaches could put students' personally identifiable information in the hands of criminals who could launch internet fraud against students` parents and students themselves. A breach at the medical institutions can put the Protected Health Information in the hands of those who sell bogus medicine and then cause the interests of patients damaged. Yet, new data security and privacy challenges are being exposed as the data security and privacy protection framework based on new information technologies missing.

2.   Related Work / Literature Survey

Data Security and Privacy Protection is an interdisciplinary complex problem. Data security refers to the use of a set of methods and technologies to protect an organization's data avoid unauthorized access, destroy, or theft from malicious users throughout its whole lifecycle [21]. But the Privacy concerns the ability of personal or organizations to seclude themselves or information about themselves and thereby express themselves selectively [23].

There is both difference and relation between the data security and Privacy-preventing. Privacy is concern about the use and governance of individual sensitive data — like setting up policies to guarantee the student's personal information is being collected correctly, shared with the right users, and utilized appropriately. Different from privacy, security concentrates more on how to protect data avoid malicious attacks, and the misuse of stolen data for profit. It means that security is essential for protecting data, but not enough to handle privacy [9].

The privacy data breach can happen in any phase of the Big-Data life cycle. But data breaches easily occur in data storage, data transmission, and data release. To solve the privacy leakage issue, predecessors have developed different solutions according to the characteristics of privacy leakage in different stages of the big data life cycle. Privacy-preserving technologies can be classified into the following categories [16].

Privacy-preserving technology based on data distortion

Data distortion-based privacy-preserving technology refers to adding some noise into the original data and then make sensitive data distorted but keeping data properties unchanged. But the statistical characteristics of these distorted sensitive data will not be changed using the privacy-preserving technology based on data distortion.

By extension, there are three kinds of data distortion-based privacy-preserving technologies. The first is randomization. It is a simple way that put some stochastic noise into the raw data and then publishes the disturbing data. The second one is the blocking and cohesion method. Blocking refers to not releasing certain specific data when the data is released, and Cohesion refers to grouping and storing the original data, and then combining them together during statistics, to fulfill the effect of privacy protection. The third one is differential privacy.

Differential privacy, as be short as DP, is a new kind of privacy-preventing model [6-7]. This method is able to unriddle two major shortcomings of the general privacy-preventing model: First of all, it gives a fairly precise attack model. In the model, the researcher doesn`t need to care about the background knowledge of malicious users, even if the malicious users have a good command of all record information except for a certain record, the privacy of the record cannot be uncovered. Secondly, there is a rigorous definition and a quantitative evaluation method has been given for the grade of privacy protection. Due to differential privacy`s advantages, the traditional privacy protection models are quickly replaced. Now, differential privacy is widely discussed in the privacy research area and has attracted the attention of many fields. Not only include computer science, but also contain database, data mining, and machine learning.

3. Privacy-preserving technology based on data encryption

Data encryption-based privacy-preserving technology means the method of using encryption technology to hide privacy in the data mining process. Two representative data encryption-based privacy-preserving technologies are the security multi-party computation and the homomorphic encryption.

The security multi-party computation belongs to a subdiscipline of cryptography. It is also called secure computation, multi-party computation (MPC), or privacy-preserving computation. The aim of security multi-party computation is to create methods to help parties involved in the computation to complete the computation while keeping the data that input by each participant private. It is different from traditional cryptographic tasks. The traditional cryptographic task is using passwords to make sure the data is secure and integrity in communication and storage. It is only useful for the user who outside the system. For the user who involves in, the traditional cryptographic task can do anything to help. But the security multi-party computation gives an idea to solve this situation. The security multi-party computation will protect each party's privacy only be known by itself [22]. In order to ensure data integrity, the blockchain will be used in the process of data transmission and storage.

The homomorphic encryption is an encryption form created by Gentry in 2009 [1]. He puts forward a feasible method which is called "fully homomorphic encryption" in mathematic. That means the encrypted data can be operated without decryption, and the

result which operates by the encrypted data is the same as the result which operates by the encrypted data after decryption [1].

4.   Privacy-preserving technology based on restricted release

The privacy-preserving technology based on restricted release is to realize privacy protection by controlling the release of original data. For example, people can release the filtered data or sensitive data with low precision and then make privacy protection.

Normally the research on restricted release-based privacy-preserving technology focuses on two aspects. One is data generalization, the other one is data anonymization. The aim of them is the same is to ensure that the risk of leakage of sensitive data and privacy is within a permissible range.

In general, data generalization has using a process to summarize data by replacing relatively low-level values with higher-level concepts, or by reducing the number of dimensions so that the data can use fewer dimensions to cover. For example, in educational institutes, when the engineer designs a sheet of the database, he can use mark grade from the letter A to the letter E instead of numeric values for an attribute student`s mark. Or, removing birth date and telephone number when summarizing the behavior of a group of students. Given the large amount of data stored in databases, it is useful to be able to protect the specific privacy value at generalized levels of abstraction [8].

Currently, there are three kinds of technologies which is K-anonymity, L-diversity, and T-closeness in data anonymization.

The earliest widely accepted privacy protection model in Privacy-preserving technology based on restricted release is K-anonymity, which has been defined in 2002. In order to dispose of the de-anonymization attacks, each data record that is released by k-anonymity must be difficult to distinguish from no less than k-1 other records (called an equivalence class). Even though the hacker gets the data that is deal with by using the k-anonymous model, he will get the records of at the minimum k different people, and thus cannot make an accurate judgment. The parameter k signifies the strength of privacy-preventing. The larger number the K gets, the stronger strength of privacy-preventing you get. But it also means the lower availability of the data and the more information will be lost [18-19].

 In 2006, Machanavajjhala et al. [14] who is working at the Cornell University noticed that the k-anonymity model has a weak point. Even though there is nothing be restricted on sensitive attributes. Hackers also can confirm the relationship between sensitive data and individuals by using background knowledge attacks, re-identification attacks, and consistency attacks. For example, the attacker obtains the k-anonymized data. if the equivalence class of the k-anonymized data is all AIDS patients, then the attacker can easily make the judgment which one in the k-anonymized data definitely has AIDS. To prohibit consistency attacks, the new privacy protection model l-diversity improves k-anonymity to ensure that the sensitive attributes in any equivalence class have at least l different values. Based on l-diversity, t-Closeness requires the distribution of sensitive attributes in all equivalence classes to be as close as possible to the global distribution of the attribute [13]. (a, k)-anonymity principle, on the basis of k-anonymity, further ensure that the percentage of records related to any sensitive attribute value in each equivalence class is not higher than a [17, 20, 24].

However, the privacy-preventing model above is still flawed and needs to be upgraded continuously [2, 3, 5, 10, 15]. Fundamentally, no single privacy protection model can effectively protect privacy. Only by using various privacy protection technologies comprehensively to form a privacy protection technical framework can privacy data be protected effectively.

5.   Problem Statement

This research aims to build a privacy protection framework based for Data Middle Platform as follows:
- To list out the effect of security leak & methods in the modern data
- To list out the effect of privacy leak & methods in the modern data
- To compare big-data and Data middle platform
- To identify the importance of Big-Data and Data middle platform in the era of IR 4.0
- To design a framework to handle the security and privacy in the modern data
- To utilize the machine learning techniques in the proposed framework
- To apply the blockchain techniques in the proposed framework
- To test the proposed framework in real-life data

6.   Solution Approach

The proposed research needs the machine learning techniques concept to handle the security and privacy in the large volume of data in addition to the blockchain due to its core characteristics like a consensus, smart contract, public, and private key notion. So, the researcher will apply machine learning and blockchain techniques to ensure no data accessibility to an unauthorized person and no one can do the unwanted operation without the accessibility rights.
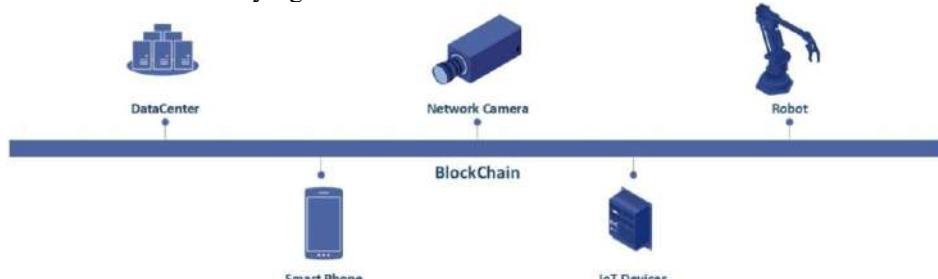
Figure 1. Different Privacy Sources Formed Blockchain

Figure 1 shows what is in the Blockchain. In IR 4.0, privacy comes from many pathways. Some privacy comes from the existing database located in the data center. But others are coming from devices such as IoT devices, mobiles, network cameras, and industrial robots. That privacy will be packed and then put in the Blockchain.
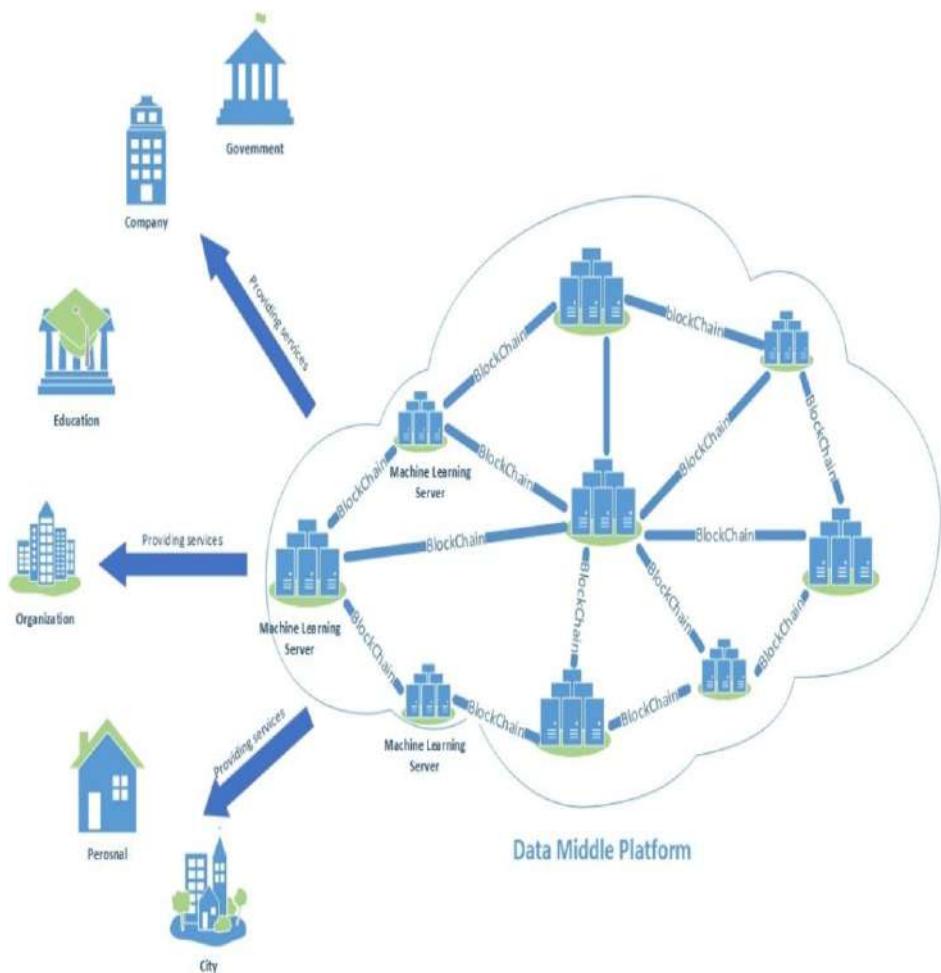
Figure 2. Framework of Privacy Blockchain in Data Middle Platform

The Blockchain will flow in the Data Middle Platform as shown in the figure 2. It will be analyzed by using Machine Learning theory. The user will get the result but doesn`t know the specific data. In this way, the privacy will be protected and then avoid different kinds of malicious attacks.

There are five steps which lists in below that will be confirmed in the research.

- A preliminary literature review study will be performed. This encompasses the background theory and pertinent topics including data security technology, privacy protection, machine learning, blockchain, Big-Data, Data Middle Platform, and the Industry Revolution 4.0.
- The literature review will be followed by the state-of-the-art in Data security and privacy-preventing for Big-Data in the Era of the Industry Revolution 4.0. This includes classifying the technologies of data security and privacy protection for identifying the relevant literature. We will use the results to guide to identify gaps in the current research on privacy-preservation in order to suggest areas for further investigation.

- Data security and privacy-preventing framework using Machine Learning and Blockchain in Big-Data to Data Middle Platform which is consisted of conceptual and logic will be provided.
- we will try to use a use case to verify the framework and then try to identify any potential defects. In this section there are several open-source architectures will be used such as Apache Hadoop, Docker, MongoDB, MySQL, Python.
- Papers will be presented at publications and conferences of proceedings for reviewed and discussion.

7. Expected Impact

This research work will benefit academicians, industry practitioners, and researchers to open the new dimension in the middle data platform and SME to Enterprises will not hesitate to adopt the middle data platform and enhance the socio-economy aspect of the enterprises.

8. Conclusion

With the rapid growth of the number of data, data security and privacy-preserving technology are not adequate in the era of the fourth industrial revolution. This is because the development of the data middle platform makes the data security and privacy-preserving technology delay. The old technology may not ensure the user`s privacy because of operational and efficiency problems. So, updating the data security and privacy-preserving framework using new technology such as machine learning and blockchain is of great urgency. Not only to help the industry protect the privacy in Data Middle Platform but also will provide clear guidance to those people who will be involved in the data governance of new initiatives related to data security and privacy-preserving.

References

[1] Gentry, Craig. A fully homomorphic encryption scheme. Vol. 20, no. 9. Stanford: Stanford university, 2009.

[2] Campan A, Truta TM, Cooper N. 2010. User-controlled generalization boundaries forp-sensitivek-anonymity. In: Proceedings of the 2010 ACM Symposium on Applied Computing - SAC '10.

[3] CAO M, Zhang L, Bi X, Zhao K. 2017. Personalized (α, l) -diversity k-anonymity Model for Privacy Preservation. Computer Science. 45(11):180–186. doi:https://doi.org/10.11896/j.issn.1002-137X.2018.11.028.

[4] CEH v11 Note Summary Module 1 (A). 2020 Nov 27. blogcsdnnet. https://blog.csdn.net/taof211/article/details/110219994.

[5] Chen XB, Han B, Huang S. 2018. Research on fuzzy t-intimacy protection method. Computer Applications and Software. 35(9).

[6] Dwork C. 2006. Differential Privacy. In: Proceedings of the 33rd International Colloquium on Automata, Languages and Programming. Venice, Italy,. p. 1–12.

[7] Dwork C. 2008. Differential privacy: A survey of results. In: International conference on theory and applications of models of computation. Berlin, Heidelberg: Springer. p. 1–19.

[8] Han, J. K. (2012). Data Mining: Concepts and Techniques. Morgan Kaufmann.

[9] Jain P, Gyanchandani M, Khare N. 2016. Big data privacy: a technological perspective and review. Journal of Big Data. 3(1). doi:10.1186/s40537-016-0059-y.

[10] Jurczyk P, Xiong L. 2009. Distributed anonymization: Achieving privacy for both data subjects and data providers. In: IFIP Annual Conference on Data and Applications Security and Privacy. Springer. p. 191–207.

[11] Karki D. 2020 Nov 9. Can you guess how much data is generated every day? Takeo. [accessed 2021 Feb 16]. https://www.takeo.ai/can-you-guess-how-much-data-is-generated-every-day/.

[12] LI B. 2018. Research on privacy protection data release algorithm based on different anonymous requirements. [Jilin University].

[13] Li N, Li T, Venkatasubramanian S. 2007. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In: 2007 IEEE 23rd International Conference on Data Engineering.

[14] Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M. 2007. L-diversity. ACM Transactions on Knowledge Discovery from Data. 1(1):3-es. doi:10.1145/1217299.1217302.

[15] Martinelli F, SheikhAlishahi M. 2019. Distributed Data Anonymization. In: 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech).

[16] Privacy protection technology in big data environment. 2015 Jun 1. Cyberspace Administration of China. [accessed 2021 Jan 22]. http://www.cac.gov.cn/2015-06/01/c_1115473995.htm.

[17] SUN J. 2019. Research and Implementation of k-anonymity based on sensitivity classification [Beijing University of Posts and Telecommunications].

[18] SWEENEY L. 2002a. ACHIEVING k-ANONYMITY PRIVACY PROTECTION USING GENERALIZATION AND SUPPRESSION. In: International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. Vol. 10. p. 571–588.

[19] SWEENEY L. 2002b. k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY. In: International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. Vol. 10. p. 557–570. [accessed 2019 May 10]. https://epic.org/privacy/reidentification/Sweeney_Article.pdf.

[20] Wang J, Li H, Guo F, Zhang W, Cui Y. 2019. D2D Big Data Privacy-Preserving Framework Based on (a, k)-Anonymity Model. Mathematical Problems in Engineering. 2019:1–11. doi:10.1155/2019/2076542.

[21] What is data security? Definition, solutions and how to secure data. 2021. www.ibm.com. [accessed 2021 Feb 16]. https://www.ibm.com/topics/data-security.

[22] Wikipedia Contributors. 2019 Dec 6. Secure multi-party computation. Wikipedia. https://en.wikipedia.org/wiki/Secure_multi-party_computation.

[23] Wikipedia contributors. (2021b, March 23). Privacy. Wikipedia. https://en.wikipedia.org/wiki/Privacy

[24] WU S. 2015. Research on the t-closeness privacy protection model supported by rough sets and clustering. [Shanxi Normal University].

[25] Zhang C, Hou L. 2020. Data middle platform construction: The strategy and practice of National Bureau of Statistics of China. In: Statistical Journal of the IAOS. Vol. 36. p. 979–986.

[26] Zhong H. 2017. The transformation of enterprise IT architecture: Alibaba's strategic thinking and structure practice of middle platform. China Machine Press.

Parallel Deep Learning Framework for Video Surveillance System

Mayuri KARVANDE [a,1] , Apoorv KATKAR [a] , Nikhil KOLI, [a] , Amit JOSHI [a] and Suraj SAWANT [a]

[a] *Dept. of Computer Engineering and IT, College of Engineering Pune (COEP), India*

Abstract. In today's world, the security of every individual has become an important aspect. There is a need for constant monitoring in public places. A Manual operating camera system is an unreliable and very basic and poor method for this purpose. Intelligent Video Surveillance is an approach where multiple CCTVs conscantly record the scenes and proper algorithms are deployed in order to detect and monitor activities. Deep Learning frameworks and algorithms like Kera's, YOLO, Convolutional Neural Networks or backbones for image detection like VGG16, Mobile net, Resnet101 have been used for human and weapon detection. The paper focuses on deep learning techniques and threading to collectively develop a Parallel Deep Learning Framework for Video Surveillance that aims at striking the right balance between accuracy and system performance or stability. Threading is used in terms of implementation of a uniquely proposed Dynamic Selection Algorithm that uses two backbones for object detection and switches between them based on the queue status for achieving system stability. A uniquely designed logistic regression filter is also implemented that boosts the system performance.

Keywords. Surveillance, deep learning, regression filter, ResNet

1. Introduction

In recent times, machine learning and deep learning algorithms have created a lot of buzz. These days surveillance systems play a vital role in public security. CCTV cameras are used in almost all sectors where security has importance. However, most of these systems are very straightforward and have a simple capacity of recording with limited capabilities. Today, when an incident related to public security happens, pre-recorded clips are analyzed by humans. As this is a post-incident analysis, it delays the response to the incident which is the most valuable time in which the crime could be solved the fastest. Surveillance systems are not used efficiently with human interventions and automation is limited. Considering the rapidly increasing availability of cameras and the advance- ments in computing environment, it is worth taking a look at a method to detect events automatically without compromising on system stability as video accounts for huge data volumes [1]. This paper aims at proposing a solution for the above problem.

---

[1] Mayuri KARVANDE ,Dept. of Computer Engineering and IT, College of Engineering Pune (COEP), Pune-411005, Maharashtra, India; E-mail: karvandema17.comp@coep.ac.in.

The paper proposes a new surveillance model using existing deep learning architectures and parallelism techniques. It aims at using the full power of the recent advancements in technology so as to decrease human intervention through the learning capabilities of the deep learning model [2]. It would help in throwing alerts which would not have been possible as it would be practically impossible for a single human to cover the wide vision of the camera and analyse it continuously throughout the day. For example, 24-hour surveillance video input at 20 fps will account for more than 15 lakh frames [3]. It is practically impossible for a human to be attentive for these large time durations, thereby making a human-based approach for the analysis of live surveillance videos in efficient and impractical. It is tedious and time-consuming. Naturally implementing such a complex system in real-time will put considerable stress on the system resources in order to work as required. When implementing such a solution, the system should be reliable. The system reliability cannot be ignored and its solution is discussed in the form of a Dynamic Selection Algorithm in the paper. It is based on the use of threading and a queue. The unique logistic regression filter analyses if there is a chance for a human to exist in the frame and then passes it to the model to analyse only if required [4]. This architecture helps us to only analyse frames that have a significant probability of a human existing. We call this architecture Filter Before using YOLO (FBYOLO). Now, the queue is fed sampled frames from the video feed. Here based on its current state, the queue switches between two deep learning models which are running on two different threads. The two powerful networks of YOLOv3 - ResNet and MobileNet have been used in this method- ology [5]. The queue architecture is explained in subsequent chapters. All the proposed objectives are implemented using one or more deep learning models and one will be selected among them based on the time required for each model to predict so as to arrive at the most relevant and practical solution [6]. We aim at a solution that can be used at various sites like malls, offices, homes, etc for systems with limited computing powers like that of a normal gaming laptop.

2. Literature Review

There has been significant work in the domain of video surveillance systems. Accuracy and time are two equally important aspects for designing any system. Different deep learning algorithms are considered and the algorithm which handles both the aspects and yields better results is selected at a time. This dynamic security-level control algorithm works towards maximizing accuracy related to recognition performance and at the same time minimizing operation time related to the system stability [6]. Surveillance is essential to analyze suspicious activities, most of which occur in crowded public places. Object recognition, crowd analysis and violence detection become the important parameters in this case. Deep learning techniques are capable of incorporating these things. Big data needs to be dealt with in this situation. Human detection and posture identification are incorporated using YOLO, VGG-16, Convolutional Neural Network (CNN). Deep CNN gives better results in terms of crowd analysis. These deep learning models can be combined and implemented on a powerful Graphics Processing Unit (GPU) to get a better Surveillance system [1]. YOLOv2 has proved to be an accurate and faster method than the Region Based CNN (R-CNN) for object detection with a

greater number of fps in the field of image processing. Around 78.6 map and 155 fps can be reached with this method [7]. Another version, YOLOv3 using K-means clustering over the frames of the dataset, called the tiny-YOLOv3, is an improved method for real-time object detection [8]. YOLOv3 is also a better technique over Faster R-CNN in terms of weapon detection. Handguns with various shapes are accurately detected in various scenes considering differment scales, rotations and angles [9]. Weapons and other theft-related small objects can also be detected using a two-level CNN binary classifier. The candidate regions selected in the first level undergo binary classification in the second level based on One-Versus-All or One-Versus-One [10].

Apart from weapon detection, human detection is also important for surveillance. Extraction of information from a foreground-background 2D grey image using Adaptive Gaussian Mixture Model (AGMM) along with the ConvNets YOLO models precisely determines the presence of a human [11]. Background subtraction, optical flow and extraction of spatiotemporal parameters using the Gaussian model helps in the detection of any human in motion [12]. In order to reduce the time required for computationally heavy real-time surveillance tasks, a lightweight CNN technique has been proposed which affordably detects humans on the edge device. This technique has been inspired by the Single Shot Multi-Box Detector (SSD) method [13]. The surveillance system should incorporate the detection of multiple people. W4 is an approach that operates on grey images and creates models of people on the basis of shape analysis, body parts and tracking people. It identifies activities taking place between objects and people using Gaussian models [14]. To achieve parallelism in surveillance, a multi-layer edged Distributed Intelligent Video Surveillance (DIVS) system has been designed using deep learning. Task-level and model-level parallel training methods are used to balance the workload [15].

3. Proposed Solution

The proposed solution is a way to provide various kinds of surveillance with maximized accuracy and performance. CCTV cameras capture and store the instantaneous footages [2]. The camera needs to be positioned such that it covers a wide area and every relevant aspect. The instantaneous video footages are supplied to the system and processed further. This work predominantly focuses on the two major objectives. The first objective is the Stealth Mode Movement Detection which is the most basic level of the surveillance system. The presence of humans in the video feed are identified and analyzed further. While the second important objective is Weapon Detection. Weapons like guns, knives and masks are often used by thieves and robbers to cover their faces and secretly ac- accomplish their goals. Most of the crimes involve gun violence [9]. Thus, the second objective recognizes weapons, particularly guns, from the video feed. These objectives are achieved using deep learning with parallel programming. The proposed methodology, named FBYOLO, is incorporated to minimize the time required to process a frame and maximize the accuracy of the prediction done by the model.

This methodology is based on the 'Dynamic Selection Algorithm' that selects one deep learning framework for optimizing performance-accuracy tradeoff [6]. Figure 2 depicts the block diagram for the system. It is as follows: the input CCTV video

footage is supplied to the model for processing. This video is then sampled into frames in order to process further. These frames are then fed to a simple logistic regression network. This regression is used as a filter. The regression filter is supposed to eliminate the true negative frames or the frames which don't have a human or a target object in it [4]. The FBYOLO recognizes the frames containing human and target weapons and separates them. The logistic regression network smartly calculates the probability of each frame based on the presence or absence of the required objects or human beings. The loss function is used in order to calculate the probabilities. The logistic loss function continuously tries to reduce the cost giving more optimistic results [16]. The frames which accurately predict the absence of the target are filtered out. The filtration saves the processing done for true negatives and thus enhances the frame processing rate of the model. This additional feature helps to get the relevant frames and increase both the accuracy and performance of the system.
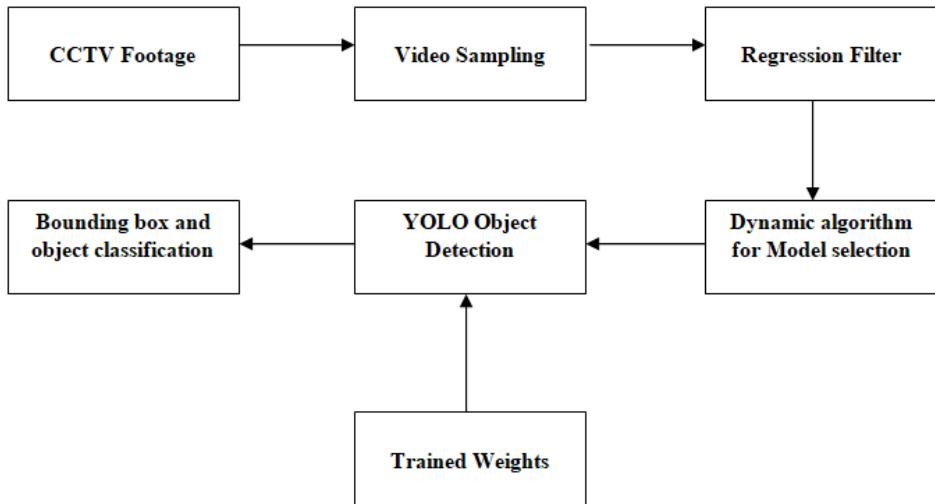
After this initial step, the filtered frames are fed to the YOLOv3 networks. Different deep learning YOLOv3 networks having different time-accuracy characteristics work on these frames [17]. The backbones of these YOLOv3 networks used in the proposed method are ResNet and Mobilenet. The YOLOv3 algorithm is an efficient and faster method for real-time object detection [8]. The YOLOv3 algorithm creates bounding boxes for the image and calculates their probabilities. Based on these probabilities, the image classes are identified [18]. The Dynamic Selection Algorithm has been designed to select the appropriate YOLOv3 backbone model. This algorithm solves the time-accuracy trade-off depending on the number of frames [6]. Figure 1 depicts the pseudocode of this proposed algorithm. All the incoming filtered input frames are fed to a queue. A thread is created in order to enqueue the frames. The total capacity of the queue is represented as $qmax$. The threshold capacity of the queue is around 60%. The queue is considered to be full when its capacity reaches 90%. A separate thread is created to dequeue the frames one-by-one for processing over the YOLOv3 networks. This thread runs parallelly with the enqueue thread. In this dynamic algorithm, the ResNet backbone is set initially. While the frames are fed to the queue, the backlog is constantly compared to the threshold value. If the current queue backlog is more than the threshold value, the backbone network is switched to MobileNet. MobileNet is faster than Resnet but slightly less accurate [19]. MobileNet is used because there are a greater number of incoming frames and hence faster processing is required [20]. Whereas, if the current queue backlog is less than the threshold value, the backbone network is switched to ResNet. In this case, slower processing is permissible due to a smaller number of incoming frames. Higher detection accuracy is achieved using ResNet [21]. ResNet has a greater number of layers in its network and hence is a deeper network than MobileNet which makes it more accurate [22]. However, if the queue is almost full, then all the frames are discarded except for one. The analysis and processing is started again with the MobileNet backbone. Thus, the appropriate network is used in the current condition of queue backlog considering the computation time and accuracy parameters. This concludes the object detection for the given sampled frame and the explanation of the block diagram. The post-processing is done on the frame to draw a box around the detected object (viz. human, weapon).

Algorithm 1 Pseudocode for Dynamic Selection Algorithm

```
 1: q = queue()
 2: qmax = queue.capacity()
 3: threshold = 0.6 * qmax
 4: qalmostfull = 0.9 * qmax
 5: backbone = ResNet
 6: while frames are fed to the q do
 7:     if q.current() <threshold then
 8:         backbone ← ResNet {more accuracy but slower}
 9:     else if q.current() >= threshold then
10:         backbone ← MobileNet {less accuracy but faster}
11:     else {q.current() >= qalmostfull}
12:         discard all frames except one frame f#
13:         backbone ← MobileNet
14:         q ← f#
15:     end if
16: end while=0
```

Figure 1. Algorithm 1 – Pseudocode for Dynamic Selection Algorithm

Figure 2. Block diagram of the proposed system

4. Experimental Setup

All the predictions are done on a gaming laptop system with configuration an Intel Core i5-7300 HQ CPU @ 2.5GHZ, 8GB DDR4 RAM @ 2400 MHz, NVIDIA GTX 1050 GPU. The operating system used is Windows 10. For rapid execution of tasks, hardware accelerators are required. A GPU that supports the CUDA framework is

required in order to achieve parallelism. The most important component of the system is the feed from a video capturing source. High-quality cameras should be positioned to get footage from proper angles. The input feed is scaled down to 512 x 512 pixels for further processing, to have a minimum output resolution of 512 x 512 pixels.

The programming environment required for the system is Python3. It is an easily accessible language with inbuilt libraries. Initially, threading was used for parallelism but it was observed that Keras is not compatible with threading. Hence, the multiprocessing library in python is used for parallelising the prediction using multiple processes as needed and explained in the Dynamic Selection Algorithm. The Queue module from the multiprocessing library is also used. Keras, a library that supports Deep Learning in Python, is used in the implementation. Tensorflow is a library that contains models of Machine Learning and Deep Learning. OpenCV is a fundamental library in real-time computer vision. It is used for video capture and analysis. It is used for human detection, object detection and image processing.

The dataset used for human detection for the first objective is TUD Brussels [23]. The size of the dataset is around 3GB and around 1476 images are used for the first objective. Out of these 1327 images were used for training. The remaining 149 images were used for testing. The dataset used for weapon detection is obtained from Soft Computing and Intelligent Information Systems which is a research group from the University of Granada [24]. It contains around 3000 images of which 1300 are used for training and 1000 for testing purpose.

5. Results and Discussion

The experiments have been carried out on the Nvidia GTX 1050 system. For human detection, images from TUD Brussels dataset have been trained. It has been observed that the frame per second rate for ResNet was 5.2fps. Whereas the rate for MobileNet is found to be 13.5fps. Thus, ResNet is slower than MobileNet. However, ResNet (88%) is found to be more powerful than MobileNet (81%) in terms of accuracy. In order to solve this issue and to maintain equilibrium between the speed and accuracy, an input video of about 8 seconds with a frame rate of 40 fps was fed to the queue at the rate of 6 frames per second. The total number of frames are around 331 out of which none of the frames is filtered out by FBYOLO since all frames are confirmed to be true positive. Metrics of the regression filter can be calculated from its confusion matrix. Left half of the Figure 4 indicates the confusion matrix. The precision, recall and accuracy calculated for the filter are 0.89, 1 and 91 % respectively. Out of 331 frames, the Dynamic Selection Algorithm chose 264 frames to be processed by ResNet, while the remaining 67 by MobileNet. The overall time for prediction is around 55.76 seconds which gives a frame rate of 6fps. This rate can be calculated using the following equation-

$$Overall\ Frame\ Rate = \cfrac{Total\ Frames}{\cfrac{ResNet\ frames}{ResNet\ frame\ rate} + \cfrac{MobileNet\ frames}{MobileNet\ frame\ rate}} \tag{1}$$

Similarly, for an overall rate of 7fps, the prediction requires 51.70 seconds. In this case, the algorithm selected ResNet for 203 frames and MobileNet for 128 frames. Thus, the Dynamic Selection Algorithm helps in running a slower but more accurate model - ResNet at a higher fps with reduced overhead by using a lighter and faster model like MobileNet to predict the backlogged frames. This ultimately results in faster and accurate prediction. For weapon detection, an input video of 6 seconds of 30fps was fed to the queue at the rate of 7 frames per second. Dynamic Selection Algorithm chose Resnet (6 fps) to predict 124 frames out of total 186 frames. For the rest 62 frames MobileNet was chosen (15 fps). Figure 3 indicates the output with FPS values for both the detectors. Similar metrics for different FPS values for Human Detector are tabulated in Table 1. For all the above results the queue size was set to 10 and the threshold value was set to 40% of $qmax = 4$. The accuracy, precision and recall score for the human and weapon detector on specific FPS values using Resnet and MobileNet are tabulated in Table 2. For weapon detector, better results are achieved on the dataset as compared to the research group from the University of Granada. From the above discussion it is quite obvious that the more accurate ResNet will take longer time to process the frames as compared to MobileNet. Hence at lower FPS values the Dynamic Selection Algorithm will choose ResNet for processing majority of the frames. As FPS value increases, ResNet will not be able to keep up with it and the Dynamic Selection Algorithm will choose the faster MobileNet to process majority of the frames. These results are tabulated in graphical form in the right half of the Figure 4.

Table 1. Human Detector Model Summary

FPS	ResNet101 (frames processed)	MobileNetV2 (frames processed)	Total Time Taken in seconds
5	331	0	65
6	264	67	55.76
7	203	128	51.7
8	160	171	43
9	114	217	41
10	94	237	38.70
11	30	301	30.72
12	8	323	29.86

Table 2. Metrics of Human Detector on TUD Brussels Dataset (Input Size = 512*512) and Weapon Detector on Guns Dataset (Input Size = 416*416)

Detector	Model	Accuracy	Precision	Recall	Frame Rate
Human Detector	ResNet101	88.50%	0.9887	0.8940	5.3
	MobileNetV2	81.52%	0.9883	0.82299	13
Weapon Detector	ResNet101	83.86%	1	0.8386	6
	MobileNetV2	81.91%	0.9989	0.8198	15
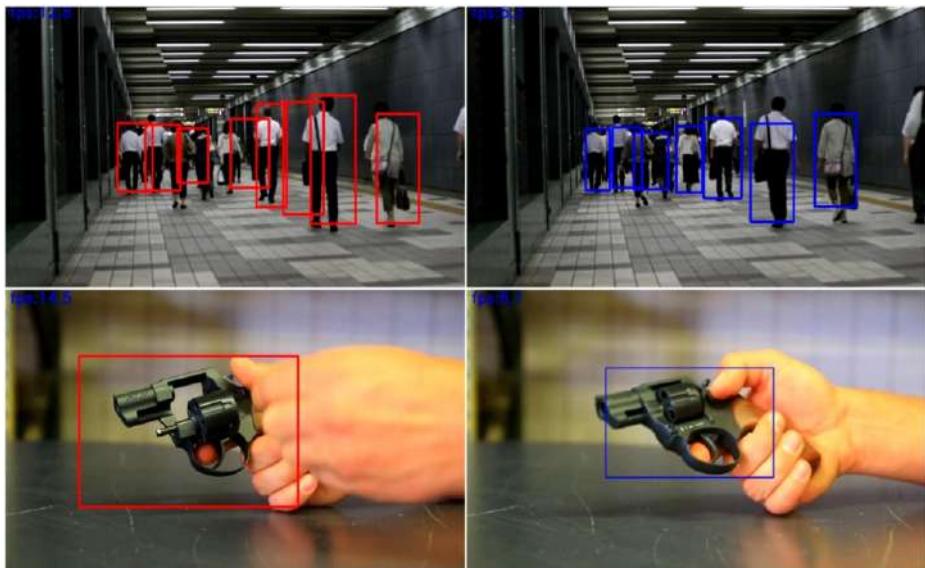
Figure 3. MobileNet vs ResNet with current FPS
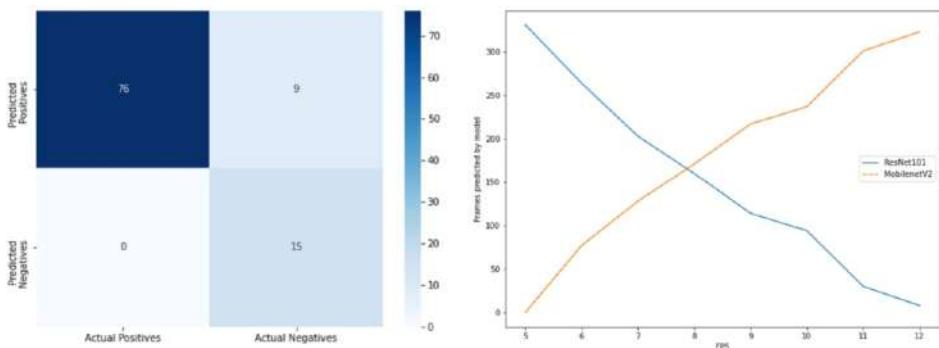
Figure 4. Left: Confusion Matrix of Regression Filter, Right: Model Selection by Dynamic Selection Algorithm

6. Conclusion and Future Scope

Initially, threading was used for introducing parallelism but it was observed that Kera's is not compatible with threading. Hence, multiprocessing was used for parallelizing the prediction using multiple processes as needed and explained in the Dynamic Selection Algorithm. It is clearly observed from the results that with increase in the fps value the number of frames predicted by MobileNet increases which hampers the accuracy of the system. It is balanced by selection using the Dynamic Selection Algorithm at suitable fps value that balances the accuracy and performance so as to achieve an efficient and stable output from the system.

Mobilenet has a small overhead as compared to ResNet. Its weights are of size 60 MB whereas ResNet has a higher overhead with its weights being of size 500MB. The accuracy of ResNet is significantly higher as compared to Mobilenet. Even on a normal system with GTX 1050, values like 7fps are achieved using ResNet with the Dynamic Selection Algorithm.

While selecting a model for a particular application one tends to use a model that has higher accuracy. But when the speed at which the model does the prediction also has an impact on the application it needs to be taken into account. Both high accuracy and faster predictions are not possible at the same time. By using the proposed Dynamic Selection Algorithm, a trade-off between speed and accuracy can be achieved by changing the threshold values and the size of the queue as required for the application. The speed of models with high accuracy can be boosted by using the Dynamic Selection Algorithm.

References

[1] Sreenu, G., Durai, M. S. (2019). Intelligent video surveillance: a review through deep learning techniques for crowd analysis. Journal of Big Data, 6(1), 1-27.

[2] Kardas, K., Cicekli, N. K. (2017). SVAS: surveillance video analysis system. Expert Systems with Applications, 89, 343-361.

[3] AlGhamdi, M. A., Khan, M. A., AlMotiri, S. H. (2015, October). Automatic motion tracking of a human in a surveillance video. In 2015 IEEE First International Smart Cities Conference (ISC2) (pp. 1-4). IEEE.

[4] Feng, J., Xu, H., Mannor, S., Yan, S. (2014). Robust logistic regression and classification. Advances in neural information processing systems, 27, 253-261.

[5] Yadav, N., Binay, U. (2017). Comparative study of object detection algorithms. International Research Journal of Engineering and Technology (IRJET), 4(11), 586-591.

[6] Kim, J., Mo, Y. J., Lee, W., Nyang, D. (2017, August). Dynamic security-level maximization for stabilized parallel deep learning architectures in surveillance applications. In 2017 IEEE Symposium on Privacy-Aware Computing (PAC) (pp. 192-193). IEEE.

[7] Du, J. (2018, April). Understanding of object detection based on CNN family and YOLO. In Journal of Physics: Conference Series (Vol. 1004, No. 1, p. 012029). IOP Publishing.

[8] Yi, Z., Yongliang, S., Jun, Z. (2019). An improved tiny-yolov3 pedestrian detection algorithm. Optik, 183, 17-23.

[9] Warsi, A., Abdullah, M., Husen, M. N., Yahya, M., Khan, S., Jawaid, N. (2019, August). Gun detection system using YOLOv3. In 2019 IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA) (pp. 1-4). IEEE.

[10] Pe´rez-Herna´ndez, F., Tabik, S., Lamas, A., Olmos, R., Fujita, H., Herrera, F. (2020). Object detection binary classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance. Knowledge-Based Systems, 194, 105590.

[11] Van Tuan, N., Nguyen, T. B., Chung, S. T. (2016, October). ConvNets and AGMM based real-time human detection under fisheye camera for embedded surveillance. In 2016 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 840-845). IEEE.

[12] Paul, M., Haque, S. M., Chakraborty, S. (2013). Human detection in surveillance videos and its applications-a review. EURASIP Journal on Advances in Signal Processing, 2013(1), 1-16.

[13] Nikouei, S. Y., Chen, Y., Song, S., Xu, R., Choi, B. Y., Faughnan, T. R. (2018, July). Real-time human detection as an edge service enabled by a lightweight cnn. In 2018 IEEE International Conference on Edge Computing (EDGE) (pp. 125-129). IEEE.

[14] Haritaoglu, I., Harwood, D., Davis, L. (2000). W4: Real-time surveillance of people and their activities. 22 (8): 809–830.

[15] Chen, J., Li, K., Deng, Q., Li, K., Philip, S. Y. (2019). Distributed deep learning model for intelligent video surveillance systems with edge computing. IEEE Transactions on Industrial Informatics.

[16] Vovk, V. (2015). The fundamental nature of the log loss function. In Fields of Logic and Computation II (pp. 307-318). Springer, Cham.

[17] Mao, Q. C., Sun, H. M., Liu, Y. B., Jia, R. S. (2019). Mini-YOLOv3: real-time object detector for embedded applications. IEEE Access, 7, 133529-133538.

[18] Redmon, J., Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[19] Bianco, S., Cadene, R., Celona, L., Napoletano, P. (2018). Benchmark analysis of representative deep neural network architectures. IEEE Access, 6, 64270-64277.

[20] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[21] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[22] Vaddi, S., Kumar, C., Jannesari, A. (2019). Efficient object detection model for real-time UAV applications. arXiv preprint arXiv:1906.00786.

[23] Wojek, C., Walk, S., Schiele, B. (2009, June). Multi-cue onboard pedestrian detection. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 794-801). IEEE.

[24] Olmos, R., Tabik, S., Herrera, F. (2018). Automatic handgun detection alarm in videos using deep learning. Neurocomputing, 275, 66-72.

163

Performance Analysis of ML Algorithms to Detect Gender Based on Voice

Raz Mohammad Sahar [a,1], Dr.T.Srivinasa Rao [a], Dr.S.Anuradha [a] and Dr.B.Srinivasa Rao [a]

[a]*Dept of computer science, GITAM. Deemed to be University, Visakhapatnam, India*

Abstract. Gender classification is amongst the significant problems in the area of signal processing; previously, the problem was handled using different image classification methods, which mainly involve data extraction from a collection of images. Nevertheless, researchers over the globe have recently shown interest in gender classification using voiced features. The classification of gender goes beyond just the frequency and pitch of a human voice, according to a critical study of some of the human vocal attributes. Feature selection, which is from a technical point of view termed dimensionality reduction, is amongst the difficult problems encountered in machine learning. A similar obstacle is encountered when choosing gender particular features—which presents an analytical purpose in analyzing a human's gender. This work will examine the effectiveness and importance of classification algorithms to the classification of gender via voice problems. Audial data, for example, pitch, frequency, etc., help in determining gender. Machine learning offers encouraging outcomes for classification problems in all domains. An area's algorithms can be evaluated using performance metrics. This paper evaluates five different classification Algorithms of machine learning based on the classification of gender from audial data. The plan is to recognize gender using five different algorithms: Gradient Boosting, Decision Trees, Random Forest, Neural network, and Support Vector Machine. The major parameter in assessing any algorithm must be performance. Misclassifying rate ratio should not be more in classifying problems. In business markets, the location and gender of people are essentially related to AdSense. This research aims at comparing various machine learning algorithms in order to find the most suitable fitting for gender identification in audial data.

Keywords. Machine Learning; Gradient Boosting; Decision Trees; Neural network; Support Vector Machine, Gender Classification.

1. Introduction

One characteristic of voice which is deeply perceived in humans is dimorphism. Specific features that differentiate human voices are speech rate, Tone, and duration. Specifically in males and females.[1].

---

[1] Raz Mohammad Sahar, Dept of computer science, GITAM. Deemed to be University, Visakhapatnam Email: raz.sahar2@gmail.com.

The thought of being dimorphic is 98.8% that constitutes the speaker's gender with its frequency. The distinction in gender, nonetheless, cannot make predictions by voice speech. Some voice pitches range between males and females, which makes it tough to identify males and females correctly.

By using Python language, recognition of the gender of a specific speaker is made possible, using methods applied for processing of speech in an asynchronous environment. Vocal cord width is the principal motive with which the difference between genders is calculated. The way a person talks and his present physical states are other motives which contribute to gender difference calculation. In a manner that one can identify a speaker as female or male, these abnormalities are used. Past researches have investigated the distinction between males and females, which includes various variables. Research shows that the major parameter for making speech analysis is derived from frequency, mean frequency, first quantile, and pitch, resulting in identification and classification. Identification of speech helps extract the information regarding gender, age, and accent in which they talk. A tremendous amount of research has been done in this area. Certain speech census is being implemented, which makes use of over period, maximum value, and mean to identify gender.

The audial data is converted into various parameters such as the strength of vocal, first quartile, third quartile, frequency, kurtosis, Spectral Flatness, spectral entropy, and so on. the previously mentioned parameters are trained and tested using various algorithms in ML to identify any genders.

In this paper, a comparative algorithmic approach that identifies gender based on different classification algorithms is proposed. The Predictions are made fundamentally on the dataset where values would be processed coming from audio files. Comparatively, results received are brought into comparison with earlier prediction results, and estimation is done by using classification algorithms to decide which algorithm gives better outcomes in identifying gender-derived specific parameters. Precise prediction of how the comparative algorithmic approach recognizes the gender-derived from the mentioned algorithms is obtained.

2.   Literature Survey

Classification of gender, processing, and gender-derived identification is being carried out for a considerable period of time. Some theories used emerged overtime to carry out gender identification. New research based on the identification of gender and identification shows that speech is turned into various parameters. The major parameters are pitch and frequency. Identification is being carried out to distinguish males and females. Firstly, the system is equipped with training data, and then data is presented and assessed for the system's outcome for the data. These results collected may differ for various algorithms and seem to give inconsistent outcomes at different periods. Dimensionality reduction is amongst the major problems encountered in machine learning. An identical obstacle is encountered when selecting gender-specific characteristics—that assist an essential purpose in the classification of the gender of an individual. This work will examine the effectiveness and importance of ML algorithms to the voice-derived gender classification problem. Gender-based identification using F0 frequency [1] This says that Random Forest befits better for speaker identification using pitch and fundamental frequency to distinguish males and females [2]. They are

further tuning based on a bucketing method to increase the effectiveness of outcomes achieved.

Voice-derived word extracting laboratory view [3] predicts that the algorithm runs finer for this identification, and it gives promising results to do extraction of vowels in samples of males. Since all samples are being trained then tested, this effectively gives a solution. This is noticed that raising the inexpressible portion in speech related to 's' pitch's value sound rises, obstructing gender exposure in males. Likewise, booming the sound portion of the speech such as 'a' reduces the value of pitch. It doesn't recognize it the instant the speaker delivers 2 varieties of tones. Identification of speech in grown-ups confirms that they can be casual, and vocal length changes and seem girlish since it is not easy to distinguish the males and females.

Certain voices of the female are challenging to investigate on how high or low it is [4]. For instance, measuring a female voice from a single aspect may hardly satisfy all necessities. This paper's (pitch) high or low sound quality in males and females [4] suggests that the voice of females should be recognized with different variables than males like Emotional, Shrill, and Swoopy. This includes parameters in which female persons can vary from each other. Therefore, preprocessing of the dataset has to be done on this before classification of gender. just as the pitch perception. [5-6]. F0 Fundamental frequency contains a compound of dialectal and non-dialectal speakers' information, and they both correspond to males and females, and it is dependent on the speaker's upper pitch and sound. [7] This guided to put a frequency f0 with not having any range experience and non-syllable outer information. It proves that the speaker's voice changes between upper and lower pitches between speakers.

Gender identification by support vector machine [8] states that gender's speech is examined by several speech techniques such as compression of speech, talking on the phone, and distinction in languages, etc. It shows that pitch of male voice, duration, and Mel of frequency approximately 100- 146Hertz and females of around 188-221Hertz. At this point, voice is classified using frequency, and it is obtained and examined. Gender Identification by Voice [9] states that gender recognition using Linear discriminant analysis (LDA) performs well. Nevertheless, even with this model, the test error rate is still larger than 10%. Gender classification by pitch analysis [10] shows that gender identification gives promising results using pitch and signal energy.

3.   Algorithms

3.1. Gradient Boosting algorithm

It is ensemble learning. The main concept in gradient boosting is that models are made in series. Gradient boosting uses many weak learners; it converts the week learners into strong learners. Boosting algorithm uses a decision tree as a weak learner then, the errors are obtained at lead nodes it obtains the errors at the leaf nodes and creates a second new tree by using values of error as values of new observation, and this process goes on for a defined number of times. The gradient boosting algorithm is shown in figure 1. Then the evaluation of the trained model is carried out against testing data so the calculation can be done to determine evaluation metrics.
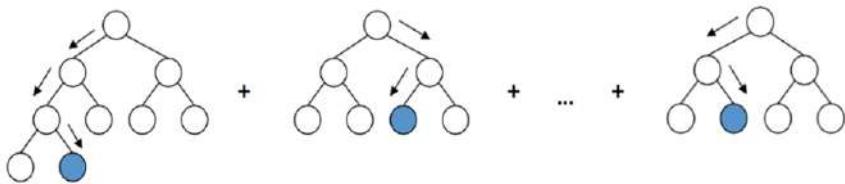
Figure 1.Gradient Boosting

3.2. Decision Tree

Decision Tree A decision tree is a classification algorithm, and it is supervised learning. A Decision tree consists of nodes, leaf nodes, and edges. Every decision tree is a structured tree in which every internal node shows an attribute as well as a feature. Here every decision rule is represented by branch, and every leaf node denotes the output or target variable. the decision tree is one of the most powerful famous tools for classification and prediction, and the decision tree gives good performance in classification related problems.

3.3. Random Forest

The random forest model is an ensemble algorithm. Ensemble algorithm combines algorithms of similar type or different type for the classification of objects or variables. Here classifier makes decision trees from the training dataset. It then checks or calculates the votes from various decision trees to choose the target variable or test object's final class. The features of random forest are more efficient working on large data sets, handles on more variables, estimates which variables are important during classification, estimates missing data, and so on.

3.4. Support Vector Machine

A Support Vector Machine is a classifier that is defined by separating with a hyperplane. It comes under supervised learning, and the hyperplane classifies the target variables or data points. The hyperplane is used for data points categorization. Data points that are on any side of the hyperplane create their own classes.

3.5. Neural Network

Neural network can be described as series of algorithms which aim to recognize underlying relationships in any data set by mimicking how the human brain functions. So, neural networks are related to neuron systems, both organic or artificial naturally. Neural networks can regulate to varying input; therefore, the network produces the greatest feasible result without redesigning the criteria of output. The idea of neural network is that it has its origins in artificial intelligence rapidly gains a reputation in the development of trading systems.

4.   Pre-processing

4.1. *Dataset*
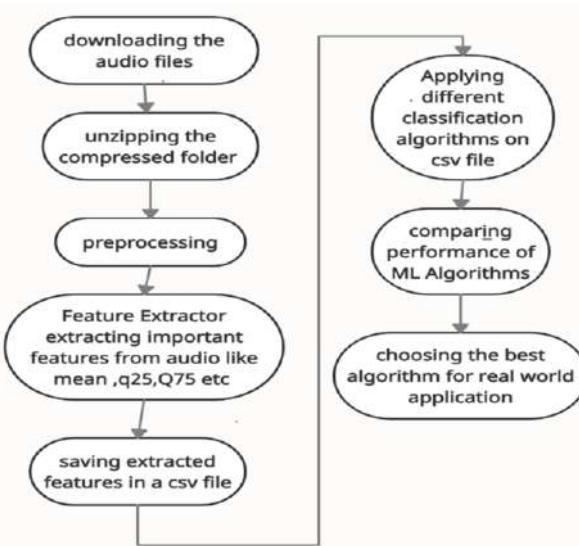
The dataset having 62,450 audios samples zipped (tgz) contains ten files can be automatically downloaded from this url= http://www.voxforge.org/ website.

4.2. *Feature Extraction*

First, all audio file's contents are read; after the essential properties are extracted and stored into a CSV file and, one can parse the README files to extraction metadata: like age, gender, and the pronunciation of the speaker. Reputably, Python comes with the package Scipy wave file is employed to obtain the audio, Scipy stats to do extraction of the prominent features, and NumPy and its FFT (Fast Fourier Transform) and fftfreq to extrapolate the audio data files to frequencies. All wav files' data are registered as amplitude in the domain of time, but the likely exciting features are those which come with a greater discriminative power male/female frequency.

In order to turn the audio to frequencies, DFT should be used, mainly the FFT algorithm. Implementation. Fourier transform receives a signal in the domain of time (set of measurements over a period of time), and it is then converted into a spectrum— a Group of frequencies with equivalent (compound) values. The spectrum never holds any information regarding time! In order to get the frequencies as well as the time at which recording was done for them, a signal of audio is divided in slight, OS (overlapping slices), and FT is applied on all (short-time Fourier transform). np.fft.fft delivers a compound range np. fft. fftfreq delivers the sample frequencies. A sample Directory consists of 10 audio tapes from a specific speaker.

Since Every wave file in the Directory is processed so that the dominating frequencies with 200ms windows sliding (1/5th of one second) can be extracted, when a wave file is 4 Secs long, extraction of a list having 20 frequencies will be done. For a sample folder (user), ten lists corresponding to the ten wav files (having twenty frequencies each) shall be collected in a list of lists. The frequencies are to be filtered to hold values in the speakers' voice within 20 hertz < frequency < 280 hertz. Also, any values in the range 50Hz are possible noise that should be filtered. fig2 shows the flow diagram of the model. as it can be seen in flow diagram, the first the audio files are automatically downloaded, then all the download audio files are unzipped, all the important voice features are extracted, then the extracted features are saved in a CSV file, five different algorithms: Gradient Boosting, Decision Trees, Random Forest, Neural network, and Support Vector Machine are used to classify the gender of a speaker, performance all the five classification algorithms are compared, and the algorithm with the best accuracy is chosen for real-world application.

Figure 2. flow diagram

4.3. Data Description

After feature extraction, 3000 audio samples of voice are obtained, as every audio sample is an. AIF file. Preprocessing of AIF format files has been carried out for audial analysis using feature extraction. So, by applying the feature extraction technique, around 22 features can be obtained from acoustic signals. The extracted and preprocessed will be stored into a file consisting of 3168 rows and twenty-one 21 columns. Using 20 features, the prediction of the output label is made. The splitting of total data is done into parts, i.e., train data set and test dataset.

acoustics features - 20 Acoustics features are listed below, and Acoustic Properties with Distribution is shown in figure 3.

- Frequency Standard Deviation
- Mode of Frequency
- Median of Frequency
- Q25 Lower Quartile
- Q75 Upper Quartile
- Kurtosis
- Frequency Centroid
- Spectral Flatness
- Spectral Entropy
- Average of Fundamental Frequency Measured Across Acoustic Signal
- Minimum Fundamental Frequency Measured Across Acoustic Signal
- Maximum Fundamental Frequency Measured Across Acoustic Signal
- Minimum of Dominant Frequency Measured Across Acoustic Signal
- Average of Dominant Frequency Measured Across Acoustic Signal
- Maximum of Dominant Frequency Measured Across Acoustic Signal
- Maximum of Dominant Frequency Measured Across Acoustic Signal

- Modulation
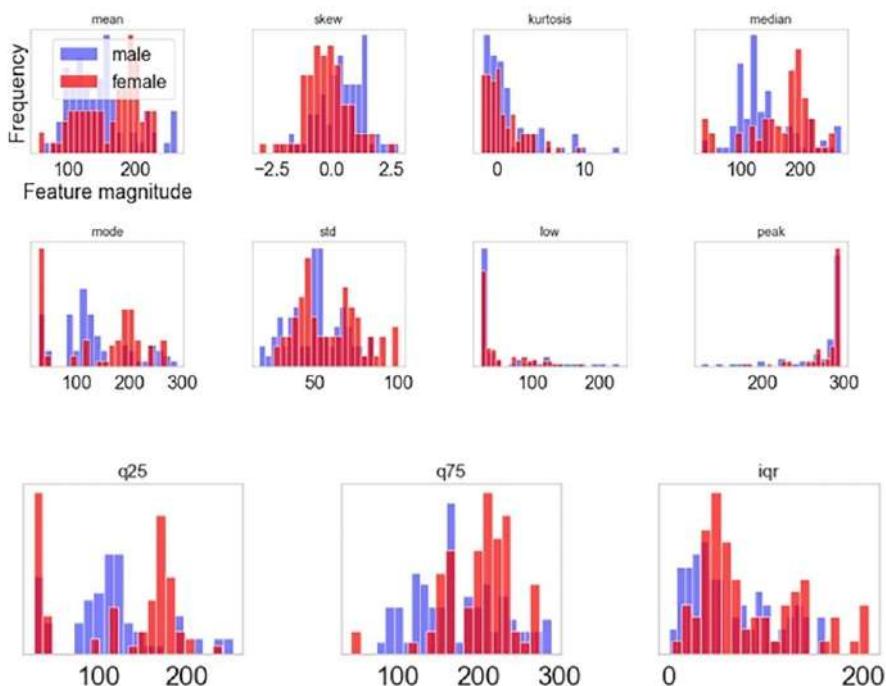
Figure 3. Acoustic Properties with Distribution

5.   Results and Discussion

The problem of gender identification based on voice is addressed in this paper. The Mel-frequency Cepstral Coefficients of voice samples are commonly used as gender recognition features. The MFCC-based identification, on the other hand, is highly complicated. This paper proposes a comparative gender classification model that uses frequency pitch, first quartile, etc., to ensure effective gender classification while keeping the system simple.

The audio dataset used for this research can be downloaded from the http://www.voxforge.org/, The models are trained and tested with around 3000 males and female's audio files. The splitting of complete datasets is done into two datasets, testing and training dataset.30% of the entire dataset is considered for the test dataset. As the more influential the training dataset, the greater the model's result. If we train our model with an extensive dataset of training, it is more likely to get vital patterns. This paper evaluates the performance of five ML algorithms to identify gender and the results obtained is given in table 1, the results show that Gradient Boosting gives good Accuracy, the importance of independent and dependent variable for each algorithm is

given in figures 4,5,6, it shows that the first quartile is the most important feature in the classification of gender.
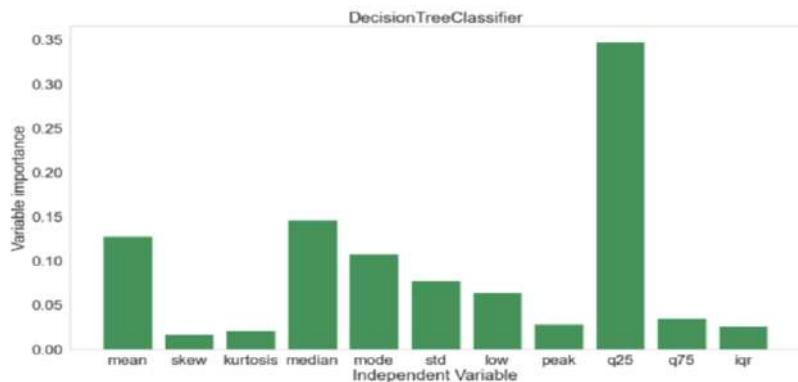
Figure 4. features importance in decision Tree
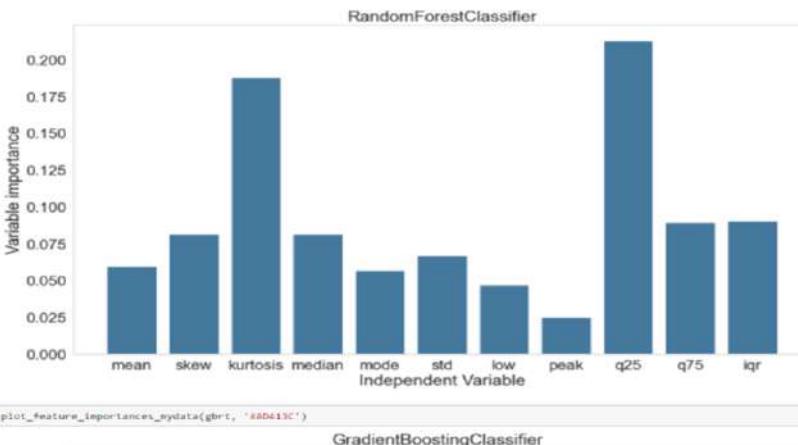
Figure 5. features importance in Random Forest
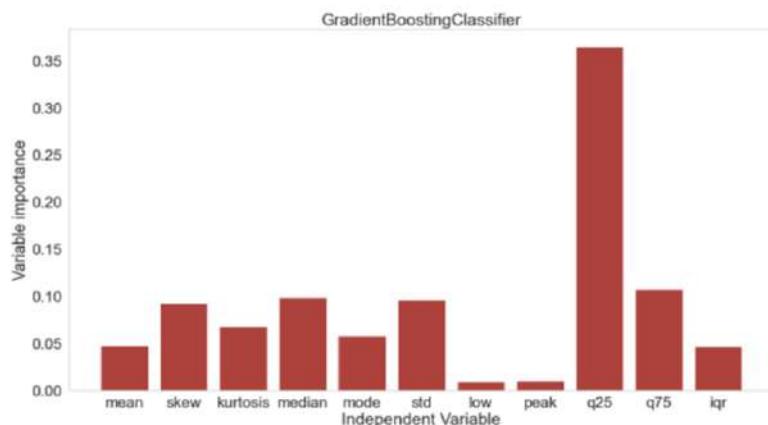
Figure 6. features importance in Gradient Boosting

Table 1. Comparison of results of different algorithms

Algorithms	Accuracy	Precision	Recall	F1 score
Random Forest	89%	81%	95%	88%
Decision Tree	82%	77%	79%	64%
Support Vector	88%	82%	87%	75%
Neural Network	89%	83%	90%	86%
Gradient Boosting	90%	82%	95%	88%

6.   Conclusion

The results obtained show that the Gradient Boosting algorithm gives promising results in gender classification. Neural network and SVM also give relatively good results. The obtained results using ML classification algorithms are just for the dataset preprocessed from the 3000 audio files, and the result may change for other datasets. Gradient Boosting has higher accuracy than other algorithms in classifying gender. One of the difficulties in classification is that the collection of samples of the audio is generally done out of loud and noisy environments, limiting classification accuracy. More effective techniques to reduce and eliminate noise can be found out, which can be considered a route for future research.

References

[1]   Ericsdotter, C., & Ericsson, A. M. (2001). Gender differences in vowel duration in read Swedish: Preliminary results. Working papers/Lund University, Department of Linguistics and Phonetics, 49, 34-37.

[2]   Whiteside, S. P. (1996). Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences. Journal of the International Phonetic Association, 26(1), 23-40.

[3]   Byrd, D. (1992). Preliminary results on speaker-dependent variation in the TIMIT database. The Journal of the Acoustical Society of America, 92(1), 593-596.

[4]   Henton, C. G. (1989). Fact and fiction in the description of female and male pitch. Language & Communication.

[5]   Bishop, J., & Keating, P. (2012). Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and speaker sex. The Journal of the Acoustical Society of America, 132(2), 1100-1112.

[6]   Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. The Journal of the Acoustical Society of America, 118(5), 3177-3186.

[7]   Muhammad, G., Alsulaiman, M., Mahmood, A., & Ali, Z. (2011, July). Automatic voice disorder classification using vowel formants. In 2011 IEEE international conference on multimedia and expo (pp. 1-6). IEEE.

[8]   Gaikwad, S., Gawali, B., & Mehrotra, S. C. (2012). Gender Identification Using Svm With Combination Of Mfcc. Advances in Computational Research, 4(1), 69-73.

[9]   Jena, B., &Panigrahi, B. P. (2012). Gender classification by pitch analysis. International Journal on Advanced Computer Theory and Engineering (IJACTE), 1.

Pervasive Computing Applications and Security – A Deep Insight

S. Magesh [a,1], Sujatha Jamuna Anand [b], Niveditha V.R[c], Y. Pavan Kumar Reddy [d],
P.S. Rajakumar [c]
[a] Maruthi Technocrat E Services, Chennai, India
[b] Loyola Institute of Technology, Chennai, Tamil Nadu, India
[c] Department of Computer Science and Engineering, Dr M.G.R Educational and
Research Institute, Chennai, India
[d] Geethanjali Institute of Science and Technology, Nellore, Andhra Pradesh, India

Abstract. Pervasive computing has made life easy with communication devices. Today devise collaboration has enhanced everywhere in this environment. It has made computing devices invisible and the services. This pervasive framework provides applications with interactions, numerous cooperation and accessibility, and integration. The proposed work enumerates the applications, pervasive security challenges. It provides security predicaments by assigning certificate credentials, access controls, trust management, and some security techniques to overcome the security paradigms in these distributed networks with IoT and the pervasive computing framework. The work also encounters security perplexities in handling the security threats and user interaction issues. Nevertheless, security techniques are listed for various pervasive applications in distinct domains such as healthcare, industries, and transforming sensitive information. The smart applications with smart environments perhaps force towards the new technologies in the pervasive computing outlook. The work also embedded with middleware with the context-based situation in these pervasive applications

Keywords. Pervasive computing, Pervasive security, Certificate credentials, distributed networks, trust management.

1. Introduction

Pervasive computing has become the growing trend of embedded computing systems and promoted with smart technologies such as smart health, smart home, smart environment in our everyday entities [1]. This pervasive environment provides effective communication and performs efficient tasks to minimize the end user's need to interact with computing technologies [2]. It can be accessed irrespective of the environment across any network handling different tasks from one node to another as it is portable. Pervasive computing systems are connected in a heterogeneous network environment everywhere with every smart device [3]. Pervasive computing has its potential in various domains such as defense, finance and healthcare. As these domains are accessing through laptops, personal digital assistants (PDA) and other smart devices, there are various difficulties at various levels [4]. According to Mark Weiser's point of view [5,6], pervasive computing is said to the next generation computing

---

[1] S. Magesh, Maruthi Technocrat E Services, Chennai, India,
 Email: mageshmtech@gmail.com.

environment provided with communication anywhere at any time for anybody. The proposed work is focused on

- Pervasive Applications in various domains with convenient access.
- Assigning security credentials to every individual deploying pervasive devices.
- Security attacks and security requirements to access the device correctly allow authorized access to modify access rights to third parties.

2. Pervasive Environment and Applications

Pervasive computing has simplified human lives in day-to-day activities by providing Smartphone users to carry out their tasks through portable and embedded computing devices. In recent days, pervasive computing has entirely changed the lifestyle while interacting with information. The multiple connected devices, as well as the environment, had demanded robust and secured information systems. These systems are embedded with various network hubs providing continuous and reliable connectivity with smart devices in an IoT environment. It focuses on sensor data providing better communication with both mobile and distributing environment [7]. The pervasive computer architecture of sensors and embedded devices is represented in **Figure 1** with various sensors embedded with IoT applications connected with low-cost device including the mobility service and connectivity. Tablets, smartphones, smart homes, wearable devices, and sensors have all evolved due to the pervasive environment.
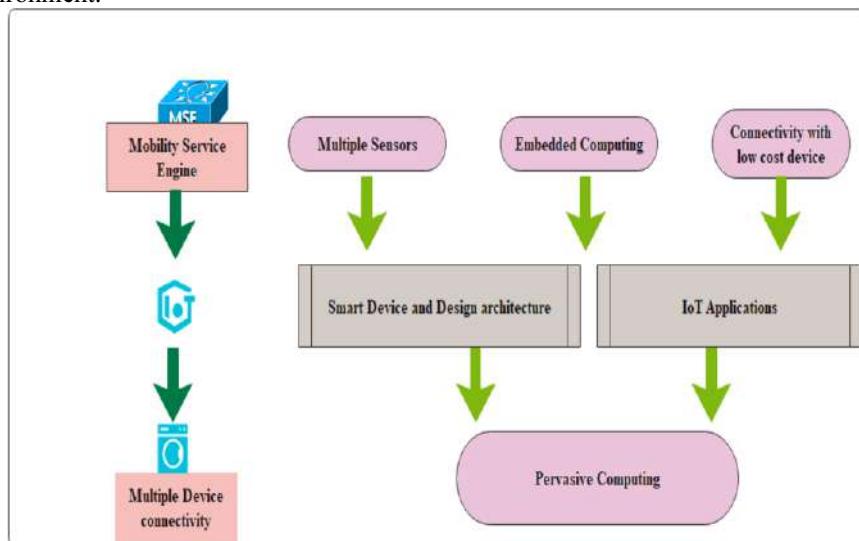
Figure 1. Pervasive Computing Environment

From simple tasks like switching the lights in the workplace, lecture and convention rooms, office cabins, and updating their emails to more complicated tasks like booking plane tickets, stock purchasing and sale, and managing banking accounts, this world has made life more accessible to the humans. By integrating these smart devices with smart environment, a new environment is launched, supporting any device

with any environment that produces the pervasive environment **(Figure 2)** representing various smart devices in smart environment.
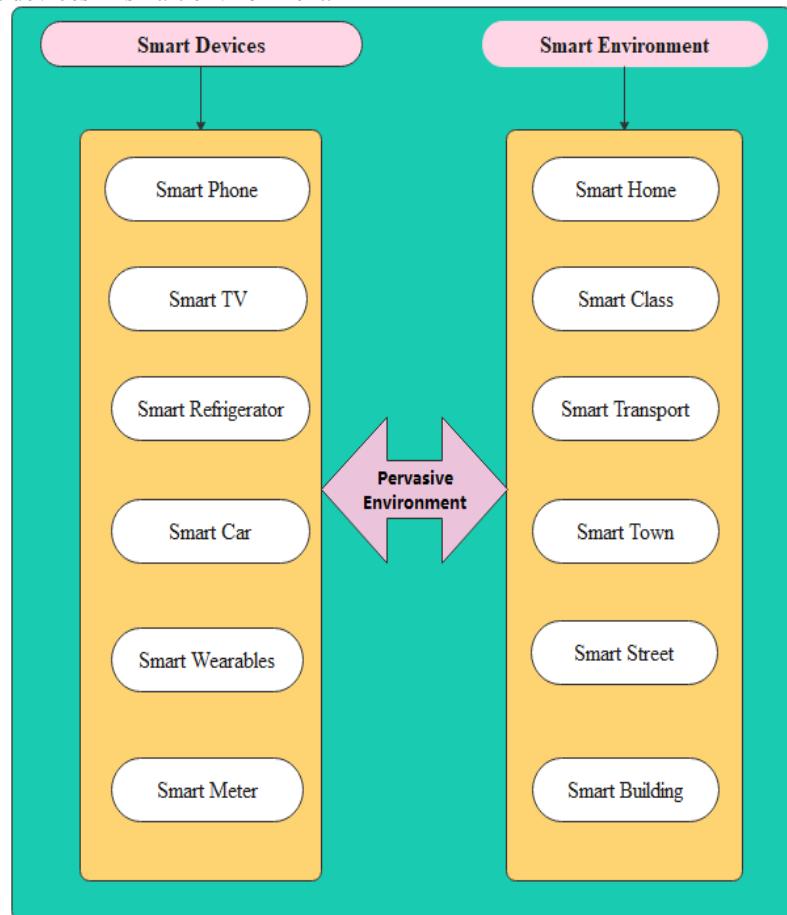
Figure 2. Various Applications of Pervasive Computing

3.   Security Challenges in Pervasive Environment

Providing security to such a widely distributed model leads to face many challenges at a certain level. This can be related to a scenario like a person who does not belong to the enterprise has access to that concern. There are several problems in providing security in this pervasive environment. The significant challenges identified are change of user data, hacking client-server information, eavesdropping, privacy loss, smart device theft, stealing sensitive data, memory isolation, data forgery, original data alteration, economic issues, utilization of unpermitted connection, exploiting sensitive data. Few additional issues on protecting the system are as follows:

- **Inconspicuous and Invisibility**: Generally, the pervasive computing seems to be unnoticeable as data are sent and received daily. This diminishes the environment making easy acceptance that paves the way for attacking the user privacy.

- **Creation of Smart Spaces**: Sensors and other devices are integrated to sense, analyze and learn every area. It is also tractable that leads to getting the user's intention with privacy risks.
- **Other Privacy Issues**: Augmentation of dynamic spaces with actuators and sensors has provided brilliant spaces and computing efficiency. The utilization of different sensors and devices with dynamic space customized for users could cause a serious vulnerability to privacy, and malicious attackers and intruders can mislead this gap. Specific security attacks also affect the pervasive environment [8]. Those attacks include attacks due to modification, impersonation, and flooding attacks. **Figure 3** addresses some of the main issues in pervasive security.
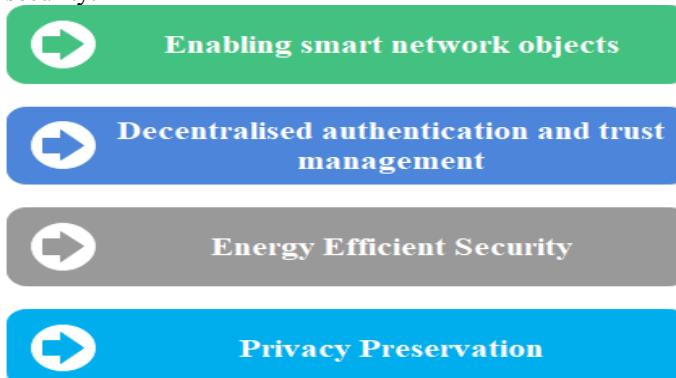
Figure 3. Security issues in Pervasive Environment

- **Attacks due to Modification**: In these attacks, packet headers are modified by sniffers and affects the integrity and the availability of the message [9].
- **Attacks due to Impersonation**: This is one type of eves dropping, providing integrity issues. It is classified under ARP spoofing attacks [10] that masquerades the connected device's destination address, and a duplicate MAC address is inserted.
- **Flooding Attacks**: Unwanted or undesired messages are sent frequently to their neighbor nodes, causing overhead. Dos, Routing Table overflow are classified under these attacks.

4. Security Predicaments

Pervasive security is an essential security technology to promote the progress of security program, including policies. The requirement of pervasive security is described below:
- **Clarity with compliance:** the main aim of security implementation is to ensure compliance by monitoring and improving users at a higher organization. In this context, security needs to be more apparent without causing aggravation to the users.
- **Managing Security Policies**: Providing a delegation certificate is an essential security requirement for access privilege. The security agent is responsible for controlling the service. At the user's request, the security agent transmits the credential with a delegation or ID certificate. The agent may also generate an

authorization certificate to use as tickets for accessing the resource. **Figure 4** shows the pervasive scenario in the health monitoring system illustrating the acquisition of patient's database using the sensor network, filtering the data using preprocessing methods, analyzing the data using queries and then diagnosing accordingly.

- **Ensuring Multilevel Layer:** The security architecture should ensure security services at different levels within the available resources, strategies, domain and context middleware.
- **Interoperability and Scalability:** Utilization of various security technologies in recent years, multiple security measures need to be implemented, and security services need to be designed so that it needs to be portable and accessible at all levels providing inbuilt devices. Besides encoded actions, delegations and privileges, an open environment built on an XML framework named DAML (DARPA XML) Agent are well suited in an open environment. A security tool named Diasuite allows defining the taxonomy of a particular application with prewritten specification [11]. The privacy of the users can be protected by replacing X.509 signatures, including the authorized person's designation.
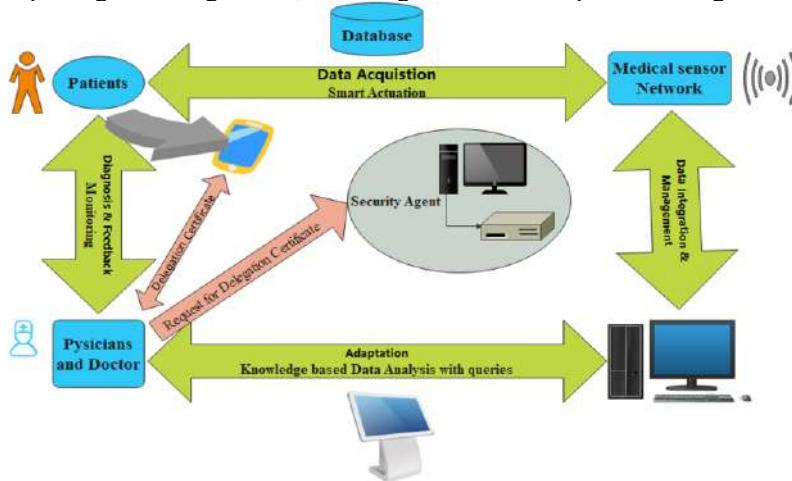
Figure 4. **Health care Monitoring in Secured Pervasive Environment**

The security predicaments need to ensure the security challenges such as Authentication and access control Confidentiality Integrity and Availability. Authentication and access control are the two security aspects to be ensured for every specific user.

5.   Discussions

Pervasive security is simply establishing the quantitative and qualitative parameters by defining the reliability, topology, some failure patterns and efficiency in designing the architecture. Specific user-centric parameters need to evaluate to monitor the system's behavior towards the intension of the users' expectations and how the system takes efforts on the users' part relying on the system. Pervasive security works everywhere, irrespective of the network resources for data transmission. The entities such as

business, technical and policies need to be secured. Today Pervasive computing has expanded transparently in many environments. Providing protection against unauthorized users, preventing access by unverified methods, accessing facilities and denial of service for unauthenticated users are some of the services rendered by Pervasive computing.

6. Conclusion

Today security for pervasive computing has become a complicated paradigm than exploiting the pervasive environment. The proposed study discusses the application of smart devices in a pervasive environment, pervasive security challenges and privacy issues. To overcome those paradigms, the study also enumerates some pervasive security requirements and provides solutions to security attacks by evaluating the security parameters in the pervasive environment. The main focus of the paper lies in the security issues in the smart environment and addressing the security predicaments in the distributed environment. Nevertheless, smart devices are deployed by many domains with ease and comfort. Due to its insecurity, higher-level protocols need to be enhanced in the environment soon.

References

[1] Ahmed, A.W., Ahmed, M.M., Khan, O.A. and Shah, M.A., 2017. A comprehensive analysis on the security threats and their countermeasures of IoT. International Journal of Advanced Computer Science and Applications, 8(7), pp.489-501.

[2] AbdAllah, E.G., Hassanein, H.S. and Zulkernine, M., 2015. A survey of security attacks in information-centric networking. IEEE Communications Surveys & Tutorials, 17(3), pp.1441-1454.

[3] Usman, A.B. and Gutierrez, J., 2018. Toward trust based protocols in a pervasive and mobile computing environment: A survey. Ad Hoc Networks, 81, pp.143-159.

[4] Hooda, M., Pathak, S. and Yadav, B., 2017, September. Pervasive Security of Internet Enabled Host Services. In 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC) (pp. 784-789). IEEE.

[5] Bdiwi, R., De Runz, C., Faiz, S. and Cherif, A.A., 2017, July. Towards a new ubiquitous learning environment based on blockchain technology. In 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT) (pp. 101-102). IEEE.

[6] Dharminder, D. and Mishra, D., 2020. LCPPA: Lattice-based conditional privacy preserving authentication in vehicular communication. Transactions on Emerging Telecommunications Technologies, 31(2), p.e3810.

[7] Gollagi, S.G., Math, M.M. & Daptardar, A.A. A survey on pervasive computing over context-aware system. CCF Trans. Pervasive Comp. Interact. 2, 79–85 (2020).

[8] Naik, A.S. and Murugan, R., 2018. Security attacks and energy efficiency in wireless sensor networks: A survey. International Journal of Applied Engineering Research, 13(1), pp.107-112.

[9] Ladas, A., Deepak, G.C., Pavlatos, N. and Politis, C., 2018. A selective multipath routing protocol for ubiquitous networks. Ad Hoc Networks, 77, pp.95-107.

[10] Bdiwi, R., De Runz, C., Faiz, S. and Cherif, A.A., 2017, July. Towards a new ubiquitous learning environment based on blockchain technology. In 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT) (pp. 101-102). IEEE.

[11] Cassou, D., Bruneau, J., Consel, C. and Balland, E., 2011. Toward a tool-based development methodology for pervasive computing applications. IEEE Transactions on Software Engineering, 38(6), pp.1445-1463.

Price Prediction for Pre-Owned Cars Using Ensemble Machine Learning Techniques

Chetna Longani[a,1], Sai Prasad Potharaju [a], Sandhya Deore [a]

[a]*Dept of Computer Engineering, Sanjivani College of Engineering, Kopargaon,*
Maharashtra, India

Abstract. The Pre-owned cars or so-called used cars have capacious markets across the globe. Before acquiring a used car, the buyer should be able to decide whether the price affixed for the car is genuine. Several facets including mileage, year, model, make, run and many more are needed to be considered before getting a hold of any pre-owned car. Both the seller and the buyer should have a fair deal. This paper presents a system that has been implemented to predict a fair price for any pre-owned car. The system works well to anticipate the price of used cars for the Mumbai region. Ensemble techniques in machine learning namely Random Forest Algorithm, eXtreme Gradient Boost are deployed to develop models that can predict an appropriate price for the used cars. The techniques are compared so as to determine an optimal one. Both the methods provided comparable performance wherein eXtreme Boost outperformed the random forest algorithm. Root Mean Squared Error of random forest recorded 3.44 whereas eXtreme Boost displayed 0.53.

Keywords. Ensemble, eXtreme Gradient Boost, Random Forest, Regression

1. Introduction

The prices of new cars are tuned by different facets as in manufacturer's cost, dealer's margin, transportation charges, GST levied on the car. More upon, the worth of a car turns down every year by 20%. Majority of the population cannot afford to buy a car from the showroom, rather prefer buying a pre-owned car. The global market for used cars valued USD 1332.2 billion in the year 2019. Market is looking forward to a stretch with a CAGR (Compound annual growth rate) of 5.5% during the next decade [3]. The used car market in India valued USD 24.24 billion during the same year. Like the global market, Indian used car market is also forecasted to register a CAGR of 15% by next five years [1]. Rapid spread of Covid-19 across the globe too has resulted in a downturn on public transport, which further has increased in demand for the pre-owned cars. A number of online sites have assembled the automotive industry at one place, so that the end user can buy or sell with a click. These sites use different algorithms for generating the price for the used cars, hence may place incompatible results to the users. More upon, these systems provide the sell and purchase mostly in urban areas. A unified algorithm must be entailed for regulating the price. This paper aims at designing a system that uses ensemble machine learning techniques to develop models

---

[1] Chetna Longani, Dept of Computer Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra, India
Email:chetnadang16@gmail.com.

That can predict prices for used cars. The model is trained for data of Mumbai region. The dataset used for the project [4] has 2454 records.

Prediction is nothing but estimation made out of observations. It uses observable phenomena so as to make future projections for the scene. The historical data forms our observations; this historical data is so large in volume that it is tedious to make conclusions just by looking at the data. Manual interpretation of the data is a soulless task. To reduce this pain and to get future predictions at a fingertip, different machine learning algorithms have to be deployed.

Machine learning when used for prediction, an algorithm is employed to train historical dataset and output a model. This model is fed with unseen data, and the prospect of a particular outcome is forecasted. Ensemble methods of machine learning blend different base models, thereby combining decisions from multiple models and providing an optimal model for prediction. Ensemble techniques use a mass of models into an account, further averages these models and delivers a proper model that can be used for prediction. In the proposed system, a random forest algorithm is deployed for predicting the prices of used cars. Further, a model is built using the eXtreme gradient boosting algorithm and prices of used cars are predicted. Performances of these two models are compared.

Various aspects that affect the price of pre-owned as in year of purchase, run of the car in terms of kilometres, its showroom price, its mileage, engine capacity, seating capacity, power capacity of car battery are considered while building the model. Other factors as in whether the seller is the owner or a dealer, is the owner first or second owner of the car, is the gear manual or automatic, the fuel used is petrol or diesel are also accounted for building the model. Fair number of perceptible attributes is employed. The motto of the developed system will be achieved if both buyer and seller will accomplish the deal at a fair price. One can buy or sell his car in a short time if he finds a fair price for his car in no time.

The organization of the paper is as follows. A review of relevant work is provided in the next section. The methodology is described in section III. Section IV includes the comparison of both the ensemble machine learning techniques to predict the car prices. Finally, the paper ends with a conclusion and setting a trend for future work.

2. Literature Survey

In the literature, few researchers applied various machine learning techniques to predict the car cost as per the given requirements. In the research article [5], the author investigated the application to predict the cost of cars which are used in Mauritius city. Author used techniques like multiple linear k-nearest neighbours, naïve bayes and decision trees, regression analysis to make the predictions of car cost. He has used multiple linear regression analysis to find out correlations between different features. In KNN, the author has taken only three attributes as marks, year and cylinder volume to classify new samples. Using this, the author wants to confirm that cars with higher values for cylinder volume have higher normalized values than cars with lower values. The main weakness he found was that continuous values with output classes are not that much controlled by decision trees and naïve bayes. Hence, the author classified the price attribute into the range of price classes. The prediction is done on two car types where samples are taken from old newspapers.

The researchers applied three machine learning technologies namely, Artificial Intelligence (AI), Support Vector Machine (SVM) and Random Forest (RF) separately [6]. Authors collected data through different web portals. They used RF for classification and regression. ANN was used for adjusting the weights between neurons. In SVM, the model is trained in such a way that the input data is classified into two categories. With these three single machine learning algorithms, authors found that these are not reliable methods for prediction of car prices. Then authors used the ensemble method that combines three ML algorithms to classify the price of a car as cheap, moderate and expensive.

Wu et al. [7] conducted a study on car price prediction. For this they used adaptive neurofuzzy inference systems (ANFIs). Three attributes were considered for forecasting the price of car as marks on the car, engine style and manufacturing year. They also did the study of ANN with back propagation and compared it with ANFIs. Through their work, they concluded that ANFIs have more accurate probability to forecast car prices than ANN with BP. The model is trained to generate relationships between any input and the outputs of complex systems, and also advises on where to sell the car. K-nearest neighbor machine learning algorithm is used for forecasting.

In the research thesis [8], the author focused on depreciation rate variation between cars with hybrid engines and those with traditional engines. Multiple variable regressions were used to analyze each independent variable on car resale values. Six segments selected by author from car and drive.com for collecting data. Through Correlation matrix age, miles, make and mpg(miles per gas) factors were used for prediction.

Author Nabarun Pal et al. in paper [9], used random forest, a supervised learning method, to forecast the prices of used cars. By selecting the most correlated feature, the model can predict the price of cars accurately. Researchers have tried both linear and random forest regression methods. From their study, they found that random forest is better than linear regression. A grid search algorithm was used to find the optimum number of trees. From that they concluded that with 500 decision trees forest accuracy is good. In case of a regression problem, they took the maximum number of features from the input data set. In case of classification, they took Square root of some features. Hence the Problem they converted into a regression problem as random forest is used for cost prediction.

Ning Sun et al. in paper[10], to analyze the price for each type of vehicle they used the optimized back propagation neural network algorithm. In order to improve the convergence speed of the network topology and also to improve accuracy of the prediction model, a back propagation neural network algorithm is used. The LB-MCM method is used for selecting the number of hidden neurons. Using this method the speed of network structure gets improved and the neuron selection efficiency also gets improved. The deviation and weights of the network gets updated and trained by a back propagation algorithm in such a way that the output vector gets closed to the expected vector. When the number of iterations reaches the highest thresholds and misclassification rate is less than the given threshold then training gets terminated.

Extreme Gradient Boosting or XGBoost [11] is one of the most popular machine learning models in current times. ForeXGBoost is the technique for prediction which uses the sliding window to extract historical sales and production of data features. In this method, leverage parallel computing is used to reduce the training overhead. XGBoost has such features so that it can significantly improve the regularization, parallel processing and in the other data prediction and classification

tasks, such as web text classification, malware classification, sales prediction, customer behaviour prediction, rate forecast and product classification.

3. Methodology

Data collection is the most prime step for any project. We have designed the system for used cars in the Mumbai region, for which the data of used cars is collected using [4] as on 15-March-2021. Total 4057 records were scrapped using the Beautiful Soup (BS4) package. The data fetched composed of null records as well. After the null records are curtailed from the data, 2454 records are retained. The features captured for each car include Year (year of purchase), Seller Type (Dealer/Individual), Driven, Nonowners (First/second/third or above owner), Fuel_Type(Petrol/Diesel/CNG), Gear Type(Manual/Auto), Used_years_2021, Kmpl(Kilometres per litre), Engg Capacity (Cubic capacity), Max Power(Brake Horse Power), Seat_Capacity(4/5/8 seater), Onroad_Price(in lakhs), Sell_Price, Model(Brand). Sell Price is the dependent variable.

We used ensemble machine learning techniques to implement the system. Ensemble techniques build multiple models, and then blend them. Thereby produces upgraded results than a single model would. We can train an ensemble and further use it to make predictions. Hence, an ensemble is a supervised learning algorithm. Using different ensemble methods, we can combine various models, thereby, moving on the path of achieving better accuracy. Suppose that you have designed an android application, before making it public you wish to know its ratings. What you can do is either ask your friends or family or colleagues to rate your app. This process would give you limited feedback. How about cumulating the reviews from fifty or more people who could be your family, friends or even strangers? Now the response that you will seek will be more assorted as the people in the closure possess different skills. This task of accumulating feedback will give you honest and accurate ratings. We can here conclude that a group of miscellaneous people make better decisions as compared to an individual. And the same is true if rather than using a single model, we use a group of diverse models. To achieve diversification in machine learning, we have ensemble techniques. Bagging, boosting, stacking and voting are famous ensemble methods. Bagging configures one ensemble model by deploying Bootstrapping as well as Aggregation, where bagging replaces observations from original datasets and creates multiple subsets holding observations. Further, a base model is fabricated for each of these subsets, which is run in parallel for each of these subsets, independent of each other. At the end, the results from these models are aggregated to give a final prediction. Unlike bagging, the boosting process works sequentially on models. Here, different models are erected; each of the subsequent models corrects errors of the previous model. The weighted mean of all the models produces the final outcome, thereby combining weak learners to form a strong learner. Each of these models contributes to boost the performance of the ensemble. Stacking is an ensemble technique that uses predictions outputted from multiple models to construct a new model, which is further used to make predictions on the test set. Random Forest is a bagging algorithm whereas XGBoost is a boosting algorithm; these algorithms are used to implement the proposed system.

Random Forest algorithm is a popular supervised machine learning algorithm that relies on the concept of ensemble learning and can be deployed for both classification and regression problems in machine learning. It operates by fabricating a

number of decision trees during the training phase, further outputs the class. The class outputted is the mode of classes if the problem under consideration belongs to classification while in case of regression outputs the mean prediction of individual trees. Being based on bagging ensemble learning, it deploys multiple uncorrelated decision trees on various subsets of a given dataset. Each of these decision trees outputs a certain prediction. Based on the average of predictions, final output is predicted. The proposed system is a regression-based task, where we have to predict output labels that should be continuous numeric values. Hence, the average of previously observed labels will give us a final prediction. The greater number of trees in the forest makes a robust forest that leads to accurate and stable prediction.

3.1. Algorithm for Random Forest

i. Select random samples from a given dataset and build multiple subsets.
ii. Build a decision tree associated for every subset.
iii. Every decision tree will output a prediction.
iv. Take the average of these predicted values.
v. The average value will be the final prediction.

Figure 1 depicts how random forest produces the predicted value of an unseen data point.

TREE 1   TREE 2   TREE n
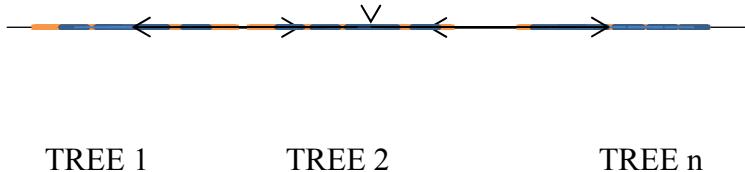
Figure 1. Test sample prediction by random forest

Being a decision tree-based ensemble machine learning algorithm, XGBoost aggregates output of several models and is one of the sturdiest techniques for fabricating predictive models. It uses a gradient boosting framework where errors are minimized by summing up weak learners applying gradient descent optimization algorithms. Boosting, being an ensemble technique, generates new models by correcting the errors made by preceding models. The process of sequentially adding models continues until no new model can be added further. Gradient Descent technique is used by the gradient boosting algorithm (GBM) to reduce the loss and to add a new model. Every decision tree takes a different subset of features; hence individual trees are different from each other. More upon, every successive tree takes the error of the preceding tree into account. In order to rectify the final yield of the model, value yielded from the new tree is added to the output of the existing sequence of trees. Process of XGboost is shown in Figure 2.
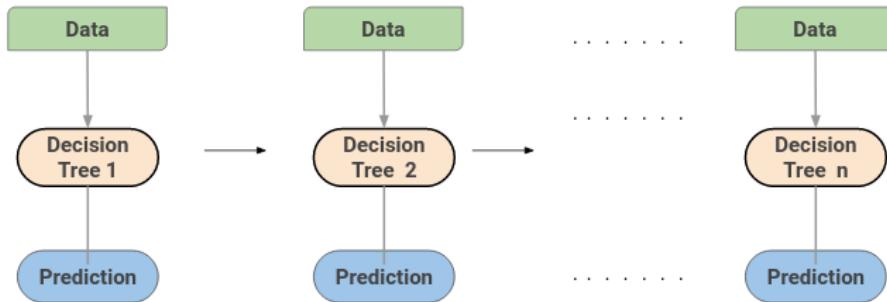
Figure 2. XGBoost prediction process

Extreme Boosting uses GBM at its core. XGBoost works on similar lines as in boosting weak learners using the gradient descent algorithm. XGBoost is used to address both classification and regression problems. XGBoost provides improvement over the base GBM framework by adding system optimization and enhancement in algorithms. The task of building a sequential tree is addressed using parallelized implementation that results in improved algorithmic performance. Using depth first approach for tree pruning, there is significant improvement in computational performance. It provides auto tree pruning so that decision trees do not grow after a certain limit. Cache awareness, out-of-core computing are offered in XGBoost algorithm which results in hardware optimization. It supports distributed computing for training very large models using a cluster of machines. The algorithmic enhancements such as Regularization, sparsity awareness, cross validation is supported by the algorithm. The algorithm also handles missing values on its own.

4. Implementation and Results Analysis

The system is implemented in python using anaconda as IDE. Sklearn tool is used. A python package sklearn. ensemble. Figure 3 Random Forest Regressor is used to implement the random forest algorithm. estimators i.e., the number of trees in the forest is set to 400, min_samples_split which is the minimum number of samples required to split an internal node is set to 5. The parameter min_samples_leaf that explores the minimum number of samples in newly created leaves is set to 5, max_features is the number of features to be considered while best split is set to auto. The parameter max_depth, the maximum depth of the tree is set to 15. Further RandomSearchCV is used to implement the fit and score method. neg_mean_squared_error is passed as a scoring method that computes mean squared error.

```python
from sklearn.ensemble import RandomForestRegressor
regressor=RandomForestRegressor()
from sklearn.model_selection import RandomizedSearchCV
rf = RandomForestRegressor()
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf}
rf_random = RandomizedSearchCV(estimator = rf,
                               param_distributions = random_grid,
                               scoring='neg_mean_squared_error',
                               n_iter = 10,
                               cv = 5,
                               verbose=2,
                               random_state=42,
                               n_jobs = 1)

rf_random.fit(X_train,y_train)
predictions=rf_random.predict(X_test)
```

Figure 3. Using RandomForestRegressor

The used car price prediction system is also implemented using a scalable, portable, accurate machine library namely XGBoost library. It is an open-source library that provides effective implementation of Gradient Boost algorithm. The parameters used are n_estimators which is the number of trees to be built with a value of 400, learning_rate which is step size shrinkage used to prevent overfitting is set at 0.05, subsample which is the percentage of samples used per tree is set as 0.7. The parameter max_depth regulates how deep each tree should grow during any boosting round is set as 20, min_child_weight is 6. The parameter objective is used to specify the loss function to be used.

```python
from xgboost import XGBRegressor
xgb = XGBRegressor()
xgb.fit(X_train, y_train)
from sklearn.model_selection import cross_val_score
score = cross_val_score(xgb, X, y, cv = 5)
params = {
    'n_estimators': n_estimators,
    'learning_rate': learning_rate,
    'max_depth': max_depth,
    'subsample': subsample,
    'min_child_weight': min_child_weight,
    'objective': objective
from sklearn.model_selection import RandomizedSearchCV
search=RandomizedSearchCV(xgb,params,scoring='neg_mean_squared_error',
                          cv=5, n_iter=10, random_state=43, n_jobs=-1,
                          verbose=2)
search.fit(X,y)
```

Figure 4. Using XGBRegressor

For this system a user interface is created using flask web framework. The interface will appear like the below figure 5. It is hosted on AWS for accessing remotely and it can be accessed through http://34.237.253.179:5001/

Figure 5. Web User Interface for prediction

5. Results

Metrics considered to test the strength of the algorithm are Mean Absolute Error, Mean Squared Error, Root Mean Squared Error. These three metrics are used to evaluate both the regression algorithms. The three metrics have lower values for Xgboost, hence it has higher accuracy over the random forest algorithm. Table 1 shows the result analysis of both the algorithm.

Table 1. Result Analysis

Algorithm	MAE	MSE	RMSE
Random Forest	1.13	11.89	3.44
XGBoost	**0.17**	**0.28**	**0.53**

Graphical representation of the result analysis is given in Figure 4. Lowest values are considered as the best result**.**
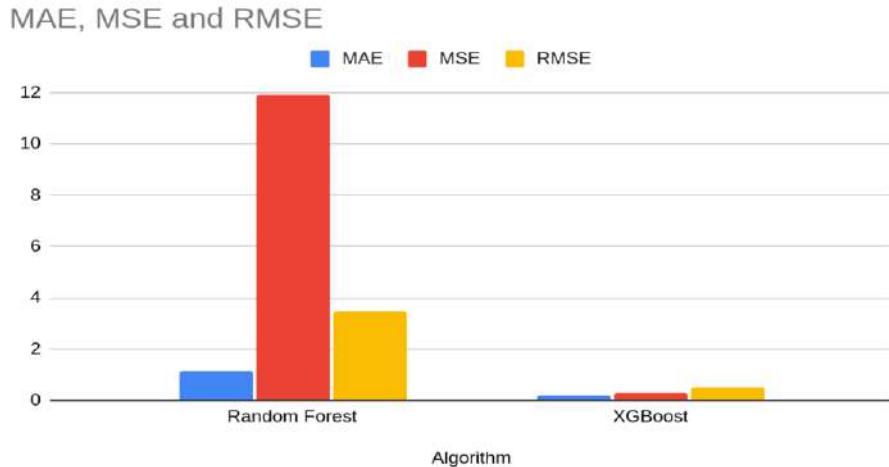
Figure 6. Graphical Representation of Result Analysis

6. Conclusion

The proposed system works well to predict a fair price for the pre-owned cars. The system skilfully projects prices for used cars in Mumbai region. The user of the system either the seller or the buyer, will get the honest price for the used car. Two popular ensemble machine learning algorithms namely Random Forest and XGBoost are deployed in order to implement a regression system for predicting used car prices. Both the techniques are comparable and offer high accuracy. Random Forest prevents overfitting by making use of more trees. With an ability to handle missing values, prevent overfitting, XGBoost is a widely used algorithm. As displayed by the results, XGBoost outperforms the Random Forest Algorithm. XGBoost is fast to execute and gives appreciable accuracy. The system proposed here is implemented for the Mumbai region only. However, it can be extended to other regions too, if the data available is in the suitable format.

References

[1]   A report by Motor Intelligence on India Used Car Market - Growth, Trends, COVID-19 Impact, and Forecasts (2021 - 2026). https://www.mordorintelligence.com/industry-reports/india-used-car-market

[2]   Sameerchand Pudaruth, Predicting the Price of Used Cars using Machine Learning Techniques, International Journal of Information & Computation Technology,ISSN 0974-2239 Volume 4, Number 7 (2014), pp. 753-764

[3]   Enis Gegic et al., Car Price Prediction using Machine Learning Techniques, TEM Journal. Volume 8, Issue 1, Pages 113-118, ISSN 2217-8309, DOI: 10.18421/TEM81-16, February 2019.

[4]   Jian-Da Wu,Chuang-Chin Hsu,Hui-Chu Chen, An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference, Expert Systems with Applications 36 (2009) 7809–7817, 2008 Elsevier Ltd.

[5]   Richardson M., 2009. Determinants of Used Car Resale Value, Thesis(BSc).,the Colorado College.

[6]   Ning Sun, Hongxi Bai, Yuxia Geng, Huizhu Shi, Price Evaluation Model in Second-hand Car System based on BP Neural Network Theory, IEEE SNPD 2017, June 26-28, 2017.

[7] Zhenchang Xia, Shan Xue, Libing Wu, Jiaxin Sun, Yanjiao Chen & Rui Zhang, ForeXGBoost: passenger car sales prediction based on XGBoost, Distributed and Parallel Databases , volume 38, pages713–738(2020)SpringerLink,25 May 2020.

Prediction of Fake Tweets Using Machine Learning Algorithms

M. Sreedevi [a, 1], G. Vijay Kumar [a], K. Kiran Kumar [a], B. Aruna [a], Arvind Yadav [a]

[a] *Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur-522502, India*

Abstract. Social networking sites will attract millions of users around the globe. Internet media is becoming popular for news consumption because of its ease, simple access and fast spreading of data takes to consume news from social media. Fake news on social media is making an appearance that is attracting a huge attention. This kind of situation could bring a great conflict in real time. The false news impacts extremely negative on society, particularly in social, commercial, political world, also on individuals. Hence detection of fake news on social media became one of the emerging research topic and technically challenging task due to availability of tools on social media. In this paper various machine learning techniques are used to predict fake news on twitter data. The results shown by using these techniques are more accurate with better performance.

Keyword. Social media, networking, Fake tweets, Data mining, Prediction

1. Introduction

In recent years, the popularity of social media has increased significantly because many people give priority to read the news through social media [1]. There are millions of news content published on the internet than traditional news paper, so social media became the main platform in terms of spreading news. However, Twitter App [2] is not only based on individual opinions, and also for an official information of a particular occasion could be delivered, such as organization could use Twitter-API like an unique and personal [3,4] user. There are variety of news can spread over social network such as satire, Hoax and fabrications. Hoax is one type of false news can spread using various sources like stories, people will easily trust this type of news. Satire is another type of false news where true is presented as a joke. So there is a chance to consider the real news as silly joke. Fabrications are international lies, these are depends on click bytes [5].

Twitter can likewise might be utilized in a different term, and it is worth might be boundless later. Knight Foundation began an examination to explore how counterfeit news spread on the stage previously, during and after the 2016 U.S. official political decision [1]. The assessment broke down more than a huge number of tweets from many Twitter accounts that connected to in excess of 600 phony and intrigue news destinations.

[1] M. Sreedevi, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation

Email: 170030759@kluniversity.in.

It recognized bunches of Twitter accounts that connected back to these destinations more than once, regularly in manners that appeared to be composed or even mechanized. Fake news is duplicate or a deceptive data introduced as news. It regularly has a point of harming the standing of an individual. First, fake news [6] is 8 intentionally written to mislead consumers, it does not make satisfactory to spot fake news from news itself.

You can report straightforwardly from an individual Tweet, List, or profile for specific infringement, including spam, oppressive or hurtful substance, wrong advertisements, self- mischief, and pantomime. For data about announcing different sorts of infringement but even to report the spam issues or fake tweets [7] we need to identify the fake tweets so to predict such fake information we are about to use 5 algorithms [6] and it would be helpful to report all kinds of harmful rumors.

2.   Literature Survey

In [8] Author suggested design of four modules are Feature Clustering using Modified K- means, user interface, log pre-processing, Naïve Bayes Classification testing and training using k-nearest neighbors' algorithm for best Précised classification of viewpoint. The design could find unrelated data and more accuracy using K means Modified with Naïve Bayes algorithm. Present research is searching on Sentimental analysis [9] regarding complete decision of highly organized web sources such as communication groups like gatherings, feedbacks are in digital form. Many ways that got improved for main problems of opinion mining's and got solved. We received a picture of which are involved in making a software system for sentimental analysis regarding the analysis. In the calculation investigates the connection got and returns a string message that incorporates the Decision, the User score, and the Tweet Score. The twitter connect is incorporated of twitter address, client screen name and status id (tweet id). Twitter has an easy-to-understand way of taken this tweet connect. Then, calculation fills in as it follows: It gets the client subtleties from our information base, If they don't have the client in the twitter stockpiling, it gets its qualities from twitter and introduce its score to nothing, It gets the tweet's subtleties from Twitter, Creates a tweet score code dependent on the got information from various API's, Creates a client score code as a number juggling mean of its earlier score and genuine tweet score.

In [10] This paper presents a down to earth examination to pick a high perform characterization technique and the unfeasible purposes behind the high performed arrangement. They introduced their framework which work to recognize counterfeit clients and phony news in the Twitter interpersonal organization by acquiring validity score. The calculation at that point utilizes a Named Entity Recognition segment, what parts the content into its after parts: it draws out the marks, the subjects, the social labels, the general tweet assessment, and the hash-label opinion. Framework engineering For the NER parsing, they utilized a public API (Application Programming Interface) called Open-Calais, while for the opinion calculation, we utilized Sentiment140. In [11] polarity of text data identification in NLP became complex due to collection of feedback about products from various resources. To avoid this complexity they performed sentiment analysis by analyzing the behavior and attitude of people by using genetic algorithm using CNN for opinion classification. They studied

existing sentiment analysis methods on twitter data and they provide comparisons theoretically of state-of-art approaches. They also analyzed document level sentiment analysis on twitter data using various methods. In [12] they performed classification approach using TSVM/SVM and C-SVM data to detect spam tweets or duplicate tweets on twitter data. Text performance on integrated cluster classification performed better results than SVM classification. Also, regular and popular pattern mining [13-15] are very much useful in prediction of fake tweets.

3.   METHODOLOGY

To predict the given news is real or fake news, we used classification algorithms like Logistic regression, Naive bayes, Stochastic Gradient Descent (SGD), Support Vector Machine (SVM) and Random Forest. After evaluating the performance of these algorithms, we conclude which among these algorithms is the best.

Data Preprocessing: Data preprocessing is one of a data mining procedure which converts the raw data into understandable, useful, and efficient formats. True information is inadequate, conflicting and it contains many errors which affects the results. So, all these issues can be resolved by performing data preprocessing.

Data Cleaning: Data cleaning is the way toward eliminating off base, copy, or inadequate information inside a dataset. When we combine multiple information sources there are numerous chances for data to be duplicated.

Data Integration: Data Integration is information pre-processing procedure that joins data from multiple sources into a reasonable data and provides unified view of the data.

Data Transformation: Data transformation is a data preprocessing technique that converts source framework into the necessary configuration of a destination system. It is the process of changing the format, structure or values of data.

Data Reduction: Data reduction is a data preprocessing strategy that decreases the measure of limit required to store the data. It can decrease the costs and increase storage capacity.
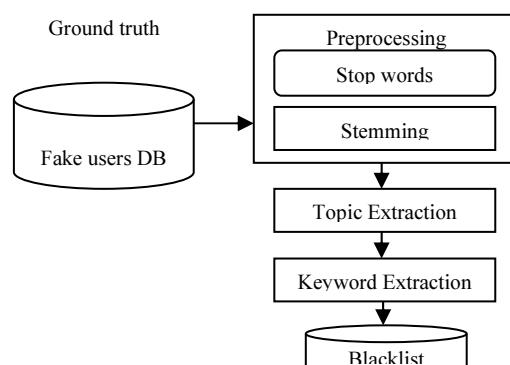
Figure 1. Block Diagram of Preprocessing

Logistic Regression: It helps to find the probability of success and failure of an event. It is a statistical model that in its basic form it models a binary dependent variable using logistic function. It is performed at the point when the needy variable is twofold. It is named as calculated relapse since it is very similar to direct relapse. The term

'strategic' is taken from the capacity logit which is utilized in the technique for grouping. In regression analysis, logistic regression is making an estimation of the parameters of a logistic model.
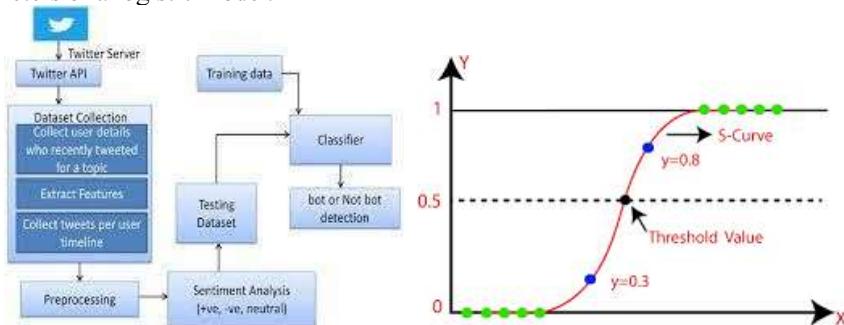
Figure 2. Process of Training data in Sentimental Analysis

Naive Bayes: Naive Bayes is used to predict the probability of different classes on various attributes. This can be utilized in text classification and it is used for problems having multiple classes. It is not sensitive to irrelevant features. It takes care of continuous and discrete information and it is mostly adaptable with number of highlights and data points. It makes real time predictions fast.

Stochastic Gradient Descent (SGD): The term 'stochastic' means a process that is linked with random probability. Stochastic gradient descent would randomly pick samples and uses the samples to calculate the derivatives. A couple of tests are randomly chosen instead of whole dataset for each iteration. It is especially useful when there are redundancies in the data.

Support Vector Machine (SVM): It is a supervised learning technique that can be utilized for relapse and order issues. It represents the datasets as points. The main objective of svm is to construct a hyperplane that divides the datasets into distinct categories. The hyperplane is a choice limit that helps in grouping information focuses, and the hyperplane ought to be at the limit edge from particular classes. This algorithm helps in removing the function that is too closely fit to a limited set of data points of the samples and provides better accuracy.
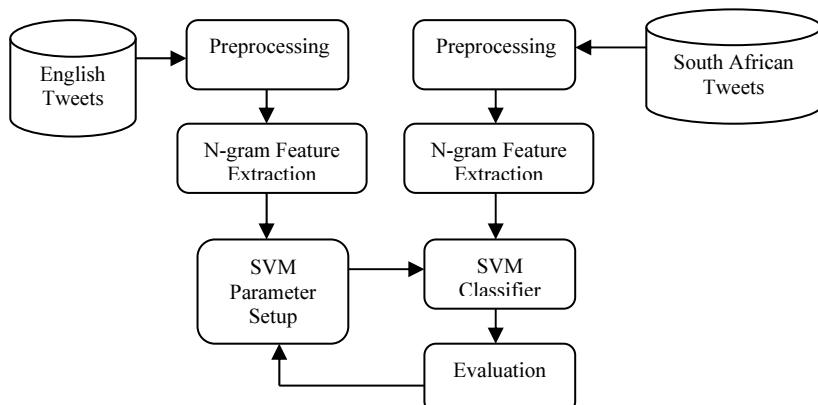
Figure 3. Comparison of various tweets in preprocessing

Random Forest: This is an outfit learning strategy for grouping and relapse that utilizes

different models of a few choice trees to get better forecast execution. It constructs various choice trees and join them to get exact and stable expectations. It runs productively on huge data sets. It can deal with many information factors without taking out factor and gauge factors which are significant in the arrangement.

4.   Results

Sample datasets are collected from Kaggle and we used python as working environment. After preprocessing the information is divided into train and test datasets. By using five different machine learning algorithms the following accuracy values are found to identify the news as fake or not is given in table 1. The least accuracy is given by Naïve Bayes and highest accuracy is given by SVM about 99.48% and SGD is about 99.28%. The values which got for prediction of fake news for 5 data mining algorithms are represented as line graph in figure 5.

Table 1. Algorithm Results

S.No.	Model	Accuracy
1	Logistic Regression	98.0%
2	Naïve Bayes	94.0%
3	Stochastic Gradient Descent	99.28%
4	Support Vector Machine	99.48%
5	Random Forest	98.0%

```
data = ['Methods': ['NAIVE_BAYES','RANDOM_FOREST',"LOG_R','SGD','SVM'],
        'Accuracy_Rate': [94.0,98.0,98.0,99.28,99.48]
        }
df = pd.DataFrame(data,columns=['Methods','Accuracy_Rate'])
df.plot(x ='Methods', y='Accuracy_Rate', kind = 'bar')
plt.show()
```
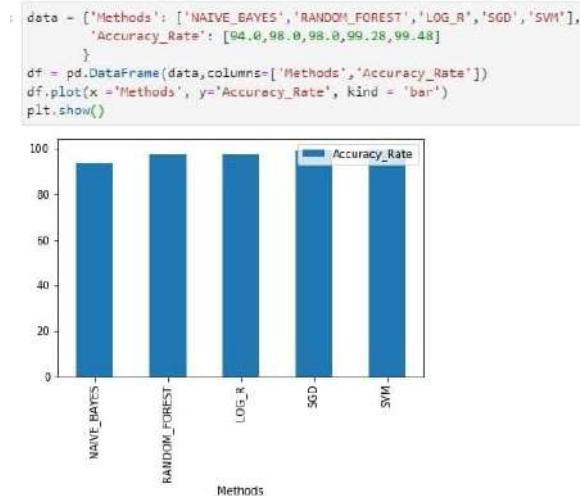

Figure 4. Accuracy Rate results on various algorithms

5.   Conclusion and Future Work

In our project, we worked upon models to determine the fake news among the datasets using data mining techniques, and this paper had about five different classification models which are compared by the accuracies. SVM got the highest accuracy and it is about 99.48%. In this paper we consider only twitter data for prediction of fake tweets. Our future scope is to predict fake

news against political affairs, economic affairs and Health affairs on facebook and other social sites data.

```
[20]: data = {'Methods': ['NAIVE_BAYES','RANDOM_FOREST','LOG_R','SGD','SVM'],
              'Accuracy_Rate': [94.0,98.0,98.0,99.28,99.48]
             }
```

```
[21]: df = pd.DataFrame(data,columns=['Methods','Accuracy_Rate'])
      df.plot(x ='Methods', y='Accuracy_Rate', kind = 'line')
      plt.show()
```
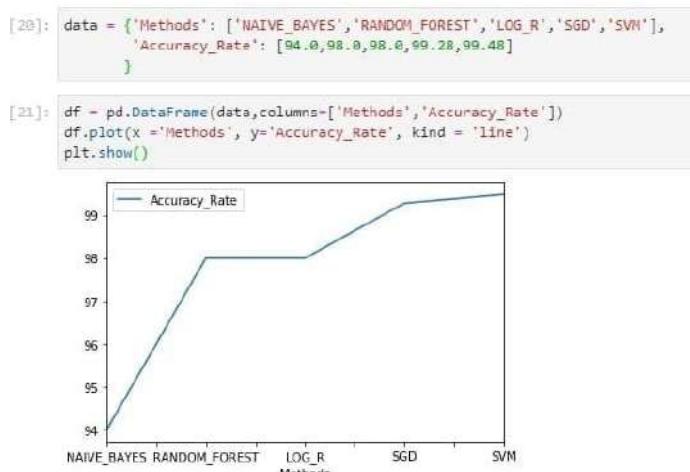

Figure 5. Comparison between accuracy rates

References

[1] Kelly Stahl, "Fake news detection in social media " B.S. Candidate, Department of Mathematics and Department of Computer Sciences, California State University Stanislaus, 1 University Circle, Turlock, CA (Confidentiality Agreement) 95382 Received 20 April, 2018; accepted 15 May 2018.

[2] Monika Verma, Sanjeev Sofat ,"Techniques to Detect Spammers in Twitter- A Survey ", Int. J. Comput. Appl 85(10):27-32, January 2014.

[3] Costel-Sergiu Atodiresei*, Alexandru Tănăselea, Adrian Iftene , "Identifying FakeNews and Fake Users on Twitter " Available online 28 August 2018. Published by: ElsevierLtd.

[4] Potharaju, S. P., & Sreedevi, M. (2017). A Novel Clustering Based Candidate Feature Selection Framework Using Correlation Coefficient for Improving Classification Performance. Journal of Engineering Science & Technology Review, 10(6).

[5] Pallavi B. Petkar, S. S. Sonawane, "Fake News Detection: A Survey of Techniques", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-9, July 2020.

[6] Diana Elizabeth Roy, Prannay S Reddy, "A Study on Fakenews Detection Using Naïve Bayes, SVM (Support Vector Machine), Neural Networks and LSTM (Long Short-Term Memory).", Journal of Advanced Research in Control & Dynamical Systems, 2019.

[7] Shlok Gilda, "Evaluating M L Algorithms for Fake News Detection" Student Conference on Research and Development (SCOReD), December 2017, pp.110-115.

[8] Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.

[9] Abdullah Alsaeedi , Mohammadd Zubair Khan, "A study on sentiment Analysis Techniques of Twitter Data" Published in IJACSA, Volume 10 Issue 2, 2019.

[10] SreeDevi, M Rama Bai, M Chandrashekar Reddy."Newspaper Article Classification using Machine Learning Techniques"ijitee.March 2020

[11] Sravya, G., Sreedevi, M. "Genetic optimization in hybrid level sentiment analysis for opinion classification", International Journal of Advanced Trends in Computer Science and Engineering, 2020, 9(2), pp. 1440–1445.

[12] Laxmi Narasamma, V., Sreedevi, M. "A Comparative Approach for Classification and Combined Cluster based Classification Method for Tweets Data Analysis", Smart Innovation, Systems and Techologies 2020, pp. 329-339.

[13] Vijay Kumar G, Sreedevi M., et al., "Incremental mining of popular patterns from   transactional databases", IJET, vol 7 issue 2.7, pp. 636-641, 2018

[14] Vijay Kumar G, Sreedevi M et al., "Regular frequent crime pattern mining on crime datasets" IJET, vol 7 issue 2.7, pp. 972-975, 2018.

[15] Vijay Kumar G, Krishna Chaitanya T. "Mining Popular Patterns from Multidimensional Database", Indian Journal of Science and Technology, Vol 9 issue 17, May 2016.

SU-CCE: A Novel Feature Selection Approach for Reducing High Dimensionality

A B Pawar [a, 1], M A Jawale [a], Ravi Kumar Tirandasu [b] and Saiprasad Potharaju [a]

[a]*Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra, India*
[b]*Department of Computer Science & Engineering, Koneru Laksmaiah Education Foundation, Vaddeswaram, AP, India*

Abstract. High dimensionality is the serious issue in the preprocessing of data mining. Having large number of features in the dataset leads to several complications for classifying an unknown instance. In a initial dataspace there may be redundant and irrelevant features present, which leads to high memory consumption, and confuse the learning model created with those properties of features. Always it is advisable to select the best features and generate the classification model for better accuracy. In this research, we proposed a novel feature selection approach and Symmetrical uncertainty and Correlation Coefficient (SU- CCE) for reducing the high dimensional feature space and increasing the classification accuracy. The experiment is performed on colon cancer microarray dataset which has 2000 features. The proposed method derived 38 best features from it. To measure the strength of proposed method, top 38 features extracted by 4 traditional filter-based methods are compared with various classifiers. After careful investigation of result, the proposed approach is competing with most of the traditional methods.

Keywords. Data mining, classifier, feature selection, preprocessing.

1. Introduction

Data mining is the adulatory area of research since a decade in many fields including medical investigations, stock market analysis, business, education, transportation, etc. With the data mining approaches the more insights of the data can be analyzed for future prediction in order to get positive results. For any field of study data is crucial part. Data may be in different forms (text, audio, video, image etc.) from field to field. Before applying data mining techniques, the type and format of data need to be preprocessed [1]. The collected data may have missing values, missing classes, imbalanced and high dimensional in nature. Missing values can be addressed using by filling the values using various techniques like average values, binning. Missing classes can be filled by calculating the distance with already classes. Imbalanced issue can be addressed with applying the oversampling and under sampling techniques [2].

---

[1]A B Pawar, Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon, India
Email: anil.pawar1983@gmail.com.

High dimensionality can be addressed using feature reduction and feature selection algorithms.In this research article, high dimensionality issue of preprocessing is addressed by proposing novel feature selection approach using correlation (Cor) coefficient and Symmetrical uncertainty.

If a greater number of independent features (may be hundreds to thousand) existed in a dataspace is called high. dimensional data [3]. Search space for the given problem will be increases as number of features are more in the dataspace. This will create a problem to face curse of dimensionality. This type of datasets consumes more amount of memory during the classification model generation. In case of lazy learners, it is very difficult for generating classification model because model is not created in advance. The classification model generated over high dimensional data may perform adversely and classifier may get confuse in classifying an unknown instance.

This high dimensionality issue is very common thing which will appear in machine learning datasets. There are some techniques existed for minimizing the number of features needed for classification as high dimensional data consists of irrelevant and redundant features. Principal component (PCA) analysis is widely used feature reduction process to identify the principal components in a classification which identifies attributes that are orthogonal to others. Apart from feature reduction, feature selection is another possible solution. Filter, wrapper, and embedded are three types of feature selection modes. Filter mode gives the rank to each feature in the dataset. Depending on the threshold (number of features to be considered to generate the model (n)), top 'n' features will be selected as per the rank given. ReliefF, Gain ratio, Chi-Squared,Information Gain, and Symmetrical Uncertainty are some of the existed methods based on filter approach. Wrapper mode gives the subset of features directly. It uses the searching algorithm (BFS, DFS. Genetic, etc) and learning algorithm for generating subset of feature. Embedded combines both these approaches [4]. This paper is based on filter methods only.

Microarray datasets has huge number of features ranging from few hundreds to few thousand, generating a classification model on such huge dataset is challenging job. Such type of dataset analysis required a proper feature selection approach in order to reduce the memory consumption problem and increasing the classification accuracy [5]. In this research colon microarray dataset, which has 2000 features is examined with the some of the existing methods and proposed method. The same method also tested on other datasets also for generalizing it.

In this research, one of the filter-based method Symmetrical uncertainty (SU) is used along with correlation coefficient (CCE) for proposing a new method. The relation between SU and Cor are given in next section. The existed literature related to the feature selection is articulated in second part. The proposed approach is described in methodology section. The experimental result analysis output and dataset description is explained in fourth section. The article is concluded with future recommendations.

2. Related Work

The proposed method is based on Symmetrical uncertainty (SU) which is filter-based approach and correlation coefficient (CCE). In the existed literature, SU and Cor is applied by many researchers on various datasets. Some of the researchers considered SU and Cor separately, few are considered combined. SU can be defined as

$$SU=2*IG/(H(F1) +H(F2)) \qquad . (1)$$

IG is Information Gain; *H(F1)* is Entropy of F1; *H(F2)* is Entropy of F2. The value of SU ranges from 0 to 1. The featurewith high SU score is considered as strong feature.

A distributed feature selection approach based on SU is proposed by the authors to reduce the features of 6 microarray datasets [6]. A quarter feature selection approach based on SU is proposed by the researcher and examined on various datasets [7]. In their approach, researchers divided the dataset in 4 clusters (25% of features). Top 25% of features derived by some of the traditional methods are compared with their approach by applying various classifiers. One of the cluster of features derived by the researchers performing better than the existed methods. Assembling approach based on SU is proposed by the authors and examined on the sonar target classification [8]. Authors divided the sonar dataset in various clusters and applied bagging and boosting to measure the classification accuracy.

Another statistical measure used in this research is Correlation Coefficient (CCE), which is used to measure the strongness /dependency between two features. The CCE between two features can be [-1, +1]. +1 indicates, two variables are strongly/positively correlated, so one attribute can be considered and other can be ignored. CCE of two random variables X and Y can be derived as below equation

$$CCE = n (\sum xy) -(\sum x) (\sum y) / \sqrt{[n\sum x2-\sum x2)]} [n\sum y2-\sum y2)] \quad . (2)$$

A FAST feature selection-based SU and CCE is proposed and tested over several datasets [9]. Authors achieved 85% success rate over traditional methods using FAST. Authors constructed the graph based on the values of SU and CCE, later applied the prims algorithm to find minimum spanning tree. The result of minimum spanning tree is considered as reduced subset of features. Our proposed system is inspired from FAST subset feature selection with different approach.

There some literature related to the colon cancer is available. Authors proposed feature selection approach based on clustering concept over the colon cancer to classify the gene expression and they achieved the best accuracy than some of the existing methods [10]. Authors proposed a method based on Particle Swarm Optimization (PSO) feature selection algorithm along with the Support Vector Machine classifier algorithm to get the best features from colon cancer microarray data, their method competing with other existing algorithms in accuracy [11]. Feature selection practices like genetic algorithm and mutual information are applied by the researchers for testing the cancer microarray data [12]. By using their approach most expected cancer connected genes are determined from large microarray dataset. A Multi-Objective Binary PSO (MOBPSO)algorithm is suggested for inspecting the cancer gene expression data [13].

After applying feature selection process in the pre- processing, classification algorithms are applied using selected features. For this research also various tree, lazy, rule, and functional classifiers are applied to test the accuracy of selected features. The selected features are compared with the features derived by some of the existing methods such as ReliefF, Gain ratio, Chi-Squared, Information Gain. These are based on the concept of information theory [14].

3. Proposed Methodology

In this research, a novel feature selection technique is proposed based on two statistical components called correlation coefficient and symmetrical uncertainty. The proposed technique is based on the steps given in algorithm.

Algorithm steps:

1. Derive the SU score of every feature and ignore the feature whose score is zero, then place remaining features it in it's descending order of SU score. Elect the middle feature's SU score as Threshold (T).

2. Create the CCE Symmetrical matrix $(CCE(X_i, Y_i))$ of initial dataset .

3. Transfigure the $CCE(X_i, Y_i)$ matrix to weighted binary matrix (WB) as per the steps given below.

$$for(i=1\ to\ n)\ for(j=1\ to\ n)$$
$$if(CCE(X_i, Y_i) > T)$$
$$WB(X_i, Y_i)=1$$
$$else\ WB(X_i, Y_i)=0$$
$$EndEnd$$

4. Calculate the total weight of each feature by summing up all 1's related to each feature.

5. Group the features which are having the same weight($W(F)$)

6. $Cluster_i=\{F_{i1}, F_{i2},...F_{ik}\}$
   /* i is the cluster id, increment i by 1 until all features are formed */

7. Select the strong feature from each cluster. (i.e a feature whose SU score is greater than all of its features are nominated as strong feature)

Example

Consider the features *f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12)*

in primary data set.

1. SU score of every feature is given in Table 1. (As per step 1)

Table 1. SU score of all features in primary data set.

SU	Rank	Fid
.19	1	f10
.19	2	f8
.19	3	f7
.18	4	f9
.15	5	f2
.09	6	f1
.07	7	f4
.06	8	f3
.06	9	f5
.02	10	f6
0	11	f11
0	12	f12

* Ignore *f11* and *f12,* as their SU score is zero

2. *Threshold (T) = .15,* as 'b' is the middle feature. (As perstep 2)

3. $CCE(X_i, Y_i)$ matrix of the primary data set is given in below Table 2. (As per setp3)

Table 2. $CCE(X_i, Y_i)$ Matrix

Feature Id	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10
f1	1	0.21	0.25	0.22	-0.23	0.01	-0.23	0.06	0.07	0.09
f2	0.21	1	-0.15	-0.9	0.03	0.04	-0.09	0.05	0.04	0.07
f3	0.25	-0.15	1	0.08	0.05	0.24	-0.09	0.26	0.01	0.13
f4	0.22	-0.9	0.08	1	0.27	0.06	0.2	-0.21	0.02	0.03
f5	-0.23	0.03	0.05	0.27	1	0.03	0.06	-0.1	0.06	0.17
f6	0.01	0.04	0.24	0.06	0.03	1	0.04	0.19	0.03	-0.08
f7	-0.23	-0.09	-0.09	0.2	0.06	0.04	1	0.03	0.19	0.02
f8	0.06	0.05	0.26	-0.21	-0.1	0.19	0.03	1	0.04	0.02
f9	0.07	0.04	0.01	0.02	0.06	0.03	0.19	0.04	1	-0.06
f10	0.09	0.07	0.13	0.03	0.17	-0.08	0.02	0.02	-0.06	1

4. Convert the $CCE(X_i, Y_i)$ matrix to weighted binary matrix(WB) and calculate the total weight of each feature.

Table 3. Weighted Binary Matrix

Feature Id	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	Sum of Weight
f1	1	1	1	1	0	0	0	0	0	0	4
f2	1	1	0	0	0	0	0	0	0	0	2
f3	1	0	1	0	0	1	0	1	0	0	4
f4	1	0	0	1	1	0	1	0	0	0	4
f5	0	0	0	1	1	0	0	0	0	1	3
f6	0	0	1	0	0	1	0	1	0	0	3
f7	0	0	0	1	0	0	1	0	1	0	3
f8	0	0	1	0	0	1	0	1	0	0	3
f9	0	0	0	0	0	0	1	0	1	0	2
f10	0	0	0	0	1	0	0	0	0	1	2

5. Group the features which are having the same weightCluster 1 with weight
4: *{f1, f3, f4}*

    Cluster 2 with weight 3: *{f5,f6,f7,f8}*
    Cluster 1 with weight 2: *{f2, f9, f10}*

6. Select the strong feature from each clusterStrong subset of
cluster is *{f1, f8, f10}*

4. Experimental Results and Discussion

The entire research experiment was carried out in the software laboratory of
computer engineering department terminal. It was configured with Linux
distribution and installed necessary packages to run proposed algorithmic steps via
terminal. The proposed algorithm is experimented on Colon cancer microarray dataset.
It has 2000 features and 2 classes. As per the steps proposed in the algorithm, firstly SU
is applied and determined the features whose score is greater than zero. We could get 138
features whose SU score is greater than zero. Then, CCE matrix of those 138 features is
constructed. Then, weight of each feature is calculated. After completion of the process,
we could get subset of 38 features. To know the strength of the proposed method, top
38 features derived by some of the existing filter-based algorithms are taken into
consideration. With these 38 features, classification models are created. For this, Jrip,
Ridor (rule based), Simple cart(sc), J48 (tree based), Naive Bayes and instant based

IBK are used. This experiment is done with popular machine learning tool WEKA with all its default settings. For generating Correlation coefficient matrix, the popular statistical program R is used. Below table 4 has classification result accuracy.

Table 4. Classification Result Analysis

	Jrip	Ridor	J48	SC	NB	IBK
IG	77.41	74.19	80.64	77.41	80.64	82.25
Chi	79.03	<u>70.96</u>	80.64	77.41	83.87	82.25
Gr	77.41	74.19	80.64	77.41	82.25	82.25
Rel	77.41	74.19	80.64	77.41	80.64	82.25
Proposed	**82.25**	<u>70.96</u>	**91.93**	**83.87**	**88.87**	79.03
ALL features (SU > 0)	75.80	64.51	82.25	75.80	53.22	77.41

The accuracy of the features derived by Information gain (IG), ReliefF (Rel), and Gain ration (Gr) is same because, top 38 features derived by those are same. The comparative analysis of proposed method and existing algorithms with various classifiers is given in Fig. 1.
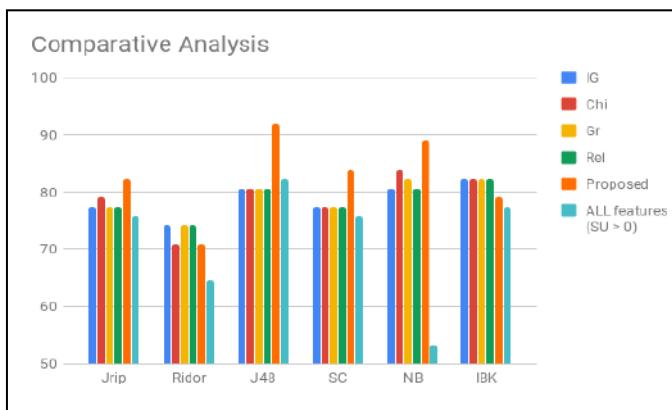
Figure 1. Comparative Analysis

Figure 1 show Comparative analysis reveals that, proposed method is performing better than all existing methods with Jrip, J48, and NB classifiers. It also competing with Chi and performing better if all the features are considered using Ridor. The proposed method is performing little lower than all existing, but recorded improved performance with IBK. The proposed method is tested with little variation on 10 real datasets and its performance is analyzed. After careful analysis majority of the cases the proposed method has displayed improved performance than existing methods.

5. Conclusion

In this research paper, a novel approach of subset selection of feature selection framework is demonstrated to reduce the dimensionality of a dataset to resolve the issues of selecting the appropriate and required strong features in high dimensionality dataset. In the experiment analysis, initially, Symmetrical Uncertainty and then and correlation coefficient are taken to select the useful, strong features. The developed method is compared with 04 existing filter-based methods as, Chi- Square, Grain Ratio, information gain, and ReliefF. For testing purpose, 06 classifiers Jrip, Ridor, J48, Simple cart, Naive Bayes, IBk are experimented on colon cancer high dimensional dataset, which has 2000 features and later demonstrated with 10 different real time data sets. After rigorous analysis, it is observed, proposed method gives promising better results over existing IG and GR on 8 data sets, also performed better on 8 data sets. Also, found more promising than ReliefF method.

References

[1] Ramírez-Gallego, S., Krawczyk, B., García, S., Woźniak, M., & Herrera, F. (2017). A survey on data preprocessing for data stream mining: Current status and future directions. *Neurocomputing*, *239*,39-57.

[2] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321-357.

[3] Rehman, M. H., Liew, C. S., Abbas, A., Jayaraman, P. P., Wah,T. Y., & Khan, S. U. (2016). Big data reduction methods: a survey.*Data Science and Engineering*, *1*(4), 265-284.

[4] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, *50*(6), 94.

[5] Al-Rajab, M., Lu, J., & Xu, Q. (2017). Examining applying high performance genetic data feature selection and classification algorithms for colon cancer diagnosis. *Computer methods and programs in biomedicine*, *146*, 11-24.

[6] Potharaju, S. P., & Sreedevi, M. (2018). Distributed feature selection (DFS) strategy for microarray gene expression data to improve the classification performance. *Clinical Epidemiology andGlobal Health*.

[7] Potharaju, S. P., & Sreedevi, M. (2018). A Novel Cluster of Quarter Feature Selection Based on Symmetrical Uncertainty. Gazi University Journal of Science, 31 (2), 456-470.

[8] Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In *Cognitive Informatics and Soft Computing* (pp. 247-256). Springer, Singapore.

[9] Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the 20th international conference on machine learning (ICML-03)* (pp.856-863).

[10] Nies, H. W., Daud, K. M., Remli, M. A., Mohamad, M. S., Deris, S., Omatu, S., ... & Sulong, G. (2017, June). Classification of Colorectal Cancer Using Clustering and Feature Selection Approaches. . In *International Conference on Practical Applications of Computational Biology & Bioinformatics* (pp. 58-65). Springer, Cham.

[11] Al-Rajab, M., Lu, J., & Xu, Q. (2017). Examining applying high performance genetic data feature selection and classification algorithms for colon cancer diagnosis. Computer methods and *programs in biomedicine*, *146*, 11-24.

[12] Pavithra, D., & Lakshmanan, B. (2017, June). Feature selection and classification in gene expression cancer data. In *Computational Intelligence in Data Science (ICCIDS), 2017 InternationalConference on* (pp. 1-6). IEEE.

[13] Chandra Sekhara Rao Annavarapu, S. D., & Banka, H. (2016). Cancer microarray data feature selection using multi-objective binary particle swarm optimization algorithm. *EXCLI journal*, *15*,460.

[14] Hall, M. A., & Smith, L. A. (1999). Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper. In *FLAIRS conference* (Vol. 1999, pp. 235-239).

A Review on Systematic Investigation of Leucocytes Identification and Classification Techniques for Microscopic Blood Smear

Mr. Pranav More [a,1], Prof. Dr. Rekha Sugandhi [b]

ªDept of CSE, MIT School of Engineering, MIT ADT University Pune, Maharashtra, India

ᵇDept of IT, MIT School of Engineering, MIT ADT University Pune, Maharashtra, India

Abstract. Healthcare services are an important part of human beings and healthcare services are changing with new and innovative technologies. In recent day's healthcare sector performing very crucial role in metamorphose of traditional health services to e-health technologies. This proposal provides an error-free and improved technology-based blood analysis service for the identification of leucocytes in blood samples of humans. Leucocytes play a vital and important character in human immune systems. This system helps to protect the body from suffering from leukemia. Leukemia, a blood cancer, nowadays is commonly found in all age persons. Leukemia is a type of disease and image processing techniques and algorithms can play a crucial role in disease diagnostic methodology. Identification of leukocytes in blood smear provides important information to pathologist as well as doctors to analyze and predicts different types of diseases, such as cancer. However, this analysis is critical and major complexities which results in errors and also takes a lot of time for analysis. Most of the time, the laboratory practitioners and doctors are interested only in leucocytes in blood smear. Medical image processing techniques strongly supports in their critical diagnosis and better results.

Keywords. leukemia, cytoplasm, leucocytes, hematic disease, nucleus

1. Introduction

The image processing and analysis is an important and emerging research area which has attracted a significant amount of interest from academic to industrial communities since last few years. Medical image recognition and processing applications widely used in the health domain for better results. The digitized and computerized system plays a vital role in various lines of research such as radioactive pharmaceuticals, x-rays (CT scans), magnetism (MRI), sound (ultrasound) and light (endoscopy, OCT), etc. The analysis of Leukocytes used in the clinical investigation of different diseases with

---

[1]   Mr. Pranav More, Dept of CSE, MIT School of Engineering, MIT ADT University Pune
Email : pranavmore2530@gmail.com

the help of digitized medical images and extracting important and accurate information. This extracted meaningful information is really valuable and useful for lab practitioners. In recent years, medical image analysis has helped to hematologists to analyze the different components of blood such as cells.

The analysis of different blood components leads to good quality, accurate, health standard, and remote disease diagnosis systems. Clinical analysis of blood samples to perform CBC (Complete Blood Count) is mandatory these days to go for any conclusion regarding the health diseases. In most of the cases, lab practitioners and doctors are responsive in white blood cells. Medical image processing techniques can promote the systematic use of health parameters in IT. This also tells about the secure sharing of human health information among health care professionals and practitioners.

Health professionals can check disease like leukemia and performs the analysis on detected components and condition of the cells like WBC. This digitization mainly segments the detected cells into two important elements such as nucleus and cytoplasm. This cell segmentation is performed using a proposed cell segmentation framework. The identification and classification of blood cancer leukemia majorly consist of several digital images processing algorithms like medical image acquisition, medical image segmentation, and feature extraction of the blood sample, feature selection, and classification of cells.

In this scenario, there are some limitations in data extraction from blood cell components due to large variations in cell size, shape of a cell, connected edges of cells, and cell position. If illumination is not balanced between the cell properties, the image color component contrast between blood cell edges and the cell background get changed based on the situation at the time of capturing process. The major important goal of this is to segment the cytoplasm and cell nucleuses using image digitization which we implemented and tested using medical image processing. In healthcare industries, medical image processing techniques use increased very fast in the recent few years. Figure 1 in this analysis, hematologists can conduct analysis on blood images and automatically process blood samples for the initial assumption of detecting the various diseases.
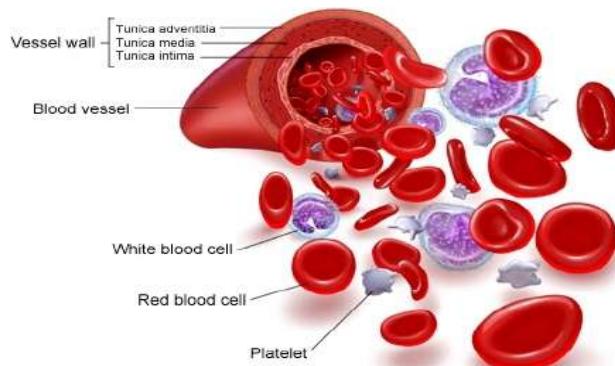
Figure 1. Types of blood cells in the human body

This methodology is used to identify the blood cells and counts the total numbers of white blood cells in the human body with higher accuracy and automated approach. This also provides complete and clear information on the nucleus versus cytoplasm ratio. This helps in the identification and classification of different kinds of white blood cells by using the internal component such as monocyte, eosinophil, basophil, neutrophil, and lymphocyte. Figure 2 in this research, a proposed methodology consists of different methods for segmentation and extraction which combines together for cell kernel and extracted cytoplasm.

neutrophil   eosinophil   basophil   monocyte   lymphocyte
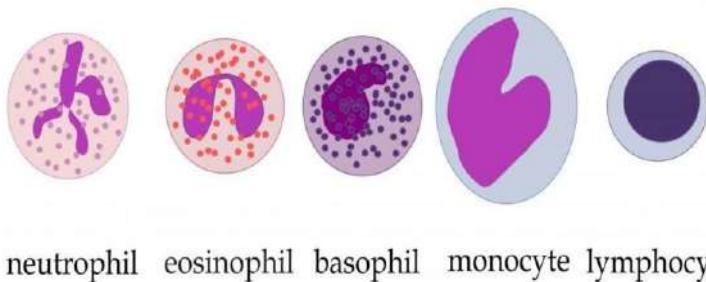
Figure 2. Types of White Blood Cells

2. Motivation

This research is very important to apply in medical field digital image processing for medical image examination and diagnosis. The major purpose of this research is to develop a framework to early diagnosing and classification of acute leukemia. Generally this can be done with the help of cell morphology, and classifies into (AML) Acute Meyloid Leukemia or (AAL) Acute Lymphoblastic Leukemia. The peripheral blood examination has been done for this research for following reasons including:

A. The blood malignant growth smear location measure is inspected dependent on the magnifying instrument morphological investigation of fringe blood slides. Further research center tests will be done dependent on the result of the underlying analysis.

B. The PB is typically utilized for an intermittent treatment assessment, since it is a lot simpler, more conservative, and less excruciating to get blood from the vein as opposed to from the BM.

3. Literature Review

AAL or AML Leukemia will be treated as a blood cancer. This will treated as most dangerous disease which influences the blood cells including White Blood Cells (WBCs). This will lead to casualty among people in many of the countries [15]. Human blood is a combination of billions of different cells in a super liquid form. Blood contains mainly three broad types of blood cells namely Erythrocytes commonly named as Red Blood Cells, Leukocytes commonly named as White Blood Cells, and

some platelets. Erythrocytes are majorly responsible for oxygen transportation; White Blood Cells perform fighting with infections and platelets are important for blood clotting. As per the statistics by the ACS (American Cancer Society), blood cancer is treated as a common cancer type, especially in children [16]. In the United States on average 29,000 cases in adults and in children near about 2000 cases are diagnosed.

Report by Agency of International Research on Cancer:

In 2016, one important report was released by the IARC and states that the frequency of children cancer higher than previously. This report completely depends on more than 90 cancer registries in 65 countries during the period 2001-2010. This indicates that near about 250,000 cases of blood cancers were diagnosed in teens and children under the age of 19. As per the report given by IARC on Cancer, [14] the following statistics in Figure 3, mentioned that there were approximately 80,000 deaths in a year from childhood cancer in the world.
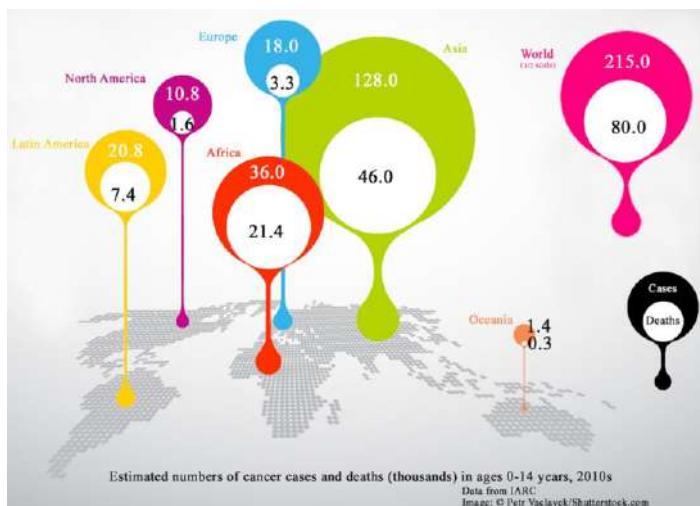
Figure 3. Statistics by International Agency for Research on Cancer-2016 [14]

Statistics by the American Cancer Society:

As per the analysis by the (ACS) Cancer Society America, leukemia is treated as common types of cancer, majorly found in children. Every year in United States near about 29,000 adults and near about 2000 new blood cancer cases are found & diagnosed. Generally, leukemia shows an impact on people in all age groups. In 85% of children, leukemia can be found in acute type. On the basis of analysis done by ACS, it has been proved that due to leukemia there is leading death ratio in children aged 1 to 14 years old [16]. According to the [17], the following graph in Figure 4, gives the total statistics of United States of cancer found in children from their birth to 15 years old. Figure 4 explains about the childhood cancer and affected areas of human body. Cancer affect the human body parts as per the as per the level and depth of a cancer.

Statistics of National Cancer Registry

The National Cancer Registry report distributed in 2008 [24], it has been demonstrated that malignancy in Malaysia is as per the human sex. In men, significantly discovered

the tumors are (incessant to least continuous): huge entrails, lung, nasopharyngeal malignancy, prostate organ, leukemia, lymphoma, stomach, liver, bladder, and other skin diseases. (Refer to Figure 5)
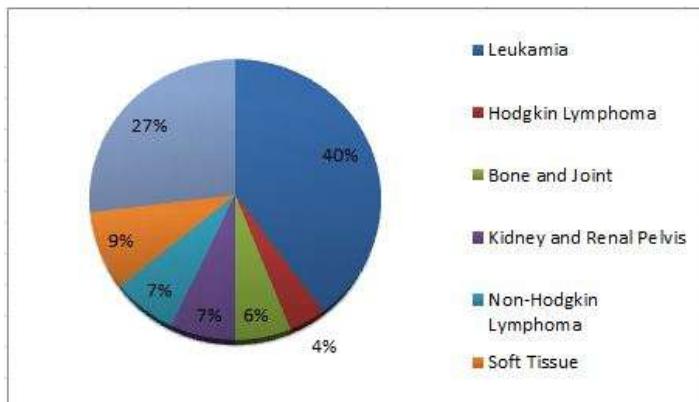
Figure 4. Analysis of US of Childhood Cancer

Statistics of National Cancer Registry

The National Cancer Registry report distributed in 2008 [24], it has been demonstrated that malignancy in Malaysia is as per the human sex. In men, significantly discovered the tumors are (incessant to least continuous): huge entrails, lung, nasopharyngeal malignancy, prostate organ, leukemia, lymphoma, stomach, liver, bladder, and other skin diseases. (Refer to Figure 5)
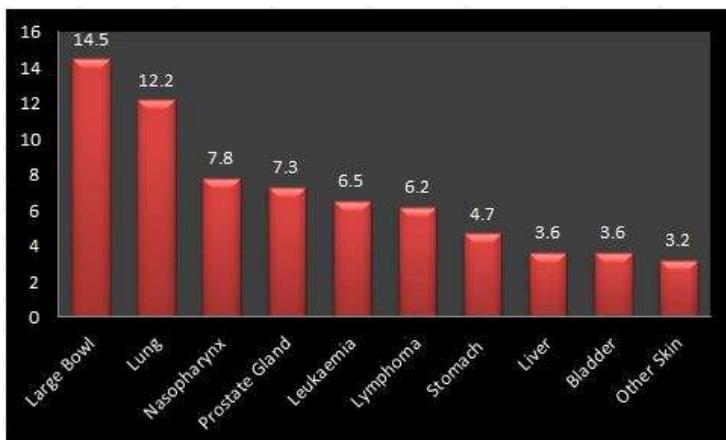
Figure 5. General category of Cancer in Malaysia in Male

The National Cancer Registry report distributed in 2008 [24], it has been demonstrated that in females, the commonest malignancies are (from generally successive to least incessant): bosom, cervix, enormous inside, ovary, leukemia, lung, lymphoma, corpus uteri, thyroid organ, and stomach. (Refer to Figure 6)
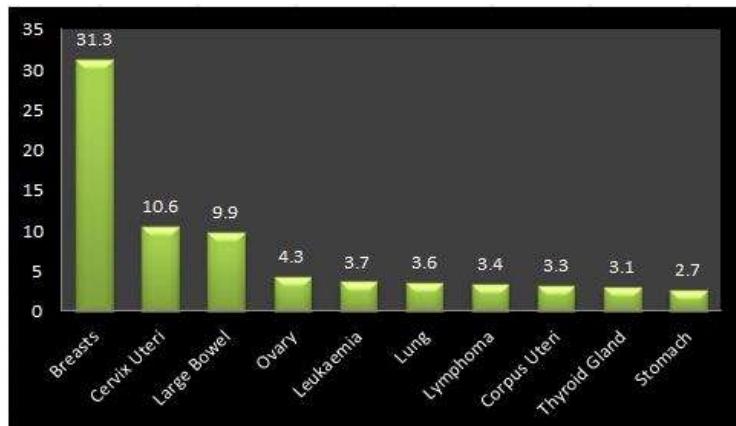
Figure 6. General category of Cancer in Malaysia in Female

4. Classification Methods

The literature review process started with studying the general challenges of Image processing such as Noise in segmented images, Cytoplasm shape, Background separation, Cell overlapping, Feature extraction, and Classification. Also how to maintain the sensitivity and specificity by using Multilevel Threshold-based image segmentation and how to increase the accuracy of classification.

Yampri proposed the use of green colour components for automatic thresholding between cytoplasm and nucleus for good contrast [08]. The available noise in the image is eliminated using the standard morphological operation process and the nucleus and cells are segmented using active contours. In some cases, most of the WBCs can have more number of nuclei in a cell, which results in decreasing of accuracy.

Some of the image processing operators analyse the boundary connection of the segmented nucleus using Watershed transformation. During this process, some series of morphological operations are performed structurally and cell size distribution data is used to distinguish the red blood cells and background of cytoplasm. A few restrictions are experienced when the states of the cytoplasm are not roundabout and the sizes change as indicated by the kinds of WBCs. [02]

In some proposed mechanism colour replacement used o find out neighbour contour cell using modified fuzzy c-mean clustering [09] technique. This approach can be cluster-based and the modification of data values is based on iteration. To test the images manual cropping of an image is required and an overall performance with other adaptive methods is evaluated.

Shift clustering [10] also used for leukocyte identification from blood images. This research used merging rules-based for mean-shift clustering and removing of boundary rules. Sholeh [11] proposed an algorithm for white blood cell image segmentation and enhancements of properties. Some operations related to morphology were used to remove unwanted noise and platelets from blood smear images. To maintain the structure of the identified cell, the opening of cell and cell closing operations were executed. Blue channel is used for the segmentation of WBC and finally, all five types

of WBCs were extracted, counted, and classified after analysis of cells. In the last step leukocytes counting is performed of segmented image and some features extraction properties such as minor axis length, major axis length, and average. Otsu segmentation an alternative methodology that can be applied directly for an image with the help of before segmentation of WBCs enhancement on blood samples [12]. Then, some of the morphological operations were used for eliminating undesired objects. Finally filling of holes can be performed.

Prinyakupt and Pluempitiwiriyawej [10] proposed the use of image thresholding and morphology for blood cell segmentation. Morphology is the process of complex mathematical computation which applies subtraction as well as addition on blood sample image. This will separate RBC, White blood cells and Platelets from the blood component. Segmentation was responsible to partition the image into two areas such as background and foreground by using an optimal threshold value.

Ravikumar [14] proposed a comprehensive investigation of two methodologies known as fast relevance vector machine (FRVM) and standard extreme learning machine (ELM) distribution technique used for segmentation. After comparing these two classification techniques identified that it performs well in local computation of maxima. In general, ELM is not satisfactory concerning performance than FRVM.

Table 1 describes the work done by various researchers in the field of medical image processing. In this table various technologies and methodologies are explained by the authors with respect to their accuracy and other remarks. Table describes the classification methods with respect to their works and methods. This table gives the idea about the work done in this field and results achieved by the researchers. Different classification methods are explained here for understanding and getting idea about the work done in this field.

Table 1. Study of Classification Techniques for Leukocyte Classification

Authors and Year	Work done	Technologies/ Algorithms used	Accuracy/ Remark
Nilanjan Ray, 2010 [25]	Formulation for Parametric Curve Fitting	Curve fixing, Curved programming, leukocyte discovery, oval fixing	Can be extended for feature correspondent.
Singh, Gautam, Bhadauria, Raman, 2016 [26]	Leukocytes Classification with the help of Naïve Bayes Classifier and Morphology Features	Otsu thresholding for segmentation, mathematical morphing, Naïve Bayes classification	80.88% accurate, Average processing 22s.
Doric, Irena, Novoselnik, Grbic, 2018 [27]	Convolutional Neural Network used for Identification and detection of cells	Segmentation and classification in 5 categories using (CNN) Convolutional Neural Network.	81.11% accurate.
Mlndhu, Dhiravida, Muthu, 2018 [28]	Convolutional Sparse Dictionary Learning used for Cell Discovery and Cell calculation	Holographic lenses free cell image in convolutional sparse dictionary learning	Checked zone has been recognized, distinguished
Harvindra Bhadauria, Anjali Gautam, 2014 [29]	Morphological Features are used for classification	Otsu's thresholding, morphing, classification of nucleus	Accurate classification of wbcs
Muhammad, Khan, Shoaib, Ali, Sajjad,2016 [30]	Classification & Segmentation using k- means algorithm, morphing, SVM for classification	Segmentation of nuclei, classification of wbcs	Accuracy higher than linear and naive Bayes classifiers.

Lalit Mohan Saini, Shubham Manik, Nikhil Vadera, 2016 [31]	Artificial Neural Network is used for counting of cells as well as classification.	Classification performed using ann. Finding of Intensity maxima for Nucleus Enhancement	Accuracy - 87.9%
Zahra Khandan Khadem Alreza, A Karimian, 2016 [32]	Classification of Leukemic Blood Image using machine vision system	Color space conversion model, watershed conversion	Accuracy 93%
Ricardo J Ferrari, Bruno C Gregorio da Silva, 2015 [33]	Intravital Video Microscopy for detection of cells based on the Eigenvalues analysis using Hessian Matrix	Detection of wbcs based on H matrix Eigen values	Improvement in IVM stabilization. Better than TM technique.

5. Result and Discussion

The proposed acute leukemia diagnostic methodology contains several phases with primary emphasis on detecting, segmenting, and then classifying acute leukemia blast cells. The PB images are first obtained as inputs to the diagnostic process and then analyzed through the proposed acute leukemia diagnostic phases. [34] The performance of the whole process is evaluated through various approaches and measures. The performance can be by using certain parameters to make certain decision of proposed framework. In the evaluation process, below some major parameters can be considered like accuracy, recall, precision, F- measure etc. The outcome of this evaluation process can be classified in (TP) True Positive, (FP) False Positive, (TN) True Negative and (FN) False Negative. In True Positive, affected leucocytes cells are identified accurately with good accuracy rate and in True Negative, non-affected leucocytes cells are identified. In False Positive, dangerous and non-affected cells are identified and in False Negative, non-dangerous but affected cells are identified.

6. Conclusion

In this paper, we discussed the background literature on the digital Image processing, including, color spaces, and image segmentation. An elaborative discussion regarding feature extraction was provided; this gives a clear idea about how work carried out here to accomplish the feature extraction methodology for the analysis of acute leukemia blast cells. This includes discussion of feature extraction and selection techniques including, Support Vector Machine. We additionally investigated existing arrangements of intense leukemia demonstrative frameworks which have been recently evolved to help and work with hematologists for the exact determination of the illness. This paper provided a report by the International Agency for Research on Cancer (IARC) and this shows that near about 250,000 cases of ALL and AML cancer are found and every year it diagnosed at the age of 19. Also provided Statistics by the American Cancer Society regarding the distribution of the more common childhood cancers for children from birth to 14 years old in the United States. The National Cancer Registry report published in 2008 gives categorized the most common types of cancer in Malaysia according to the gender i.e. male and female.

References

[1] Minetto, Dorini and Leite. White blood cell segmentation using morphological operators and scale-space analysis. SIBGRAPI 07: Proceedings of the XX Brazilian Symposium on Computer Graphics and Image Processing. 2007.

[2] Deng, Liao. An accurate segmentation method for white blood cell images. IEEE International Symposium on Biomedical Imaging. 2002; 245–248.

[3] Yang L, Meer P, Foran DJ. Unsupervised segmentation based on robust estimation and color active contour models. IEEE Trans.Inf. Technol.Biomed. 9. 2005; 475–486.

[4] BM Silva, JJPC Rodrigues, I de la Torre Díez, M López-Coronado, and K Saleem. Mobile-health: A review of current state in 2015. J. Biomed. 2015; 265-272.

[5] Yi F, Chongxun Z, Chen P, Li L, White blood cell image segmentation using on-line trained neural network. In: Proceedings of 27th International Conference on Engineering in Medicine and Biology Society. 2006; 6476–6479.

[6] Yampri P, Pintavirooj C, Daochai S, Teartulakarn S. White blood cell classification based on the combination of Eigen cell and parametric feature detection. In: IEEE Conference Industrial Electronics and Applications. 2006; 1–4.

[7] Chinwaraphat S, Sanpanich A, Pintavirooj C, Sangworasil M, Tosranon P. A modified fuzzy clustering for white blood cell segmentation. In: Proceedings of the Third International of Symposium on Biomedical Engineering. 2008; 356–359.

[8] FI Sholeh. White blood cell segmentation for fresh blood smear images. Proc. Int. Conf. Adv. Computer Sci. Inf. Syst. 2013; 425-429.

[9] S Nazlibilek, D Karacor, T Ercan, MH Sazli, O Kalender and Y Ege. Automatic segmentation, counting, size determination and classification of white blood cells. 2014; 58-65.

[10] J Prinyakupt and C Pluempitiwiriyawej. Segmentation of white blood cells and comparison of cell morphology by linear and naive Bayes classifiers. Biomed. Eng. On Line. 2015; 63.

[11] S Ravikumar. Image segmentation and classification of white blood cells with the extreme learning machine and the fast relevance vector machine. Nanomed, Biotechnol. 2016; 985-989.

[12] Kothari R, Cualing H and Balachander T. Neural Network Analysis of Flow Cytometry Immunophenotype Data. IEEE Transactions On Biomedical Engineering. 1996; 803-810.

[13] American Cancer Society. Childhood Leukemia. Retrieved on 23/10/2018 from http://www.cancer.org/cancer/leukemiainchildren/detailedguide/childhoodleukemia-diagnosis. 2013.

[14] American Childhood Cancer Organization. Childhood Cancer Statistics. Retrieved on 19/11/2018 from http://www.acco.org/Information/AboutChildhoodCancer. 2013.

[15] NIC Marzuki, NH Mahmood and MAA Razak. Segmentation of white blood cell nucleus using active contour. 2015; 115-118.

[16] BC Ko, JW Gim and JY Nam. Automatic white blood cell segmentation using stepwise merging rules and gradient vector flow snake. Micron. 2011; 695-705.

[17] F Zamani and R Safabakhsh. An unsupervised GVF snake approach for white blood cell segmentation based on nucleus. Proc. 8th Int. Conf. Signal Process. 2006; 1453-1456.

[18] Dohner H, Estey EH, Amadori S, Appelbaum FR, Buchner T, Burnett AK, Larson RA. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert. 2010; 453-474.

[19] Gokbuget N and Hoelzer D. Treatment of adult acute lymphoblastic leukemia. Paper presented at the Seminars in hematology. 2009; 64-75.

[20] Esposito F and Malerba D. Machine Learning in Computer Vision. Applied Artificial Intelligence. Published in An International Journal. 2001; 693-705.

[21] Lai DT, Pakkanen J, Begg R and Palaniswami M. Computational Intelligence and Sensor Networks for Biomedical Systems. Encyclopedia of healthcare information systems. 2008.

[22] LB Dorini, R Minetto and NJ Leite. Semiautomatic white blood cell segmentation based on multiscale analysis. IEEE J. Biomed. Health Information. 2013; 250-256.

[23] SF Bikhet, AM Darwish, HATolba and SI Shaheen. Segmentation and classification of white blood cells. Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. 2000; 2259-2261.

[24] Lim G C C, Rampal S, & Yahaya H. Cancer Incidence in Peninsular Malaysia- The Third Report of the National Cancer Registry, Malaysia. National Cancer Registry. 2008.

[25] Nilanjan Ray. A concave cost formulation for parametric curve fitting: Detection of leukocytes from intravital microscopy images. IEEE International Conference on Image Processing. 2010.

[26] Anjali Gautam, Priyanka Singh, Balasubramanian Raman, Harvendra Bhadauria. Automatic classification of leukocytes using morphological features and Naïve Bayes classifier. IEEE Region 10 Conference. 2016.

[27] Filip Novoselnik, Ratko Grbic, Irena Galic, Filip Doric. Automatic White Blood Cell Detection and Identification Using Convolutional Neural Network. IEEE International Conference on SST. 2018.

[28] S Dhiravida Janani, R Muthu Selvi, G Mlndhu. Blood Cell Detection and Counting Using Convolutional Sparse Dictionary Learning. IEEE International Conference on Current Trends towards Converging Technologies. 2018.

[29] Anjali Gautam, Harvindra Bhadauria. Classification of white blood cells based on morphological features. IEEE International Conference on Advances in Computing, Communications and Informatics. 2014.

[30] Muhammad Sajjad, Siraj Khan, Muhammad Shoaib, Hazrat Ali, Zahoor Jan, Khan Muhammad, Irfan Mehmood. Computer Aided System for Leukocytes Classification and Segmentation in Blood Smear Images. IEEE International Conference on Frontiers of Information Technology. 2016.

[31] Shubham Manik, Lalit Mohan Saini, Nikhil Vadera. Counting and classification of white blood cell using Artificial Neural Network. IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems. 2016.

[32] Zahra Khandan Khadem Alreza, Alireza Karimian. Design a new algorithm to count white blood cells for classification Leukemic Blood Image using machine vision system. IEEE International Conference on Computer and Knowledge Engineering. 2016.

[33] Bruno C Gregorio da Silva, Ricardo J. Ferrari. Detection of Leukocytes in Intravital Video Microscopy Based on the Analysis of Hessian Matrix Eigenvalues. SIBGRAPI Conference on Graphics, Patterns and Images. 2015.

[34] S. Agaian, M. Madhukar, A. T. Chronopoulos, Automated screening system for acute myelogenous leukemia detection in blood microscopic images, IEEE Systems journal 8. 2014.

DSAE- Deep Stack Auto Encoder and RCBO- Rider Chaotic Biogeography Optimization Algorithm for Big Data Classification

Anilkumar V. Brahmane [1], Dr B Chaitanya Krishna

Department of Computer Science and Engineering, Koneru Lakshmaiah Education
Foundation, Vaddeswaram, AP, India

Abstract – In today's era Big data classification is a very crucial and equally widely arise issue is many applications. Not only engineering applications but also in social, agricultural, banking, educational and many more applications are there in science and engineering where accurate big data classification is required. We proposed a very novel and efficient methodology for big data classification using Deep stack encoder and Rider chaotic biogeography algorithms. Our proposed algorithms are the combinations of two algorithms. First one is Rider Optimization algorithm and second one is chaotic biogeography-based optimization algorithm. So, we named it as RCBO which is integration is ROA and CBBO. Our proposed system also uses the Deep stack auto encoder for the purpose of training the system which actually produced the accurate classification. The Apache spark platform is used initial distribution of the data from master node to slave nodes. Our proposed system is tested and executed on the UCI Machine learning data set which gives the excellent results while comparing with other algorithms such as KNN classification, Extreme Learning Machine Random Forest algorithms.

Keywords – Deep learning, Big Data Classification, Apache Spark

1. Introduction

Data mining algorithms are used for extracting the meaningful data from the big data. Various applications are available in day-to-day life where these data mining algorithms are playing important role for classifications and clustering.

Some of such algorithms are KNN- K- Nearest Neighbor classification, ELM- Extreme Learning Machine, Random Forest algorithm and many more. We study all these above algorithms and find out that each one of these is having some advantages and disadvantages. No doubt that these algorithms are giving the accurate results. But as the Vs (Volume, velocity, variety, veracity **and** value) related to big data is changing from platform to platform, from application to applications the data which is generating is heterogeneous in nature and many times its difficult to achieve the accurate

---

[1] Anilkumar V. Brahmane,Research Scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, India.

Email: brahmaneanilkumarcomp@sanjivani.org.in.

classification. Benefits of KNN are; 1) No Training Period 2) Since the KNN calculation requires no preparation prior to making forecasts, new information can be added flawlessly which won't affect the precision of the calculation. 3) KNN is extremely simple to execute. Drawbacks of KNN are; 1) doesn't function admirably with enormous dataset 2) Doesn't function admirably with high measurements 3) Need highlight scaling 4) Touchy to loud information, missing qualities and exceptions. Similarly, ELM – extreme machine learning is a proficient learning calculation for the single secret layer feed forward neural organizations. Contrasted and the other customary neural organization calculation it has the benefit of over-fitting issues and moderate preparing speed.

For accurate classification neural network plays crucial role. Deep learning is essential for a more extensive group of AI techniques. Learning can be regulated, semi-directed or unaided. After training the deep neural network efficiently we can achieve the accurate results. And as far as the data analytics is concern the Apache Spark is the widely used data processing and data analysis tool for big data. The purpose our research is for accurate data classification using Apache Spark based on Optimization algorithms. We used the RCBO that is Rider Chaotic Biogeography Based Optimization. This is the hybrid method made up of ROA- Rider Optimization Algorithm and CBBO – Chaotic Biogeography Based Optimization. The paper consists of; section 1 is Introduction to big data classification. Section 2 is related work and pros and cons. Section 3 proposed big data classification methodology. Section 4 is result analysis. Section 5 is conclusion.

2.   Related work – Pros and Cons

Enough work was carried out in past related to classification of big data. We studied some of this existing system and tried to learn the Pros and Cons about it. Some of it is listed here.

S. Ramírez-Gallego, mentioned in his paper about the incremental use of instance selection for big data [1]. This paper is having the pros that it constantly improves the performance. The Cons is that this paper is not addressing anything about huge size and redundancy problems. M.Duan address in his paper about ELM – extreme learning machine based on Spark Framework- SELM [2]. The advantage of this system is that it focuses on highest speedup due to SELM. The disadvantages are that more memory is needed for this system. Elsebakhi, E., mentioned in his paper about new large-scale machine learning classifier [3]. Good thing in his paper that it mentioned about saving the computational time. Cons about this system, that the performance of system is better with minimal values only. W. Lin focused on ensemble random forest algorithm with parallel computations [4]. This system possesses the strong point it uses SVM which yield performance. Problem with this system is that prediction is not accurate. Hernández address in his paper about machine learning to optimize parallelism in big data [5]. This system accurately predicts the execution time. But suffered due to complexity. Ramírez-Gallego mentioned in his paper about distributed discretization algorithm [6]. The advantage of this system is that its accuracy and simplicity. But concept of drift is affecting this system. Karim addresses in his research work about ASP-Tree Construction Algorithm [7]. The best point about this system it reduces the search time and space time. The problem with the system is that difficult to generate

gigantic synthetic dataset. B. Zhao proposed the LDA training system [8]. The pros about this system that it's having good scalability. The cons are that auto-tuning is major challenges of this system.

After learning and understanding the above system thoroughly we come to many conclusions and eventually reach to the point where we found the challenges in the existing system which are listed below. We also focus how to overcome these challenges.

1. Apache Spark is outstanding for its capacity to manage gigantic data using circled memory with an open-source stage. Regardless, the display of the Spark is trying issue while procuring the best yield from the Spark as the settings of Spark game plan using colossal limits configuration ominously impact the show in a gigantic degree.
2. The lively augmentation of advancement provoked the partner of tremendous data which came to appoint where the normal data taking care of are insufficient to offer the lucid and computational game plans.
3. In the programmable gathering condition challenged a couple of obstructions. From the start, different applications should be improved meanwhile and should have been taken care of with data enlisting which is a troublesome endeavor. Likewise, the flexibility of issue in central yet a tangled task. Moreover, the gathering masterminds the enlisting resources in an amazing manner which intensifies the block in the application. Along these lines, the clsuet enrolling needs a functioning response for perform different calculations.
4. The huge issue in changing data and isolating features joins obtainment of obvious data and encoding the data in numerical portrayal. Be that as it may, the extraction of basic data from the substance data and managing picture or sound data is a critical issue.
5. Attaining a totally suitable data model using Spark and MPI to give versatility, interoperability, and execution improvement adept for consistent data processing is a huge issue not fulfilled by any standard stage.These challenges are summarized as below;
6. Apache Spark configurations which can be reconfigure to achieve the better performance.
7. Cluster computing to support large scale data.
8. Cluster's environment has to be re- written in parallel manner.
9. Extracting useful features from big data.
10. Data model availability which suitable for Apache Spark architecture.

After finding above challenges, we decided our objectives which are;

1. To distribute the data equally among the salve's nodes.
2. To develop the feature vector comprising of optimized feature vector based on optimal weight.
3. To develop training algorithm using deep learning for accurate classification.
4. To get the accurate big data classification.

2.1 Existing algorithm problem

Apache Spark, popularly known for big data processing capability, is a distributed open-source platform that uses the concept of distributed memory to facilitate big data processing proficiently. From the aspect of performance, it is still a big challenge to obtain the best output from Spark, since the Spark configuration settings with large parameters configuration affect its performance at large extent [9]. The rapid development of technology has led to generation of large-scale data, and it has reached a point where sequential and traditional data processing model are not able to provide all the analytical and computational solutions. Hence, to overcome these challenges a number of clusters and distributed computing approaches and frameworks were put forward to support large-scale data intensive applications [10]. Programmable Clusters environment has brought several challenges: Firstly, many applications need to be rewritten in a parallel manner, and the programmable Clusters need to process more types of data computing; Secondly, the fault tolerance of the Clusters is more important and difficult; Thirdly, Clusters dynamically configure the computing resources between shared users, which increases the interference of the applications. With the rapid increase of applications, Cluster's computing requires a working solution to suit different calculations [11]. Common challenges during data transformation and feature extraction include: Taking categorical data (such as country for geolocation or category for a movie) and encoding it in a numerical representation. Extracting useful features from text data, Dealing with image or audio data [12]. Achieving a data model fully compatible for Spark and MPI that provides scalability, performance and interoperability suitable for scientific data assimilation remains a challenge not fully satisfied by any existing platform, and this is the goal of our framework [13]. The emerging industrial big data have '5V' characteristics (volume, velocity, variety, veracity, and value), which challenges the traditional prognostics models [14].

2.2 Proposed system benefits

The benefits of the big data classification are demonstrated in this section.

2.2.1. Banking
The process of managing and evaluating the data of banks and other financial services organizations contains huge amount of client data which includes personal and security information. In banking and finance, reference data, trade and market data, transaction data can be structured or unstructured based on the collected information. Thus, the classification of big data enables to manage all the data in one place in a structured manner.

2.2.2. Cloud Computing
The communication between the servers using information technology produces huge amount of data. These data needs to processed and stored for proper functioning. Thus, the cloud is used as an online storage model for processing huge amount of data.

2.2.3. Healthcare
The big data contains huge amount of data that is available for healthcare providers to monitor the health risks. Thus, the healthcare information and the rising care for health has adapted a big space in making strategic business decisions.

2.2.4. Data mining
The big data classification uses data mining for predicting the meaningful data instead of taking unnecessary data. It utilizes the data and analytics for identifying the best practices for the classification.

2.2.5. Stocks
It is simpler to analyze the trending stocks as per the classified result. The classification can help to check the interest of the people approaching the stocks

2.3   Technical clearness in proposed work

In proposed work the specialized angle is clarify underneath;

1.   We are utilizing the Apache Spark. Apache Spark is having starting hubs and last hubs.
2.   Our proposed work is to play out the order utilizing two stages; highlight choice and large information arrangement.
3.   These element choice and arrangement is acting in the underlying hubs and last hubs of Apache Spark.
4.   Big Data arrangement is begun in starting hub of Spark. We additionally called this underlying hub as expert hub.
5.   This ace hub accumulates the Big information from different sources over the web. This expert hub conveys this information among slave hubs.
6.   The slave hubs will play out the component choice utilizing ideal highlights choice utilizing proposed proficient improvement strategy. The name of this streamlining procedure is RCBO-Rider Chaotic Biography improvement
7.   These chose highlights by utilizing RCBO will be given to definite hubs of Spark for where the Deep learning calculation will be utilized to prepare the framework. In our propose work this part is significant and we carry out it utilizing Proposed RCBO based profound stack auto encoder.
8.   This Proposed RCBO based profound stack auto encoder will be train to arrange the large information. Here the outcomes from all slave hubs as ideal element chose will be given to group the information. Our proposed RCBO based profound stack auto encoder will play out the exact.
9.   We proposed the RCBO which is the novel procedure comprise of two things 1) Rider Optimization Algorithm (ROA) and 2) tumultuous biogeography-based-streamlining (CBBO).

3. Proposed system

This section explains the working of proposed system. As shown in fig. 1 the architecture of our proposed system consists of Apache Spark framework. The Apache Spark consists of Master node and Slave nodes.
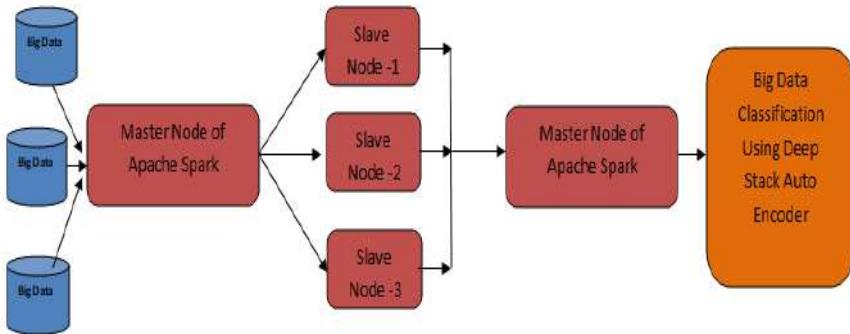
Figure 1. Architecture of the proposed system

As shown in the figure 1 the resources for the big data are various like agricultural data, stock market data, educational data, government data, employee data etc. This big data is collected and given to master node of the Apache Spark. Here in Master node the Big data is partition equally among the slave nodes of Apache Spark. These two nodes also known as initial nodes and final nodes of Spark. In the slave nodes the proposed Rider Chaotic Biogeography Optimization – RCBO is used to select the optimal features of the big data. Each slave nodes submit these selected optimal features to master node. In the master node the actual classification of big data take place. For accurate classification again the proposed RCBO based deep stack auto encoder algorithm is applied on optimal features.

Sources of Big data – Big data can be collected through various sources like government employee, agriculture, healthcare, clients etc.
Master node of Apache Spark- In the first phase the big data collected from various sources are given to master node of Apache Spark where it partitioned among slave nodes. In the second phase the selected optimal features are again integrated together from each slave nodes and fed to master node where the classification will take place.
Slave node of Apache Spark – The partitioned big data is given to slave nodes where the proposed RCBO algorithm is used to select the optimal features from big data. Each slave nodes runs the proposed RCBO algorithm for selecting the optimal features.
Deep stack auto encoder – The selected optimal features will be taken as the input and the neural network will be trained to classify the accurate data. Deep stack auto encoder is the ANN which is having encoder and decoder which decode the input and encode the output.
RCBO – rider chaotic biogeography-based optimization algorithm and deep stack auto encoder are both very novel and effective algorithms for our system.

3.1 Rider Chaotic Biogeography Based Optimization – RCBO

3.1.1. Essential BBO Algorithm

Biogeography-based improvement (BBO), recommended by Simon, is a novel populace based streamlining strategy for tackling worldwide advancement issues. It depends on the idea of biogeography, which is the investigation of the relocation, speciation, and elimination of species. In biogeography, natural surroundings imply a

biological region which is occupied by a specific plant or creature species and is geologically secluded from different environments. Every one of the territories is considered as a person with its environment appropriateness list (HSI) to gauge the decency for living. A living space with a high HSI demonstrates that it is more fit as living spots for natural species and will in general have countless species while an environment with a low HSI shows that it is less appropriate for species to live there and will in general have few species.

The elements of the development of the species among various living spaces is primarily administered by boundaries called migration and resettlement rate and these two boundaries relies on the quantity of species in the natural surroundings.

Mayhem hypothesis is a part of arithmetic zeroing in on the investigation of confusion—conditions of dynamical frameworks whose evidently arbitrary conditions of turmoil and inconsistencies are regularly administered by deterministic laws that are exceptionally touchy to beginning conditions. Chaos hypothesis is an interdisciplinary hypothesis expressing that, inside the obvious irregularity of turbulent complex frameworks, there are fundamental examples, interconnectedness, steady input circles, redundancy, self-closeness, fractals, and self-organization. The butterfly impact, a hidden standard of bedlam, portrays how a little change in one condition of a deterministic nonlinear framework can bring about huge contrasts in a later state (implying that there is delicate reliance on introductory conditions). A similitude for this conduct is that a butterfly fluttering its wings in China can cause a storm in Texas.

Little contrasts in starting conditions, for example, those because of blunders in estimations or because of adjusting mistakes in mathematical calculation, can yield broadly wandering results for such dynamical frameworks, delivering long haul forecast of their conduct unimaginable in general. This can happen despite the fact that these frameworks are deterministic, implying that their future conduct follows a novel evolution and is completely dictated by their underlying conditions, with no arbitrary components involved. at the end of the day, the deterministic idea of these frameworks doesn't make them predictable. This conduct is known as deterministic tumult, or essentially confusion. The hypothesis was summed up by Edward Lorenz as:

Mayhem: When the present decides the future, however the surmised present doesn't roughly decide what's to come.

Turbulent conduct exists in numerous regular frameworks, including liquid stream, heartbeat inconsistencies, climate and climate. It likewise happens immediately in certain frameworks with fake segments, for example, the financial exchange and street traffic. This conduct can be concentrated through the examination of a tumultuous numerical model, or through logical procedures, for example, repeat plots and Poincaré maps. Mayhem hypothesis has applications in an assortment of controls, including meteorology, anthropology, social science, physics, natural science, software engineering, designing, financial aspects, science, environment, pandemic emergency management, and theory. The hypothesis framed the reason for such fields of study as perplexing dynamical frameworks, edge of mayhem hypothesis, and self-get together cycles.

3.1.2. CBBO Algorithm

Because of the adaptability and power in tackling advancement issues, BBO calculation has effectively stimulated extraordinary interest. Notwithstanding, a few defects actually exist on this calculation, like the huge number of cycles to arrive at the

worldwide ideal arrangement and the propensity to meet to nearby best arrangements. To beat these defects of the old style BBO calculation, CBBO, which coordinates BBO with bedlam hypothesis, was proposed in our work. After the change activity of every age, lead the tumultuous pursuit to pick better arrangements into future. Along these lines, our proposed calculation exploits the qualities of the turbulent variable to make the people of sub ages circulated ergodically in the characterized space and consequently to keep away from the untimely of the people.

3.1.3. BBO Algorithm

The biogeography-based headway (BBO) estimation is gotten from the biogeography discipline, which is generally established on the transport of species in nature. Species have certain rules according to which development is coordinated among segregated islands through various obstacles. Species can comprehend developments among these islands by drifting, using the breeze, and various ways. The means in the calculations and detail flowchart are appeared underneath.
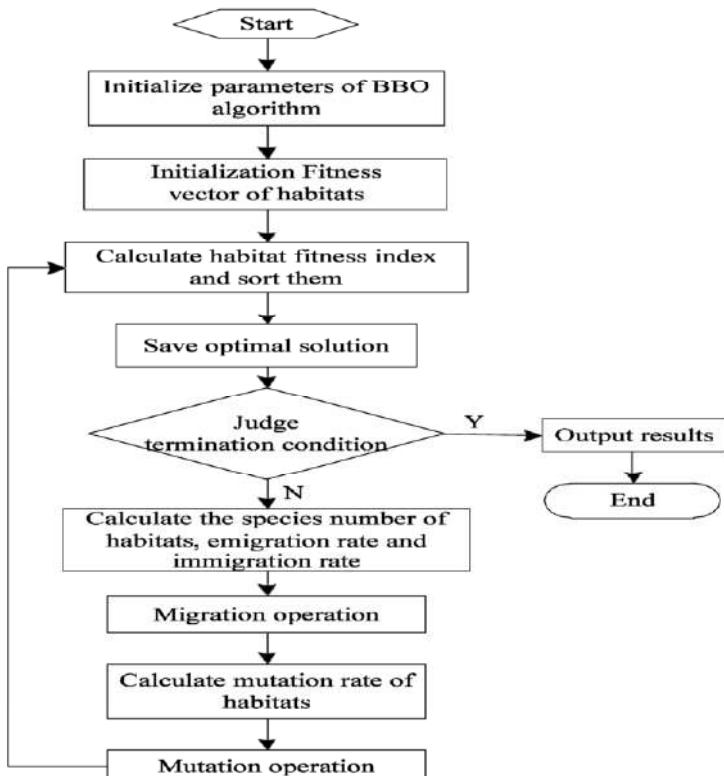
Figure 2. Steps for optimization algorithm

3.1.4. ROA Algorithm

The ROA considers a couple of rider gatherings, who travel to a typical objective area for turning into the champ of the race, to form its thought. The quantity of gatherings

considered is four, where the quantity of riders in each gathering is chosen similarly from the absolute number of riders. The four gatherings of riders are sidestepping rider, supporter, overtaker, and assailant. Every single gathering follows various techniques to arrive at the objective, as follows.

1. The detour rider means to arrive at the objective by bypassing the main way.
2. The supporter follows the main rider in a large portion of the pivot.
3. The overtaker follows his own situation to arrive at the objective, as per the close by area of the main rider.
4. The assailant takes the situation of the rider to arrive at the objective point, using the most extreme speed.

Despite the fact that the riders follow a predefined procedure, the principal variables to arrive at the objective are the right riding of the vehicle by legitimate treatment of the directing, stuff, gas pedal, and brake. For each time moment, the riders change their situations toward the objective by changing these boundaries and follow the predefined system dependent on the current achievement rate, which is contrarily relative to the distance between the situation of the riders and the objective.

The main rider is characterized dependent on the achievement rate at the current time moment. This cycle is proceeded, until the riders go into off time, which is the most extreme time given for the riders to arrive at the objective. After the off time, the rider, who is the main rider, is named as the champ of the riding race. By following this anecdotal idea, another advancement calculation is being created, as portrayed in figure 3.
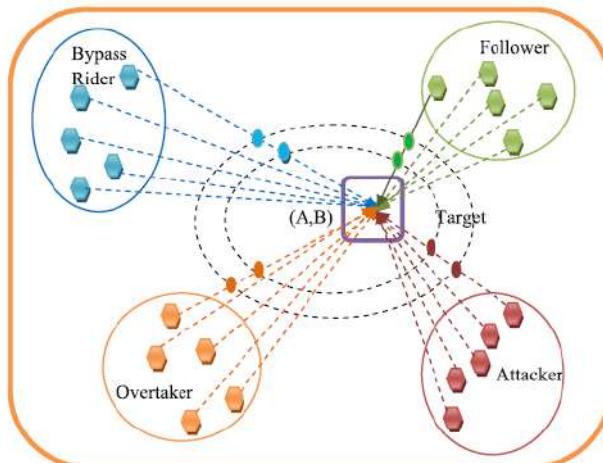
Figure. 3. Overview of ROA

Our main contribution to this system is the implementation of RCBO. This RCBO is the integration of two algorithms one is ROA- Rider optimization algorithm and second is CBBO- Chaotic Based Biogeography Optimization.

In every slave node of Apache Spark Feature selection is done by using RCBO. Optimal features are selected using following steps;

1) Big data is partition in to different subsets.
2) Subsets are equal to slave nodes, where the feature selection is carried out.
3) Solution vector is created which consist of selected features.
4) The optimal features are selected depending upon the fitness function.
5) Fitness function is based on minimization problem.
6) Solution provides less MSE is selected as the efficient solution.

3.2 Proposed system algorithm for selecting the optimal features

Our proposed system for big data classification is implemented with the help of combinations of two optimization algorithms. They are ROA – Rider Optimization Algorithm and CBBO – Chaotic Biography based Optimization Algorithm. The for the optimal feature's selection the proposed RBCO algorithm is used. Rider optimization algorithm is based on the concept of a group of riders riding towards achieving their goal. Chaotic Biography optimization algorithm is based on the concept of species migration.

1) Initialize the population
2) Calculate the fitness function
3) Find out the optimal solution

The first step is to initialize the all-available solutions which are calculated based on the fitness functions. This population of the available solution can be represented by the set of solutions.

The fitness function is used to calculate the best solution. The Mean Square Error this fitness function is use to calculate the optimal solution.

To find out the optimal solution using proposed RCBO, we use the characteristics of CBBO and characteristics of ROA algorithms. By integrating the characteristics of these tow algorithms, we get the proposed RCBO algorithm.

Based on the selected optimal solutions which are nothing but the optimal features the feature vector is created. This features vector is consisting of the optimal solutions. So, we can form a set of optimal solutions which is called as features vector.

3.3 RCBO based Deep stack auto encoder

After getting the optimal features as a result of RCBO, these optimal features are again feeds to master node of the Apache Spark. In master node all collected optimal features from all slave nodes are integrated and the classification process is carried out on this selected and integrated optimal feature.

For accurate classification we used the training neural network, which is deep neural network we term it as deep stack auto encoder. Fig .4 shows the architecture of training neural network for accurate classification.
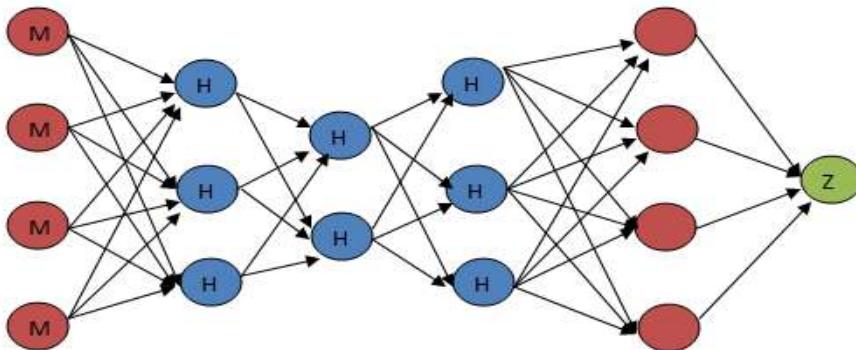
Figure 4. Training the neural network using Deep stack auto encoder

As shown in the figure 4 the input to the neural network is the optimal features which are selected initially using the RCBO. Multiple hidden layers are used to shuffle the input vector to yield the better accurate results. The training to this deep stack auto encoder is done using the RCBO – rider chaotic biogeography-based optimization algorithm.

3.4 Proposed steps for big data classification using the deep stack auto encoder

The feature vector which is obtained in optimal feature selection is given as input to big data classification. In classification process the deep stack auto encoder is used as shown in the figure 2. This is the ANN which is trained to give the accurate classification. This auto encoder is having the input layer, hidden layers, and output layer. The auto encoder is having encoder and decoder which gives the accurate classification.

The proper weights and the bias is given to this neural network for producing the accurate results The proposed RCBO is used here again to train the neural network. This step is executed in the master node of apache spark.

4.   Result analysis

4.1. Dataset

We use the Cover Type Data Set which is the UCI machine learning dataset. The details about the dataset are given below;

Table 1. Characteristics of the dataset

Data Set Characteristics:	Multivariate	Number of Instances:	581012	Area:	Life
Attribute Characteristics:	Categorical, Integer	Number of Attributes:	54	Date Donated	1998-08-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	282406

4.2. Informational index Information

Anticipating woodland cover type from cartographic factors just (no distantly detected information). The genuine timberland cover type for a given perception (30 x 30 meter cell) was resolved from US Forest Service (USFS) Region 2 Resource Information System (RIS) information. Free factors were gotten from information initially acquired from US Geological Survey (USGS) and USFS information. Information is in crude structure (not scaled) and contains parallel (0 or 1) sections of information for subjective autonomous factors (wild regions and soil types).

This investigation region incorporates four wild regions situated in the Roosevelt National Forest of northern Colorado. These regions address timberlands with negligible human-caused unsettling influences, so that current woodland cover types are more an aftereffect of environmental cycles instead of backwoods the board rehearses.

Some foundation data for these four wild regions: Neota (region 2) presumably has the most noteworthy mean elevational worth of the 4 wild regions. Rawah (region 1) and Comanche Peak (region 3) would have a lower mean elevational esteem, while cache la Poudre (region 4) would have the most reduced mean elevational esteem.

Concerning essential significant tree species there, Neota would have tidy/fir (type 1), while Rawah and Comanche Peak would likely have lodgepole pine (type 2) as their essential species, trailed by tidy/fir and aspen (type 5). Store la Poudre would will in general have Ponderosa pine (type 3), Douglas-fir (type 6), and cottonwood/willow (type 4).

The Rawah and Comanche Peak regions would will in general be more commonplace of the by and large dataset than either the Neota or Cache la Poudre, because of their arrangement of tree species and scope of prescient variable qualities (rise, and so on) Cache la Poudre would most likely be more extraordinary than the others, because of its moderately low height territory and species creation.

4.3. Attribute Information

Given is the trait name, characteristic sort, the estimation unit and a concise depiction. The timberland cover type is the characterization issue. The request for this posting compares to the request for numerals along the columns of the data set.

Table 2. Attribute Information

Name	Data type	Measurement	Description
Elevation	quantitative	meters	Elevation in meters
Aspect	quantitative	Azimuth	Aspect in degrees azimuth
Slope	quantitative	degrees	Slope in degrees
Horizontal_Distance_ To_Hydrology	quantitative	meters	Horz Dist to nearest surface water features
Vertical_Distance_To _Hydrology	quantitative	meters	Vert Dist to nearest surface water features
Horizontal_Distance_ To_Roadways	quantitative	meters	Horz Dist to nearest roadway
Hillshade_9am	quantitative	0 to 255 index	Hillshade index at 9am, summer solstice
Hillshade_Noon	quantitative	0 to 255 index	Hillshade index at noon, summer soltice
Hillshade_3pm	quantitative	0 to 255 index	Hillshade index at 3pm, summer solstice
Horizontal_Distance_ To_Fire_Points	quantitative	meters	Horz Dist to nearest wildfire ignition points
Wilderness_Area	qualitative	0 (absence) or 1 (presence)	Wilderness area designation
Soil_Type	qualitative	0 (absence) or 1 (presence)	Soil Type designation
Cover_Type (7 types)	integer	1 to 7	Forest Cover Type designation

We tested and executed our system using the UCI machine learning data sets forest cover type data set. The analysis parameters we used are; Accuracy, Sensitivity, Specificity. We compare our results with ELM, K-NN and random forest algorithms. Every time our results are better than these classification algorithms.

We tested the system using tanning data. We got the better results are compare to ELM, K-NN and random forest algorithms. Similarly, we tested the system using number of features selected again we got the better results as compare to ELM, K-NN, and random forest algorithm. These results of our system are shown in the fig. 5.

For the purpose of performance analysis the proposed system is executed and tested with respective accuracy, sensitivity, and specificity.

Our proposed system is executed on above data set. We get following results.

1. Performance analysis – After comparing the 50% of training data we observed that our proposed algorithms give better exactness. The various layers of deep stack auto encoder give the better results.
2. The responsiveness also gives the better results at various layers.
3. The specificity also gives the better results at various layer of learning algorithm.
4. Performance analysis - We compare the proposed system upon selecting the features. We observed that the proposed system gives the better results at various layers of learning algorithm.
5. We compare the proposed system upon selecting 5 features and we get better results for exactness, responsiveness and specificity.
6. Comparative analysis - We compare the proposed system against the algorithms such as ELM, Random Forest, K-NN algorithms.
7. We test the results using 50 % training data against above algorithms; we observed our proposed system gives better results for parameters such as exactness, responsiveness and specificity.
8. Comparative analysis - We tested our system using 5 selected features and observed better results against such
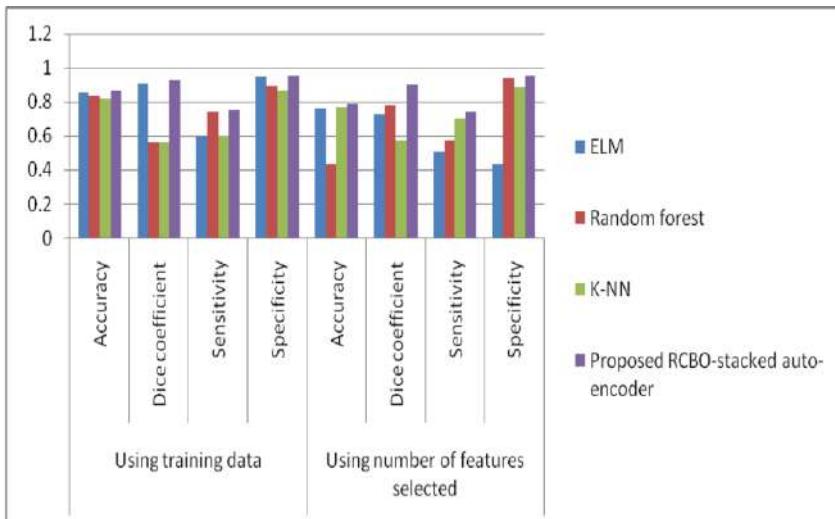
Figure 5. Result Analysis

5. Conclusion

In this paper we propsed the system using Rider Chaotic Biogeography Based Optimization algorithm and using the deep stack auto encoder for accurate classification of big data. We use the Apache Spark framework for gathering the big data in the master node of Apache Spark and then finding the optimal features of this

big data in the slave node of Apache Spark. Our proposed RCBO – algorithm is the integration of the ROA – rider chaotic biogeography optimization and CBBO- Chaotic Biogeography based Optimization algorithm. We proposed the deep stack auto encoder as a training deep neural network for getting the accurate classification of the big data. We tested out system with existing algorithms like ELM, K-NN, and random forest algorithms. We obtained the better results as compare to these existing algorithms.

References

[1] S. Ramírez-Gallego, B. Krawczyk, S. García, M. Woźniak, J. M. Benítez and F. Herrera, "Nearest Neighbor Classification for High-Speed Big Data Streams Using Spark," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, no. 10, pp. 2727-2739, October 2017.

[2] M. Duan, K. Li, X. Liao and K. Li, "A Parallel Multi classification Algorithm for Big Data Using an Extreme Learning Machine," in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 6, pp. 2337-2351, June 2018.

[3] Elsebakhi, E., Lee, F., Schendel, E., Haque, A., Kathireason, N., Pathare, T., Syed, N. and Al-Ali, R., "Large-scale machine learning based on functional networks for biomedical big data with high performance computing platforms," Journal of Computational Science, vol. 11, pp. 69-81, 2015.

[4] W. Lin, Z. Wu, L. Lin, A. Wen and J. Li, "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis," in IEEE Access, vol. 5, pp. 16568-16575, 2017.

[5] Hernández, Á.B., Perez, M.S., Gupta, S. and Muntés-Mulero, V., Using machine learning to optimize parallelism in big data applications. Future Generation Computer Systems, vol. 86, pp.1076-1092, 2018.

[6] Ramírez-Gallego, S., García, S., Benítez, J.M. and Herrera, F., "A distributed evolutionary multivariate discretizer for big data processing on apache spark," Swarm and Evolutionary Computation, vol. 38, pp. 240-250, 2018.

[7] Karim, M.R., Cochez, M., Beyan, O.D., Ahmed, C.F. and Decker, S., "Mining maximal frequent patterns in transactional databases and dynamic data streams: a spark-based approach," Information Sciences, vol. 432, pp.278-300, 2018.

[8] B. Zhao, H. Zhou, G. Li and Y. Huang, "ZenLDA: Large-scale topic model training on distributed data-parallel platform," in Big Data Mining and Analytics, vol. 1, no. 1, pp. 57-74, March 2018

[9] M. A. Rahman, J. Hossen and V. C, "SMBSP: A Self-Tuning Approach using Machine Learning to Improve Performance of Spark in Big Data Processing," 2018 7th International Conference on Computer and Communication Engineering (ICCCE), Kuala Lumpur, pp. 274-279, 2018.

[10] S. Srivastava, A. Nigam and R. Kumari, "Work-in-Progress: Towards Efficient and Scalable Big Data Analytics: Mapreduce vs. RDD's," 2017 International Conference on Information Technology (ICIT), Bhubaneswar, pp. 272-275, 2017.

[11] Z. Han and Y. Zhang, "Spark: A Big Data Processing Platform Based on Memory Computing," 2015 Seventh International Symposium on Parallel Architectures, Algorithms and Programming (PAAP), Nanjing, pp. 172-176, 2015.

[12]]J. Fu, J. Sun and K. Wang, "SPARK – A Big Data Processing Platform for Machine Learning," 2016 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), Wuhan, pp. 48-51, 2016.

[13] S. Caíno-Lores, J. Carretero, B. Nicolae, O. Yildiz and T. Peterka, "Spark-DIY: A Framework for Interoperable Spark Operations with High Performance Block-Based Data Models," 2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT), Zurich, pp. 1-10, 2018.

[14] J. Yan, Y. Meng, L. Lu and C. Guo, "Big-data-driven based intelligent prognostics scheme in industry 4.0 environment," 2017 Prognostics and System Health Management Conference (PHM-Harbin), Harbin, pp. 1-5, 2017.

Predicting COD and BOD Parameters of Greywater Using Multivariate Linear Regression

Samir Sadik Shaikh [a, 1], Dr. Rekha Shahapurkar [a]
a Computer Science and Engineering Department, Oriental University, Indore, India

Abstract. Greywater reuse furthermore, reusing can be an incredible method to get non-consumable water. Since it contains broke down pollutions, greywater can't be utilized straightforwardly. As an outcome, it is critical to decide the nature of water prior to utilizing it. Body estimations require five days to finish, while COD estimations require only a couple of hours. Not exclusively improve models for evaluating water quality are required; however, a more coordinated methodology is additionally getting more normal. Most of these models require a wide scope of information that isn't in every case promptly available, making it a costly and tedious activity. Because of different issues in the enlistment with estimation included in water quality boundaries like BOD as well as COD, the principal objective of this investigation is to track down the best multivariate direct relapse models for foreseeing complex water quality outcomes. The code was written in Python for multi-variable information sources, and a Linear Regression Model was created. The projected COD versus estimated COD chart shows that the noticed and expected qualities are practically the same. The R-squared worth was 0.9973. A plot of extended BOD as an element of COD is likewise remembered for the outcome.

Keywords. ANN, BOD, COD, Greywater, Multivariate Linear Regression.

1. Introduction

Alternative water management strategies have been set up in dry territories because of the absence of fresh water. Almost 97% of the world's absolute water supply is found in the seas, yet only 3% of it appropriate for direct use [1]. Greywater is squandered water that is generally made by kitchen sinks, showers, clothing or clothes washers, cooling outlets, and other comparable gadgets. As indicated by information, greywater age fluctuates somewhere in the range of 39 to 85 percent in various nations [2] Greywater treatment and reuse will incorporate non-consumable water for latrine flushing, cultivating, vehicle cleaning, and floor washing, in addition to other things. Table 1 shows characteristics of greywater [3, 4].

---

[1] Samir Sadik Shaikh, Computer Science and Engineering Department, Oriental University, Indore
Email: samirshaikh81@gmail.com.

Table 1. Characteristics of Greywater

Parameters	Units	Values
pH	---	7.3 - 8.1
EC	$\mu S\ cm^{-1}$	489 - 550
Turbidity	NTU	20.6 - 38.7
Total Suspended Solids (TSS)	$mg\ L^{-1}$	12 - 17.6
Nitrate (NO_3^-)	$mg\ L^{-1}$	0.5 - 0.63
Total Nitrogen (TN)	$mg\ L^{-1}$	42.8 - 57.7
Phosphate (PO_4^{3-})	$mg\ L^{-1}$	1.52 - 3.36
BOD	$mg\ L^{-1}$	56 – 100
COD	$mg\ L^{-1}$	244 – 284
Total Caliform (TC)	CFU/100 mL	3.74×10^4 to 3.8×10^4
Na	$mg\ L^{-1}$	43.8 – 48.1
K	$mg\ L^{-1}$	8.3 – 15.2
B	$mg\ L^{-1}$	1.3 – 1.5
Cl^-	$mg\ L^{-1}$	7.4 – 12.9

The Clean Water Act was sanctioned in the mid-1970s, trailed by the making of the USEPA, which finished in the characterization of wastewater quality for natural benefit dependent on four principle rules [5]:

- **Physical Properties:** e.g. pH, turbidity, temperature, colour, and odour.
- **Solids:** e.g., Total Solids (TS), Total Suspended Solids (TSS), Total Dissolved Solids (TDS), Total Volatile Solids (TVS), and Total Fixed Solids (TFS).
- **Organics:** e.g., Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Total Organic Carbon (TOC), and Oil and Grease (O&G).
- **Nutrients:** e.g., TN (Total nitrogen) and TP (Total phosphorous).

Various water quality lists have been utilized in numerous ordinary investigations comparable to different water sources like lakes, waterways, and dam supplies [6-8]. The Trophic State Index (CTSI), set up via Carlson in 1977 [9], is ordinarily utilized by water the board offices and associations throughout the planet. The CTSI is a typical and valuable water quality record that has been utilized as the essential measurement in numerous examinations [10-11]. CTSI is determined utilizing three separate water quality variables: chlorophyll-a (Chl-a) fixation, total phosphorus (TP) focus, and Secchi depth (SD). Substance tests, estimations, and careful estimations of water tests are utilized to decide the centralizations of chlorophyll-an and complete phosphorus. Secchi depth, then again, can be physically estimated in repositories without the utilization of compound examinations or present-day innovation, however, it is likely the most unpredictable boundary because of its reliance on temperature (counting season, turbidity, and different variables) [12].

Because of the trouble of ascertaining chlorophyll-an and complete phosphorus fixations, various water quality records are used, together with turbidity, electrical conductivity, natural oxygen request focus or biochemical oxygen demand focus (BOD), chemical oxygen demand fixation (COD), and all-out total dissolved solids (TDS). A considerable lot of these are the most normally utilized boundaries for surveying water quality in Artificial Intelligence (AI) methods [13-17]. Chang and Liu (2015) suggested a fluffy back spread neural organization model to order the level of eutrophication because of the shakiness of trophic status dictated by TDS, BOD, and COD due to temperamental turbidity. Not exclusively improve models for evaluating water quality are required; however, a more methodical methodology is additionally getting more normal. Lately, assortments of AI-based approaches have been used towards address water quality issues; also AI holds a ton of guarantees around here (Chau, 2006). The utilization of AI to acquire useful connections among information dependent on chronicled info and yield information is at the core of AI. Fake neural organizations, choice trees (DTs), straight relapse, and the assistance vector machine are the most regularly utilized information-digging calculations for this reason in related works. In ANN-based applications [18-21] and SVM-based applications [22-25], some of them are utilized independently for the forecast.

In a few settings, the practical connection between covariates (otherwise called input factors) and reaction factors (otherwise called yield factors) is of extraordinary interest. When demonstrating complex sicknesses, for instance, potential danger factors and their impacts on the infection are explored to decide hazard factors that can be utilized to improve preventive or mediation techniques. Fake neural organizations can estimate any complex useful relationship. Rather than summed up straight models [26], it isn't critical to characterize the type of connection among covariates and reaction factors ahead of time. Thus, fake neural organizations are a compelling factual instrument. They are GLMs' immediate augmentations, and they can be utilized similarly. The neural organization is prepared utilizing noticed information, and it iteratively adjusts its boundaries to gain proficiency with a guess of the relationship [27].

In any interaction industry, execution lists like biochemical oxygen interest (BOD) and compound oxygen interest (COD) are used to decide the nature of wastewater created. Body plus COD are characteristic boundaries in place of sewer water quality. The body stays an expected pointer for the measure included in biochemically degradable natural matter found in a water test aimed at homegrown wastewater. COD estimations should be possible surprisingly fast versus five days for BOD estimations, regardless of the way that COD qualities are consistently higher than BOD esteems. The at present accessible technique for figuring BOD and COD is tedious and defenseless against estimation blunders. To deal with the accepted procedures for water quality protection, a few water quality models, like ordinary unthinking methodologies, have been made. Most of these models require a wide scope of information that isn't in every case promptly available, making it a costly and tedious activity [28]. Lately, the Artificial Neural Network (ANN) procedure has acquired in prevalence. Dogan et al. [29] investigated the capacity about the ANN model on the way to increase the exactness of natural oxygen request assessment (BOD). By contrasting the discoveries

with noticed BOD levels, the limit of an ANN technique in BOD assessment in the Melen River was investigated in this report. Utilizing the ANN strategy with COD, water release, suspended strong, complete nitrogen, and all-out phosphorus, MSE of 708.01, normal supreme relative mistakes of 10.03 percent, and a coefficient of assurance of 0.919 were gotten. Rene and Saidutta [30] utilized ANNs to assess BOD and COD fixations dependent on quantifiable water quality lists. The ANN's capacity to anticipate BOD was better than COD, as per their outcomes. Akratos et al. [31] utilized an ANN model and plan conditions to foresee BOD and COD evacuation in even subsurface stream planned wetlands. The discoveries of the ANNs and the model plan condition were fundamentally the same as test proof from the writing. The outcomes showed that utilizing the ANN cycle, a reasonable connection could be gotten. COD evacuation was found to be unequivocally connected with BOD expulsion. What's more, a COD evacuation expectation condition was created.

Due to various issues in the registration and measurement of water quality such as BOD and COD, the main goal of this study is to find the best multivariate linear regression models for predicting complex water quality results.

2. Material and Method

2.1. Case Study

Throughout the span of 11 months, the informational index for this examination was gathered through the kitchen sink (May 2020-and in the long run picked for model development dependent on estimated estimations of various factors and their correlative investigation. The body is estimated by hatching a fixed water test for five days and ascertaining the oxygen misfortune March 2021). The examples were assembled and shipped off Vashi's Water Quality Testing Lab, and a Cumulative Report of Water Quality was acquired (see Figure 1). Components (factors) like pH, complete suspended strong (TSS), absolute suspended (TS), and water temperature (T) that influence water quality (BOD and COD) were distinguished from start to finish. In the event that examples are not weakened until hatching, microscopic organisms will drain the entirety of the oxygen in the jug before the test is finished. The test outcomes were determined utilizing the Standard Procedures of the American Public Health Association [32].

From the cumulative water quality report, values of std. deviation and deviation coefficients were calculated, as represented in Table 2. The value of SDx and CV are calculated as:

$$SDx = \sqrt{\frac{\sum_{i=1}^{n}(Xmean - Xi)^2}{n}} \qquad (1)$$

Figure 1. Cumulative Water Quality Report.

$$CV = \frac{SDx}{Xmean} \tag{2}$$

Table 3 shows the model domain boundary of the water quality parameter. Xmean, Xmax, Xmin, SDx, and CV denote the data set's mean, maximum, minimum, standard deviation, and deviation coefficient, respectively (derived from cumulative report and Table 1). Table 2 shows that the CV value for pH (0.06) is the lowest and it is highest for TSS (0.53).

Table 2. Calculation of Standard deviation (SDx) and Deviation coefficient (CV)

Sample No.	Temp (°C)	pH	Total Solids (mg/L)	Total Suspended Solids (mg/L)	C.O.D (mg/L)	B.O.D (mg/L)
SS/R/01/20	4	0.16	14981.76	11859.21	8949.16	2061.16
SS/R/02/20	0.04	0	707.56	2714.41	750.76	0.16
SS/R/03/20	5.76	0.25	25728.16	22171.21	15775.36	3294.76
SS/R/04/20	0.36	0.01	129.96	146.41	12.96	153.76
SS/R/05/20	1.21	0.04	2342.56	835.21	1128.96	547.56
SS/R/06/20	8.41	0.36	38966.76	36062.01	24211.36	4678.56
SS/R/07/20	1.21	0.04	7673.76	9044.01	5241.76	707.56
SS/R/08/20	3.24	0.09	14065.96	13712.41	8911.36	1648.36
SS/R/09/20	9	0.25	32616.36	25312.81	19432.36	4569.76
SS/R/10/20	5.76	0.16	22380.16	19071.61	13548.96	2872.96
SS/R/11/20	13.69	0.36	44352.36	32436.01	26049.96	6496.36
SS/R/12/21	25	0.49	74310.76	49773.61	42189.16	11793.96
SS/R/13/21	10.89	0.49	54943.36	52854.01	34819.56	6304.36
SS/R/14/21	14.44	0.64	73657.96	72846.01	46915.56	8172.16
SDx	**2.712537**	**0.488438**	**170.4736**	**157.8514**	**133.0782**	**61.70288**
X_{mean}	23.5	7.9	672.6	298.1	285.4	158.6
CV	**0.115427**	**0.061828**	**0.253455**	**0.529525**	**0.466287**	**0.389047**

Table 3. Water Quality Properties

Parameters	Unit	Xmin	Xmax	Xmean	SDx	CV
T	°C	18.5	27.3	23.5	2.71	0.1
pH	---	7.2	8.7	7.9	0.48	0.06
TS	mg/L	400	944	672.6	170.47	0.25
TSS	mg/L	75	568	298.1	157.85	0.53
COD	mg/L	80	502	285.4	133.1	0.47
BOD	mg/L	50	249	158.6	61.7	0.39

3. Multivariate Linear Regression

In a forecast issue, straight relapse (LR) is a relapse model that was intended to decide the connection among autonomous and subordinate factors [33]. This investigation utilizes multivariate straight relapse, which is one of the numerous types of direct relapse. The most essential strategy for setting up a connection between autonomous factors (noticed or estimated), otherwise called indicators or regressors, and a reliant variable, otherwise called the reaction variable, is various direct relapse (MLR). A summed up articulation for the model can be composed as follows:

$$Y= \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + C \qquad (3)$$

Where Y is the dependent variable, β_1, β_2, β_3, and β_4 are the coefficients of X_1, X_2, X_3, and X_4 respectively, and C is the block. The direct relapse strategy is like the condition of a straight line, given by $Y = ax + b$.

Statistical methods, for example, regression models, stand as the most effective tools aimed at examining every relationship amongst dependent and independent variables in lesser samples [21]. The MLR is a process for modeling the linear relationship among one or more independent variables and a dependent variable. MLR is based on least squares. In the best model, the sum of square error between observed and predicted parameters should be a minimum value. BOD and COD estimation also can be performed using linear models which explain the linear relationship between parameters. MLR is based on the principle of least squares. The sum of square errors between observed and predicted parameters should be as low as possible in the best model. Linear models that describe linear relationships between parameters can also be used to estimate BOD and COD. In addition, as shown in equation 4, the same input variables used in MLR models can also be used in linear models.

$$Y = \beta_1 T + \beta_2 pH + \beta_3 TS + \beta_4 TSS + e \qquad (4)$$

Where, Y represents COD values, $\beta_1, \beta_2, \beta_3, \beta_4$ as well as e are constant coefficients coming from the linear regression model, T, pH, TS also TSS are input factors which will determine the predicted value of COD for our model. Also, we will estimate the values of BOD using COD.

4. Results and Discussions

Google gives an online Google Colaboratory that can be utilized to compose and execute AI calculations in Python utilizing the online code supervisor. Along these lines, there is no compelling reason to introduce the libraries of python on a work area or PC. For the execution of the examination, we have utilized Google colab where the code was written in python and the Linear Regression Model was made for multivariable sources of info (T, pH, TS, TSS) by bringing in linear_model utilizing sklearn. OLS (Ordinary Least Square) technique was utilized to create the aftereffects of direct relapse as demonstrated in Figure 2.

Thus from the report, it can be seen that the values of intercept (marked as constant) and coefficients of $T, pH, TS,$ and TSS (marked as $x1, x2, x3,$ and $x4$) can be used to predict the values of COD for the described model. Substituting these values in equation (4) yields the following;

$$Y = -3.1004*T + 7.3189*pH + 0.6417*TS + 0.1808*TSS -185.5697 \qquad (5)$$

Along these lines, if a model is portrayed by equation (5) it will give us a best-fit model. The condition was executed in dominant utilizing T, pH, TS, TSS as info factors to anticipate COD as yield, characterized by equation (5). A portion of the outcomes that appeared in figure 3 are acquired by plotting single free factor ($T, pH, TS,$ and TSS) against the anticipated estimations of COD. It is done so in light of the fact that a straight fit can be best seen in situations where we have single autonomous and ward factors. In this manner, the element of perception will be a 2D plane (which is administered by the connection $p+1$ where p is the quantity of autonomous factors). As the quantity of autonomous variable expands, the element of noticed plane expansions in the same extent ($p+1$), consequently fitting model on a straight line gets

unpredictable. Such models are acknowledged utilizing the dissipate plots as demonstrated in figure 4.

```
                              OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 6.432e+05
Date:                Sun, 21 Mar 2021   Prob (F-statistic):           1.54e-24
Time:                        06:16:43   Log-Likelihood:               -0.39479
No. Observations:                  14   AIC:                             10.79
Df Residuals:                       9   BIC:                             13.98
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        -185.5697     28.918     -6.417      0.000    -250.986    -120.153
x1             -3.1004      2.323     -1.334      0.215      -8.356       2.155
x2              7.3189      4.296      1.704      0.123      -2.399      17.037
x3              0.6417      0.088      7.299      0.000       0.443       0.841
x4              0.1808      0.054      3.376      0.008       0.060       0.302
==============================================================================
Omnibus:                        0.394   Durbin-Watson:                   1.619
Prob(Omnibus):                  0.821   Jarque-Bera (JB):                0.034
Skew:                           0.106   Prob(JB):                        0.983
Kurtosis:                       2.889   Cond. No.                     2.70e+05
==============================================================================
```

Figure 2. OLS Regression Report.

(a)

(b)

(c)

(d)
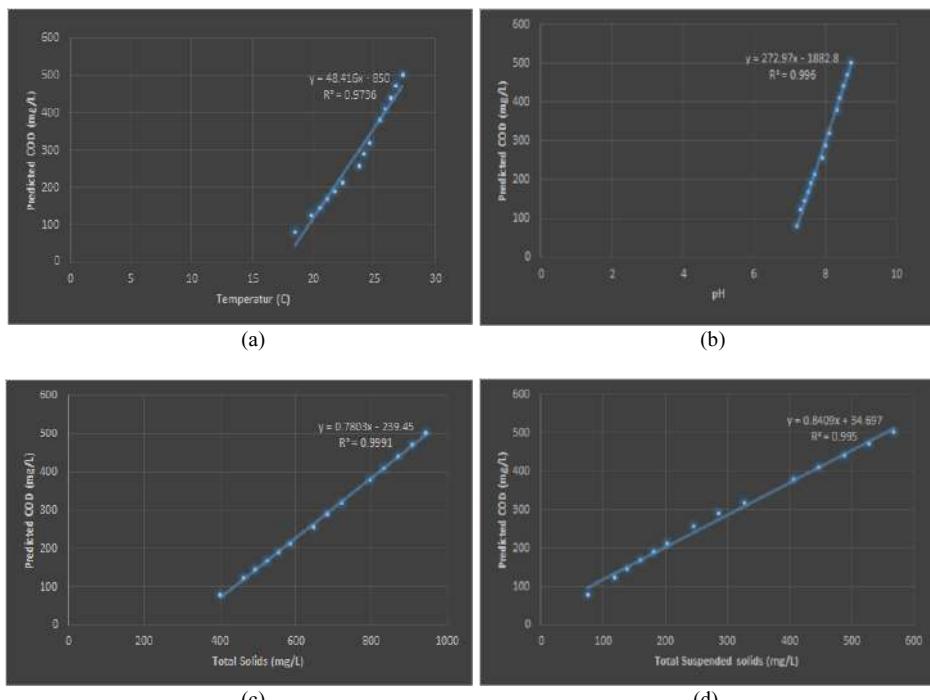
Figure 3 (a) Predicted COD v/s Temperature **(b)** Predicted COD v/s pH **(c)** Predicted COD v/s TS
(d) Predicted COD v/s TSS

(a)	(b)
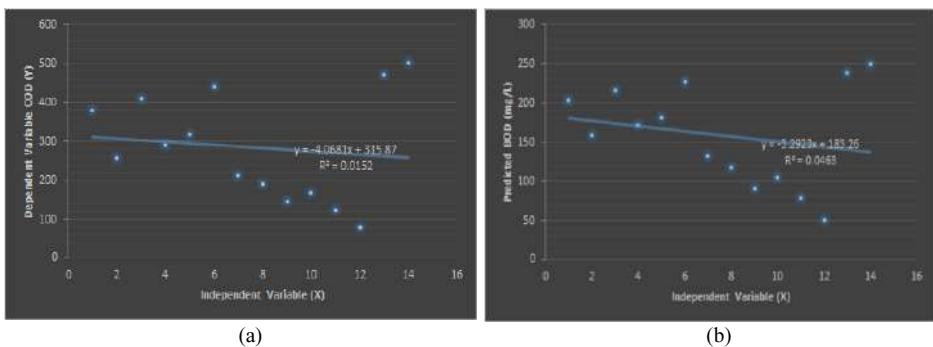
Figure 4 (a) Predicted COD v/s T, pH, TS, TSS **(b)** Predicted BOD v/s T, pH, TS, TSS

The main objective of a linear regression model is to estimate the difference between the predicted and observed (measured) value of the variable with the intention of validates the usefulness of the model. A response of predicted COD v/s measured COD is represented in figure 5 (a). Chemical investigation for COD measurement takes few hours, while BOD measurement takes 05 days; therefore it is also possible to predict BOD using values of COD Figure 5(b) represents Predicted BOD as a function of COD.

(a)	(b)
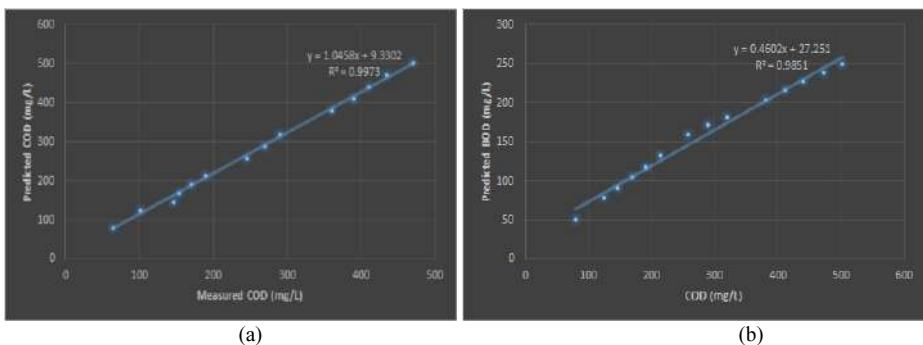
Figure 5 (a) Predicted COD v/s Measured COD **(b)** Predicted BOD v/s COD

5. Conclusion

As per information, greywater age fluctuates somewhere in the range of 39 and 85 percent in various nations. Greywater treatment and reuse can be utilized to give non-consumable water to latrine flushing, cultivating, vehicle and floor cleaning, and different employments. In any interaction industry, execution records like biochemical oxygen interest (BOD) with synthetic oxygen interest (COD) are utilized to decide the nature of wastewater produced (COD). The utilization of modern techniques like compound tests, conditions, and complex water test investigations is expected to gauge these amounts. The COD test requires a couple of hours, while the BOD test requires five days. Examinations in the lab are both tedious and costly. Not exclusively improve AI models for surveying water quality should be made, yet there is likewise an expanding interest for a more incorporated methodology. As of late, the Artificial Neural Network (ANN) strategy has acquired prominence. Instead of ANN, measurable

procedures like relapse models are the best techniques for exploring any connection among reliant and free factors with a restricted example size. The code was written in Python with the guide of Google Colaboratory, and a Linear Regression Model for multi-variable sources of info was created. For the model, the accompanying perceptions were made:

- The R-squared coefficient for the model was equal to1.00
- As the number of independent variables grows, the dimension of the observed plane grows in lockstep ($p+1$), making model fitting on a straight line more difficult. Scatter plots are used to build such models.
- The graph of predicted COD v/s measured COD shows a close approximation between observed and predicted value. The R-squared value was 0.9973.
- It is also possible to predict BOD using observed values of COD.

In our future work, we aim to analyse the data set by using other machine learning techniques such as ANN, SVM, etc., and comparing the attainment of the models based upon MAE also values based on RMSE.

References

[1] Oteng-Peprah M, Acheampong M, DeVries N, (2018) "*Greywater Characteristics, Treatment Systems, Reuse Strategies and User Perception—a Review*". Water Air, & Soil Pollution; volume 229, Article number: 255.

[2] Samayamanthula D, Sabarathinam C, Bhandary H, (2019), "*Treatment and effective utilization of greywater*", Applied Water Science; volume 9, Article number: 90.

[3] Manna S, (2018), "*Treatment of Gray Water for Reusing in Non-potable Purpose to Conserve Water in India*". International Journal of Applied Environmental Sciences, Volume 13, Number 8, pp.703-716.

[4] Sushmitha M, Chanakya H, Khuntia H, (2019) "Efficient Grey Water Treatment and Reuse Options for India—A Review". Springer.2019; 225-232.

[5] Barzegar G, Wu J, Ghanbari F, (2019) "Enhanced treatment of greywater using electrocoagulation/ozonation: investigation of process parameters". Process Safety Environment Prot.; 12:125–132.

[6] Parparov A, Hambright D, Berman T, (2014). "*Water Quality Assessment*", Ecology and Management. Springer Netherlands, Dordrecht, pp. 607-615.

[7] Sofwan N.M, Fong S.S, (2014). "*Evaluation of Various Water Quality Indices for Water Quality Assessment of Sg. Sarawak, In: Aris, From Sources to Solution*", Proceedings of the International Conference on Environmental Forensics 2013. Springer Singapore, Singapore, pp. 227-231.

[8] Walker D, Jakovljević D, Savić D, Radovanović M, (2015). "Multi-criterion water quality analysis of the Danube River in Serbia: A visualisation approach". Water Research 79, 158- 172.

[9] Carlson R.E, (1977). "*A trophic state index for lakes*". Limnology and Oceanography 22, 361-369.

[10] Prasad A.G.D, Siddaraju, (2012). "Carlson's Trophic State Index for the assessment of trophic status of two lakes in Mandya district". Advances in Applied Science Research 3, 2992-2996.

[11] Taheri Tizro A, Ghashghaie M, Nozari H, (2016). "*Assessment of Carson Trophic Index in Dam Lake: a Case Study of Ekbatan Dam*". Iranian Journal Of Health Sciences 4, 25-33.

[12] Chang C.L, Liu H.C, (2015). "Applying the Back-Propagation Neural Network model and fuzzy classification to evaluate the trophic status of a reservoir system". Environmental Monitoring and Assessment 187, 560-567.

[13] Chen W.C, Chang N.B, Chen J.C, (2003). "Rough set-based hybrid fuzzy-neural controller design for industrial wastewater treatment". Water Research 37, 95-107.

[14] Najah A, El-Shafie A, Karim O.A, El-Shafie A.H, (2013). "*Application of artificial neural networks for water quality prediction*". Neural Computing and Applications 22, 187-201.

[15] S.K, T J, A.H.M, M. A.S, (2014). "Prediction of Water Quality of Euphrates River by using Artificial Neural Network Model (spatial and temporal study)". International Research Journal of Natural Sciences 2, 25-38.

[16] Seshan H, Goyal M.K, Falk M.W, Wuertz, S, (2014). "Support vector regression model of wastewater bioreactor performance using microbial community diversity indices: Effect of stress and bioaugmentation". Water Research 53, 282-296.

[17] Verma A.K, Singh T.N, (2013). "*Prediction of water quality from simple field parameters*". Environmental Earth Sciences 69, 821-829.

[18] Chau K.W, (2006). "A review on integration of artificial intelligence into water quality modelling". Marine Pollution Bulletin 52, 726-733.

[19] Hamed M.M, Khalafallah M.G, Hassanien E.A, (2004). "*Prediction of wastewater treatment plant performance using artificial neural networks*". Environmental Modelling & Software 19, 919- 928

[20] Maier H.R, Dandy G.C, (2000). "Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications". Environmental Modelling & Software 15, 101-124.

[21] Wu W, Dandy G.C, Maier H.R, (2014). "Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling". Environmental Modelling & Software 54, 108-127.

[22] Guo Y, Wang G, Zhang X, Deng W, (2014). *"An improved hybrid ARIMA and support vector machine model for water quality prediction"*, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 411-422.

[23] Tarmizi A, Ahmed A.N, El-Shafie A, (2014). "Dissolved Oxygen Prediction Using Support Vector Machine in Terengganu River. Middle-East" Journal of Scientific Research 21, 2182-2188.

[24] Zhao W.G, Wang L.Y, (2010). "Water quality evaluation based on Multiclass Fuzzy Support Vector Machine", Advanced Materials Research, pp. 708-711.

[25] Jui-Sheng Chou, Chia-Chun Ho, Ha-Son Hoang, (2018*), "Determining quality of water in reservoir using machine learning"*, Ecological Informatics.

[26] P. McCullagh. J. Nelder. (1983) "*Generalized Linear Models*". Chapman and Hall, London.

[27] Frauke Günther, Stefan Fritsch, (2010), "*neuralnet: Training of Neural Networks*", The R Journal Vol. 2/1.

[28] Hamid Zare Abyaneh, (2014), "Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters", Journal of Environmental Health Science & Engineering, 12:40.

[29] Dogan E, Ates A, Ceren Yilmaz E, Eren B (2008), "Application of Artificial Neural Networks to Estimate Wastewater Treatment Plant Inlet Biochemical Oxygen Demand". Environ Prog, 27(4):439–445

[30] Rene ER, Saidutta MB, (2008), "Prediction of Water Quality Indices by Regression Analysis and Artificial Neural Networks". Int J Environ Res, 2(2):183–188

[31] Akratos CS, Papaspyros JNE, Tsihrintzis VA, (2008), "An artificial neural network model and design equations for BOD and COD removal prediction in horizontal subsurface flow constructed wetlands". Chem Eng J, 143:96–110.

[32] APHA (American Public Health Association): (1995), "*Standard Methods for the Examination of Water and Wastewater*". 19th edition. Washington, DC.

[33] Chou J.S, Yang K.H, Lin J.Y, (2016c), "Peak Shear Strength of Discrete Fiber-Reinforced Soils Computed by Machine Learning and Metaensemble Methods". Journal of Computing in Civil Engineering 30, 1-17.

Customer Segmentation Using Machine Learning

Prof. Nikhil Patankar [a,1], Soham Dixit [a], Akshay Bhamare [a], Ashutosh Darpel [a] and
Ritik Raina [a]

[a] *Dept. Of Information Technology Sanjivani College of Engineering, Kopargaon-423601 (MH), India*

Abstract. Nowadays Customer segmentation became very popular method for dividing company's customers for retaining customers and making profit out of them, in the following study customers of different of organizations are classified on the basis of their behavioral characteristics such as spending and income, by taking behavioral aspects into consideration makes these methods an efficient one as compares to others. For this classification a machine algorithm named as k-means clustering algorithm is used and based on the behavioral characteristic's customers are classified. Formed clusters help the company to target individual customer and advertise the content to them through marketing campaign and social media sites which they are really interested in.

Keywords. Machine learning, Customer segmentation, K-means algorithm

1. Introduction

Today many of the businesses are going online and, in this case, online marketing is becoming essential to hold customers, but during this, considering all customers as same and targeting all of them with similar marketing strategy is not very efficient way rather it's also annoys the customers by neglecting his or her individuality, so customer segmentation is becoming very popular and also became the efficient solution for this existing problem. Customer segmentation is defined as dividing company's customers on the basis of demographic (age, gender, marital status) and behavioral (types of products ordered, annual income) aspects. Since demographic characteristics does not emphasize on individuality of customer because same age groups may have different interests so behavioral aspects is a better approach for customer segmentation as its focus on individuality and we can do proper segmentation with the help of it.

2. Literature Survey
[1] A solution is proposed as distinguish the customers group into two groups named as premium and standard with the help of machine learning methods named as NEM, LiRM and LoRM [2].

---

[1] Prof. Nikhil Patankar, Sanjivani College of Engineering, Kopargaon, India.
Email: patankarnikhilit@sanjivani.org.in

Tushar Kansal, Suraj Bahuguna, Vishal Singh, Tanupriya Choudhury. "Customer Segmentation using K-means Clustering", International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS).2018, In this paper customer segmentation on Telecom customers is achieved by using information such as age, interest, etc. with the help of cluster analysis method.

3. Use Case Diagram
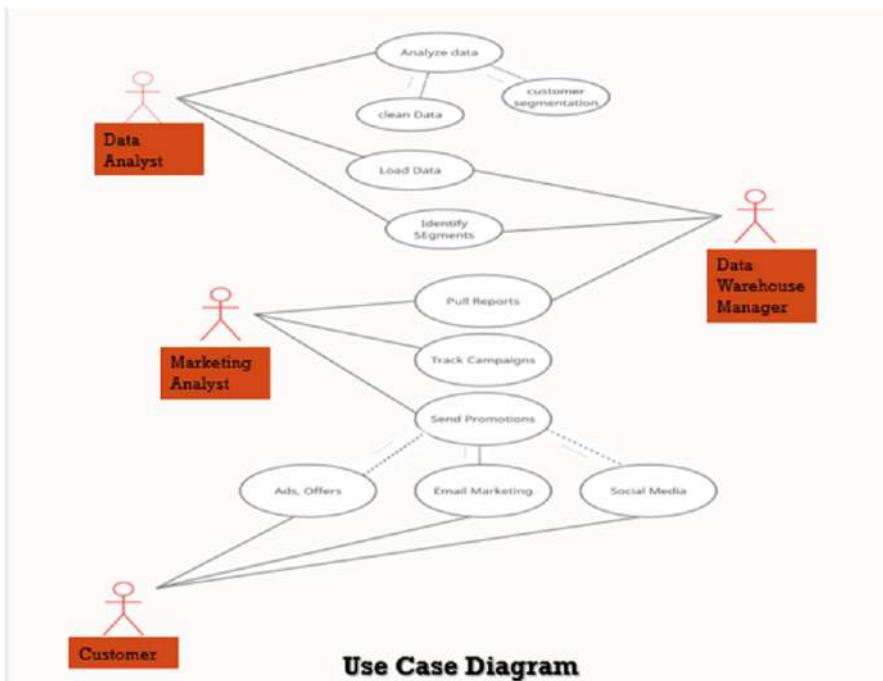
Figure 1. Use case Diagram

Use case diagram of proposed system consist of 4 users 1. Data Analyst 2. Marketing Analyst 3. Data Warehouse Manager 4. Customer in figure 1
And 6 use cases,
1. Analyze Data: analyst has the access to loaded data and analyst clean the data and perform analysis to form clusters.
2. Load Data: analyst log into database & view data & load into memory to work on it.
3. Identify Segments: analyst form report for segmented customer data and send to data warehouse and marketing analyst can access that data to form marketing strategies.
4. Pull Reports: marketing team can view & make edits on the reports, data for report is pulled from DW system.

5. Track Campaigns: The customer's interaction tracked by marketing team for success report.
6. Send Promotions: Marketing team send promotions through mail, social media ads, paid ads, coupons.

4. K-means Clustering Algorithm

K-means Clustering is a clustering Algorithm in which we are given with data points with its data set and features and the mechanism is to categories those data points into clusters as per their similarities.
The algorithm forms K clusters based on its similarity. To calculate the similarity K-means uses Euclidean distance measurement method.
Steps
   i.   In first step, we randomly initialize k points.
  ii.   K-means classifier categorizes each data point to its nearest mean and rewrite the mean's coordinates.
 iii.   Iteration is continuing up till all data points are classified.

5. Proposed System

In our system we including annual income and total spending as a feature for classification in figure2
1. **Data Gathering**: first, Data analyst fetch data required for analysis from database, format data i.e., remove all NA values from data & make data ready for processing.
2. **Feature Extraction**: Selects features which makes model more accurate, in our case features are annual income and spending score for efficient analysis.
3. **K-means Classifier**: After that, K means classifier performs clustering with respect to features provided to it,
4. **Hyper Parameter Tuning**: during forming groups to select optimal no of clusters we applied hyper parameter tuning which is achieved by Elbow method to choose optimal no of clusters.
    below graph is for elbow method which shows curve is getting flatter after 5 which indicates that 5 is optimal no of clusters we can form for better classification.
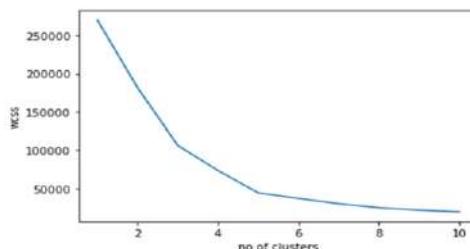
Figure 2. Elbow Method

5. **Data Visualization**: With the formed clusters marketing team can make different strategies for better targeting customers in figure 3.
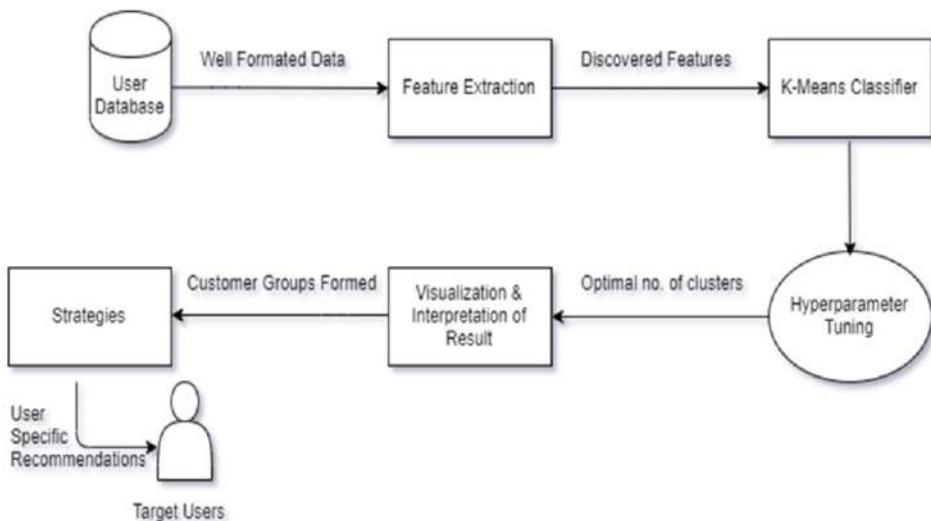
Figure 3. Flow of operation

6. Results

After analysis of data and classifying customers with features annual income and spending score, we got clusters of customers & with formed clusters marketing team form strategies for customers specific recommendation to make value out of them in figure 4.
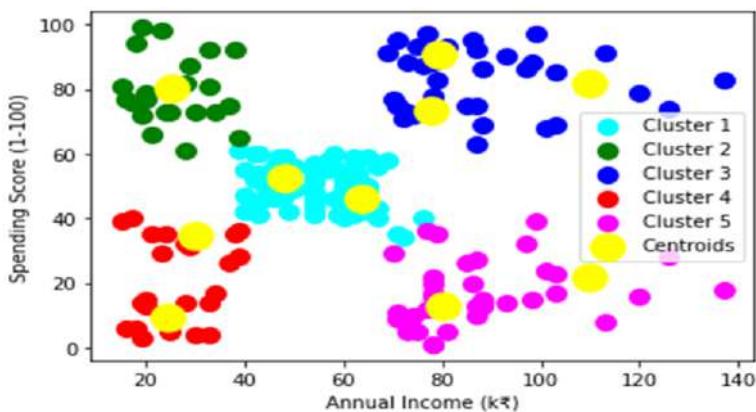
Figure 4. Final Cluster Formed

7. Drawback of System

1. Marketing will become expensive.

2. Because of having less no. of customers in a segment problem of limited production occurs.

8. Conclusions

Customer segmentation is performed on the company's customers data and with the help of K-means clustering machine learning algorithm customers are divided using features like total spending and annual income, this study also proves that the dividing customers on the basis of behavioral characteristics is a better solution for existing customer segmentation problem and K-means clustering algorithm is identified as a good choice for this approach.

References

[1] Sukru Ozan, "A Case Study on Customer Segmentation by using Machine Learning Methods", IEEE, Year: 2018.

[2] Jayant Tikmani, Sudhanshu Tiwari, Sujata Khedkar "Telecom customer segmentation based on cluster analysis An Approach to Customer Classification using k-means", IJIRCCE,Year: 2015.

[3] Chinedu Pascal Ezenkwu, Simeon Ozuomba, Constance kalu Electrical/Electronics and Computer Engineering Department, University of Uyo, Akwa Ibom State, Nigeria "Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services", IJARAI,Year: 2015.

[4] Potharaju, S. P., Sreedevi, M., Ande, V. K., & Tirandasu, R. K. (2019). Data mining approach for accelerating the classification accuracy of cardiotocography. Clinical Epidemiology and Global Health, 7(2), 160-164.

[5] Yogita Rani and Dr. Harish Rohil "A Study of Hierarchical Clustering Algorithm", IJICT, Year: 2013.

[6] Omar Kettani, Faycal Ramdani, Benaissa Tadili "An Agglomerative Clustering Method for Large Data Sets", IJCA, Year: 2014.

[7] Snekha, Chetna Sachdeva, Rajesh Birok "Real Time Object Tracking Using Different Mean Shift Techniques–a Review", IJSCE, Year: 2013.SulekhaGoyat"The basis of market segmentation: a critical review of literature", EJBM, Year: 2011.

[8] Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.

[9] Tanupriya Choudhury, Vivek Kumar, Darshika Nigam, Intelligent Classification and Clustering Of Lung and Oral Cancer through Decision Tree and Genetic Algorithm, International Journal of Advanced Research in Computer Science and Software Engineering,2015

[10] Tanupriya Choudhury, Vivek Kumar, Darshika Nigam, An Innovative and Automatic Lung and Oral Cancer Classification Using Soft Computing Techniques, International Journal of Computer Science and Mobile Computing,2015

[11] H. Mehta, V.S. Dixit and P. Bedi," Refinement of recommendations based on user preferences".

[12] H. Mehta, S.K. Bhatia, V.S. Dixit and P. Bedi," Collaborative personalized web recommender system using entropy-based similarity measure".

[13] Rivedi, A., Rai, P., DuVall, S. L., and Daume III, H. (2010, October). ´Exploiting tag and word correlations for improved webpage clustering in Proceedings of the 2nd international workshop on Search and mining user-generated contents (pp. 3-12). ACM.

[14] Potharaju, S. P., & Sreedevi, M. (2017). A Novel Clustering Based Candidate Feature Selection Framework Using Correlation Coefficient for Improving Classification Performance. Journal of Engineering Science & Technology Review, 10(6).

[15] Domavicius, G., and Tuzhilin, A. (2015). Context-aware recommender systems. In Recommender systems handbook (pp. 191-226). Springer US.

[16] K. Windler, U. Juttner, S. Michel, S. Maklan, and E. K. ¨Macdonald, "Identifying the right solution customers: A managerial methodology," Industrial Marketing Management, vol. 60, pp. 173 –186, 2017.

[17] R. Thakur and L. Workman, "Customer portfolio management (cpm) for improved customer relationship management (crm): Are your customers platinum, gold, silver, or bronze?" Journal of Business Research, vol. 69, no. 10, pp. 4095 – 4102, 2016.

A Session Key Based Security Mechanism for Cyber Physical System

Sandip Thite [a,1] and J. Naveenkumar [a]

[a] *Bharati Vidyapeeth University Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India*

Abstract. In recent years extensive research is going on for the development of applications which convert physical devices into smart devices. Industry 4.0 adopt the technologies under Cyber Physical Systems (CPS) for the development of such types of smart devices. Increase in the use of such type of smart devices without any security mechanism causes an open invitation for cyber attackers to perform cyber-attacks on such devices. Even current security algorithms are not efficiently work due to some constraints of smart devices. The goal of this research paper is to provide effective solution against different cyber-attacks on CPS applications. This paper proposed session key-based security mechanism which is used for the prevention of cyber-attacks and authentication of cyber devices.

Keywords. Security, Cyber Physical System, Replay attack, Internet of things, Session key, Cyber-attacks, man-in-the middle attack

1. Introduction

Cyber Physical system (CPS) is defined as a new generation of electronics system which works with integration of physical system and computational algorithm. It is broadly used in the development and deployment of smart devices.[1] A basic architecture of the CPS represent four different components, which include physical system, computation, communication and information system. Physical system is represented by the basic static system that work manually. A next component of CPS is Computation system. It is used to convert physical system into automated system. To convert into automated system it uses set of instructions into coded form. This code will be executed to perform automated operations of physical device. The next important component of CPS is communication. To establish communication in between devices it uses communication protocols under wireless (IEEE 802.11) and wired (IEEE 802.3) environment. Even now a days for short range devices Zigbee (IEEE 802.15.4) protocols also used. Communicating devices establish ad hoc network in between them. Information exchange for execution of task is also performed by communicating devices.

The use of cyber physical system in different application is increased rapidly but at the same time negligence of security of such type of devices also happened. Due to this it is open invitation for attacker to perform attacks on these devices [2][7]. From last two years if we going through different security reports, we found that cyber-attacks increase significantly on CPS [4].

---

[1] Sandip Thite,Bharati Vidyapeeth University Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune ,Email: sandipthite83@gmail.com.

The main reason behind that people taking more care of their devices like desktop computer or laptop by installing antiviruses, firewall or they implement current security mechanism in it.

2. Background

Basic Architecture of CPS represent four different components which include Physical system, computation system, communication system and Information system. These four components are used for development of CPS. At the same time these components having some loophole to perform cyber-attacks. These attacks have two parts Physical attack and Cyber-attack. Attacker founds different attack points in CPS which includes Fake devices, Weak protocol of communication, Fake access point, Spoof user interaction and manufacturer infrastructure, weak application programming interfaces etc.

One of the most important parameters for securing CPS infrastructure is device identity and mechanisms for authentication. But many CPS devices do not have the required computation power, memory or storage   to support and implement current security algorithm [8]. Today's strong encryption and schemes of authentication are based on cryptographic suites such as Advanced Encryption Suite (AES) for confidentiality, Rivest- Shamir-Adleman (RSA) for digital signatures and key exchange and Diffie-Hellman (DH) for key negotiations and management [3].   While these algorithms are robust. They require high computation resource that may not exist in all CPS based devices. Consequently, authentication and authorization will require systematic study and reengineering   to accommodate security needs of new CPS connected networks [9].

Secondly, existing authentication and authorization protocols also require a degree of user intervention in terms of configuration and provisioning [5]. However, many CPS devices will have limited access, thus requiring re-designing of the new techniques and protocols that can support tiny, low memory and low computational power CPS devices. Furthermore, preservation of privacy has been a concern since the dawn of the Internet. Identity management in the CPS is important characteristics required in the security framework [6]. The main aim of this research paper is to present a general and flexible security framework that provides robust security for CPS-based applications in diverse and user-centric environments. We also proposed Session key-based security mechanism which prevent attacker to perform cyber-attacks on the system.

3. Secure CPS Flexible Framework

Figure 1 shows secure CPS flexible framework. Bottom layer of the CPS framework is physical layer which describes hardware used for the development of security mechanism for cyber physical systems. Microcontroller interface includes Raspberry pi or Arduino. Both the interfaces are small in size but capable to handle heavy tasks. These interfaces acts as a Gateway server. While at the client side we can choose communication, interface depends on environment. Where we can use IEEE 802.11, 802.15 or 802.15.4. It is totally depends on wireless environment which we used. Whether it is short range or long range. Depends on physical distance between

Gateway controller and client node we can choose suitable hardware interface like ESP8266, ESP32, XBEES2C or BLE4.0.

Second layer is provides device drivers for hardware interfaces which is used for security mechanism. These device drivers are used to activate all hardware which we used to implement security mechanism in cyber physical system. It also work for power management of all the hardware. Handling of Input/Output devices also done in this layer.
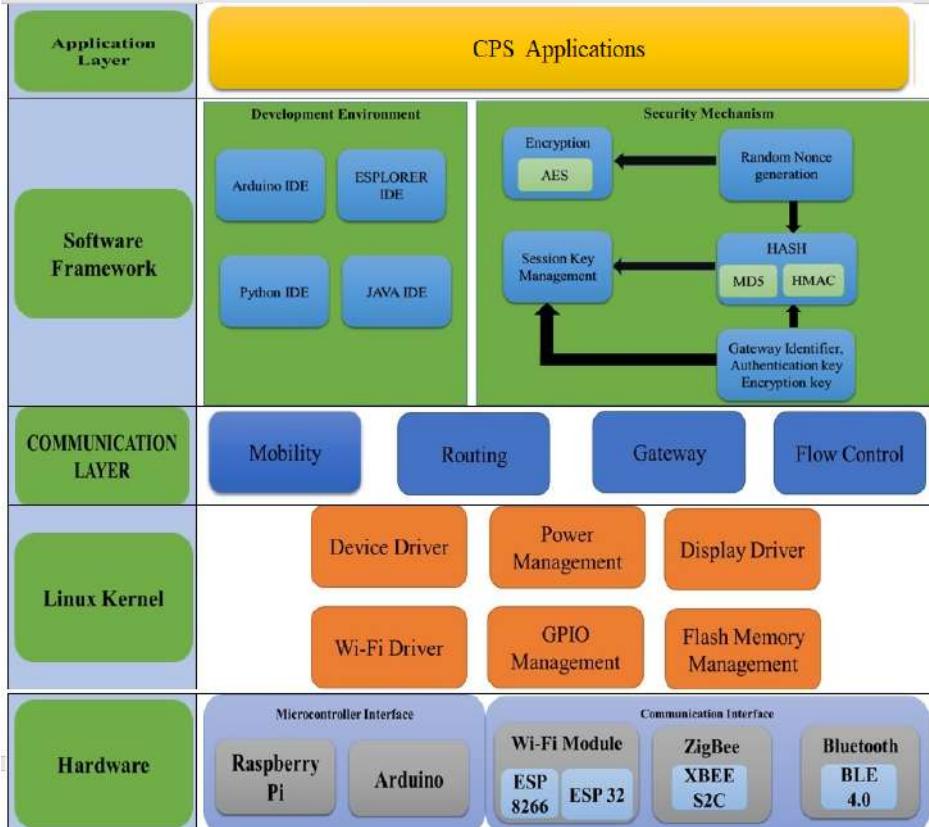
Figure 1. Secure CPS flexible framework.

Communication layer is a bridge in between Application user, application software and hardware. Which works for the mobility of the application. It provides connection establishment between gateway server and client node. It also used to exchange messages in between the nodes for authentication and authorization purpose. Software framework layer contain two modules. First module works for development environment for security mechanism. It is used for execution of code building procedure. While second module is actually used to code development for security mechanism.

The top most layer is Application layer, which includes applications that we used in cyber physical system. Security mechanism is specially developed in a way that where it support client server environment.

4. A Session key based Security Mechanism

The CPS consists of various small capacity devices which include microcontroller sensors etc. which works under wireless environment.

We proposed a new security mechanism for CPS. It is a session key-based mechanism. It is a lightweight mechanism developed for constraint-based devices which has a issues like low computational power, less energy, less memory etc. Where following mentioned parameters are used for successful execution of mechanism

MK = Microcontroller key

PK1 = Public key for authentication

PK2= Public Key for Encryption

S1, S2 =Random nonce for Session key generation

AT_D= Authentication token at smart device

ST_D= Session token at smart device

AT_M = Authentication token at Microcontroller

ST_M = Session Token at Microcontroller

SK_D = Session key at smart device

SK_M = Session key at Microcontroller
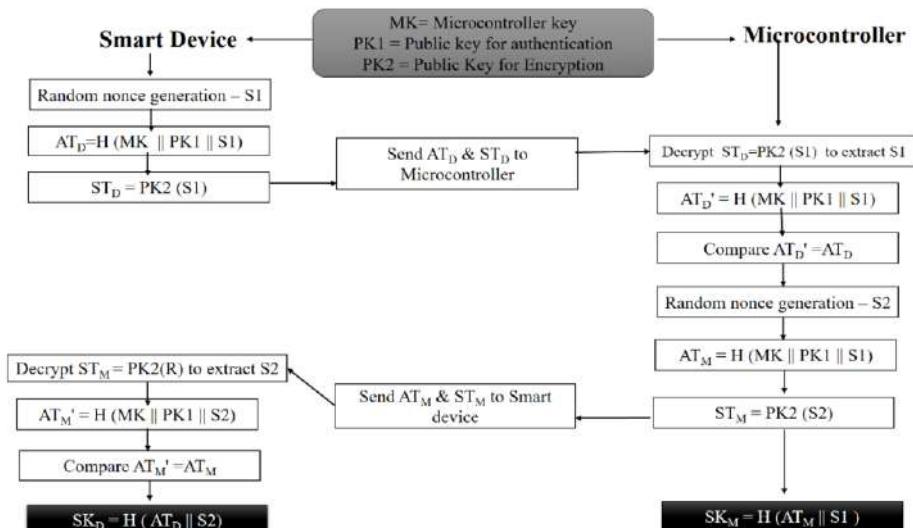
Figure 2. Session key based Security Mechanism

The figure 2 shows key generation algorithm in between smart device and microcontroller device which controls all smart devices. As shown in figure MK, PK1 and PK2 are common parameters used in between smart device and microcontroller for the implementation of security mechanism for generation of session key.

Stepwise execution of Security mechanism is divided into two parts where part 1 execute at smart device and part 2 execute at microcontroller for the generation of session key.

Step 1: S1 is a random number generated by using random nonce generation function at smart device.

Step 2: Authentication token (AT_D) is generated by implementing Hash function(H) at smart device by using three parameters MK, PK1 and S1.

Step 3: Session token (ST_D) is generated at smart device by using Public key for encryption and random nonce S1.

Step 4: Authentication token (AT_D) and session token (ST_D) sent to microcontroller from smart device through communication channel.

Step 5: At microcontroller random nonce S1 is extracted by decrypting Session token (ST_D) and by using common parameter PK2.

Step 6: Authentication token (AT_D') is regenerate again at microcontroller by implementing hash function on MK, PK1 and S1.

Step 7: Newly generated Authentication token (AT_D') compare with Authentication token (AT_D) sent by smart device. If both are same then execute step 8 else authentication not done with smart device so discard smart device.

Step 8: S2 is a random number generated by using random nonce generation function at Microcontroller.

Step 9: Authentication token (AT_M) is generated by implementing Hash function(H) at microcontroller by using three parameters MK, PK1 and S1.

Step 10: Session token (ST_M) is generated at microcontroller by using Public key (PK2) and random nonce S2.

Step 11: Authentication token (AT_M) and session token (ST_M) sent to Smart device from microcontroller through communication channel.

Step 12: At smart device random nonce S2 is extracted by decrypting Session token (ST_M) and by using common parameter PK1.

Step 13: Authentication token (AT_M') is regenerate again at smart device by implementing hash function on MK, PK1 and S2.

Step 14: Newly generated Authentication token (AT_M') compare with Authentication token (AT_M) sent by Microcontroller. If both are same then execute step 11 else authentication not done with microcontroller so communication will not be established with Microcontroller.

Step 15: Session Key (SK_D) is generated at smart device by implementing Hash function on Authentication token (AT_D) and S2.

Step 16: Session key (ST_M) is generated at microcontroller by implementing Hash function on Authentication token (AT_M) and S1.

Step 17: Session key (ST_D) at smart device and session key (SK_M) at microcontroller authenticate both the devices.

5. Experimental Setup

Experimental setup implemented on home automation system where for microcontroller we used Raspberry Pi Model B 4. Home devices like FAN, Bulb, TV connected with ON / OFF switch where we used Wi-Fi module ESP32 for connectivity with microcontroller. In between microcontroller and smart device, we used Access point (IEEE 802.11) to establish communication in between these devices.

6. Results

A new session key is generated at every communication session in between microcontroller and smart devices. New session key at every session causes difficulty for attacker to capture authentication information. To test the system, we performed Man in the Middle attack on system. To perform this attack we used Ettercap tool. Which is open source and free security tool to perform man-in-the-middle attacks on network. We also developed other unsecure infrastructure of home automation system. After execution of the system, it was found that unsecure system easily breakable under Man In the Middle attack while the secure system developed with Session key based security mechanism not breakable   under Man In the Middle attack by using Ettercap.

7. Conclusion

Increase in the use of Cyber physical system with less focus on security causes attacks on such system. So better security mechanism is basic requirement for the development of Cyber physical system. In this paper we proposed Secure Cyber Physical System Framework which shows that we can implement independent security mechanism on CPS. Hardware compatibility issues removed with secure CPS Framework. We also proposed a security mechanism which generate a new session key at every communication session causes difficulty for cyber attacker to get authentication information which will be used to perform attacks on the system.

References

[1]   Thite S., Thakore D. (2020) A Survey on the Internet of Things: Applications, Challenges and Opportunities with India Perspective. In: Kumar A., Paprzycki M., Gunjan V. (eds) ICDSMLA 2019. Lecture Notes in Electrical Engineering, vol 601. Springer, Singapore. https://doi.org/10.1007/978-981-15-1420-3_138

[2]   J. Granjal, E. Monteiro and J. Sá Silva, "Security for the Internet of Things: A Survey of Existing Protocols and Open Research Issues," in IEEE Communications Surveys & Tutorials, vol. 17, no. 3, pp. 1294-1312, thirdquarter 2015, doi: 10.1109/COMST.2015.2388550.

[3]   Q. Jawadwala and K. Patil, "Design of a novel lightweight key establishment mechanism for smart home systems," 2016 11th International Conference on Industrial and Information Systems (ICIIS), Roorkee, 2016, pp. 469-473.

[4]   Banerjee, A., Venkatasubramanian, K.K., Mukherjee, T., Gupta, S.K.S.: Ensuring Safety, Security, and Sustainability of Mission-Critical Cyber–Physical Systems. Proceedings of the IEEE 100(1), 283–299 (2012).

[5]   S. Thite, D. Thakore, "Key Establishment Algorithm for Secure Cyber Physical System to Prevent Cyber-attacks ", International Journal of Innovative Technology and Exploring Engineering Volume-9 Issue-2, December 2019.

[6]   Mehmet Hazar Cintuglu , Osama A. Mohammed , Kemal Akkaya , A. Selcuk Uluagac "A Survey on Smart Grid Cyber Physical System Testbeds" in IEEE Communications Surveys & Tutorials, Vol-19, Issue-1, 2017 DOI 10.1109/COMST.2016.2627399

[7]   Sandip Thite, J. Naveenkumar. (2021). FASSSTeR: A Novel Framework Aligned with ISA and ISO Standards for Cyber Physical System Safety, Security, Sustainability and Resiliency. Design Engineering, 2021(02), 399 – 411

[8]   Y. Chen, S. Kar, and J. M. Moura, "Dynamic attack detection in cyber physical systems with side initial state information," IEEE Transactions on Automatic Control, vol. 62, no. 9, pp. 4618–4624, 2017.

[9]   A. Y. Nur and M. E. Tozal, "Defending cyber-physical systems against dos attacks," in 2016 IEEE International Conference on Smart Computing (SMARTCOMP), May 2016, pp. 1–3

Real Time Human Gesture Recognition: Methods, Datasets and Strategies

Rameez Shamalik[1] and Dr.Sanjay Koli[2]

[1]*Research Scholar,G. H. Raisoni College of Engineering and Management, Pune,India*
[2]*Professor, Department of E&TC, Dr.D.Y. Patil School Of Engineering,*
Pune,India

Abstract. Gestures are universal means of communication without any language barrier. Detecting gestures and recognition of its meaning are key steps for researchers in computer vision. Majority of the work is done in sign language already. Sign language datasets are compared with respect to their usability and diversity in terms of various signs. This paper highlights the available datasets from three dimensional body scans to hand action gestures. Their usability and strategies used to achieve the desired results are also discussed. Major neural networks are evaluated in terms of varied parameters and feutures. A Methodology for effective gesture recognition in real is proposed. Lastly Results achieved through an Open CV in combination with Sci-kit learn library based technique for gesture recognition are presented and analyzed in terms of efficacy and efficiency.

Keywords. Gesture Recognition, Datasets, Open CV, Hand Gestures, Sci-Kit Learn

1. Introduction

Gestures facilitate a user to interact with given environment. Gestures not only allow front end interaction with others with no language barrier but also remote interactions with a smart screen, virtual-reality and augmented-reality objects. Digital computing is made possible thanks to gesture-based human-computer interfaces[1] from everyday life physical objects like lights, mirrors, doorknobs, notebooks to the specific and focused tasks like driving a car or a medical surgery. Gesture study helps to further understand semantic meaning with contextual information about a person's behaviour and reactions. Among a plethora of various gestures which include hand, body and head gestures, hand gestures stand class apart as they simply carry more information due to multiple combinations than the other types of gestures.

Gestures play a vital role to help people with speaking and hearing disabilities as it replaces voice as the primary means of communication[2]. Dynamic gesture recognition through video processing is tedious due to the complications in the gesture background, ambiguity of video devices and the non-uniformity in a collected data. Enhanced requirements for computing devices especially in the recent lockdowns increased the volumes and necessity of easy to use computer interfaces.

2.   Related Theory

There is a rapid transition in video processing techniques based on requirements of various application[3]. Three Dimensional(3D) videos are gaining increasing importance amongst various techniques[4]. Image based gesture recognition is the most basic techniques[5], also Haar cascade is repeatatively used for static hand and face recognition[6].Adaboost algorithm based fingure tracking and contours detection is also recently achieved[7]. Intelligent cars are being designed having a simple RGB camera in front connected to wifi for real time gesture control[8].

Computer vision plays an important role not only in gesture recognition but also in facial identification[9], Crop disease detection[10] and even in accident detection in various weather situations[11].

Datasets of particular project are the core of the research topic. Efficient and accurate data with respect to the application needs to be selected. When it comes to gestures, sign language datasets emerge first but a lot of work is already done upon them with greater accuracy. Some of the examples include American sign language dataset and Brazilian sign language dataset[12].

MPI-FAUST[13] dataset includes a complete three dimensional (3D) scans of hundreds of subjects highlighting their gestures**. DensePose**[14] is another dataset made by Facebook research team, it maps all human pixels of an RGB image to the 3D surface of the human body. **DensePose- COCO**[15] is an upgraded version of the prior, to work on 3D estimation of pose in wild. When it comes to Hand gestures, the most effective dataset is found to be **EgoGesture**[16] developed by Chinese Academy of sciences. It has both images and videos of 26 different signs made by hands with depth. **NVGesture**[17] is a very similar dataset for Dynamic Hand Gesture Dataset is a resource for researchers in the field of dynamic hand gesture recognition. **Jester**[18] is another dataset which is a collection of labelled video clips that show humans performing pre-defined hand gestures. **MSR 3D** dataset[19] is a sequence of 100 images captured from 8 cameras showing the breakdancing and ballet scene developed by Microsoft. **DHG** i.e. Dynamic Hand Gesture dataset[20] contains sequences of 14 handgestures performed in two ways: using one finger and the whole hand.

All of the above mentioned datasets are of size ranging from 1 GB to 30 GB in size as they contain image as well as video sequences to be worked upon. Also an extra depth feature is also involved to achieve greater accuracy. While working with such a huge amount of data, Annotations become difficult for each frame, hence pre-processing is limited to negligible. Deep Neural Networks are the valid solution to achieve better results in such scenario. Neural networks starting from simple Convolutional Neural Networks (**CNNs**) to **VGG-16, AlexNet, ResNet** etc. are used to process the data for higher accuracy. Basic comparison is highlighted in table 1.

Table 1. Comparison between Major neural networks with key features

Neural Network	Special feature	Parameters	Accuracy
VGG Net	Static kernels	138M	92.3%
AlexNet	Deep structure	60M	84.7%
ResNet	Short connections	62M	60.3%

3.   Methodology

First part in gesture recognition is to find the favorable pixels in the given video frame or image. Gestures especially hand ones has lot to do about edges. The pixel density is always high upon region of interest(ROI).The Canny algorithm performs brilliantly when it comes to a function that finds edges in the input image and marks them in the output map edges. Edge detection helps in finding both depth and geometry of a frame. It involves noise reduction and finding intensity gradient of the image.

The next step especially in hand gesture recognition is to find the hand region from all the edges available. It is important to eliminate unwanted regions in the video sequence. It is done with help of Background Subtraction which uses concept of running averages. 30 frames are looked upon over a given video sequence to calculate it. A difference image is thus obtained by calculation of absolute difference between the background model and the current frame to highlight foreground portion[21].

Motion Detection and Thresholding play a vital role in featuring the visibility of hand in question in order to get rid of rest of the portion which is colored as black and hand particularly becomes white. This exercise can also help understand the depth in the frames available as logically hand gestures should be in the foreground.

Contours help in understanding the size and the shape of the hands by making a boundary around the particular gesture made by hand. Bounding box over the contour helps in segmentation of the ROI from the frame.

4.   Results and Discussion
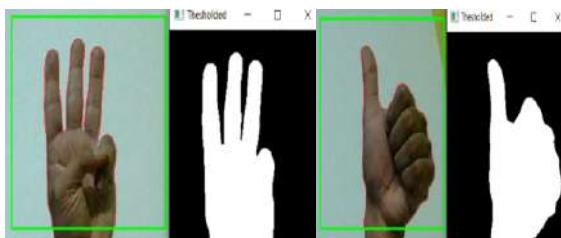
Figure 1. Various Hand gesture capture with threshold frame

Figure 1 shows three steps happening together which includes background subtraction happening in the bounding box shown with green borders. Bounding box helps with segmentation as whole video frame need not to be processed but just a specific area. Finding exact contours is the key to proper gesture detection as it forms the exact shape of the gesture. Proper thresholding is important for highlighting the gestures in a separate window where the background is shown as black and foreground i.e. a gesture as white in a specified bounding box only. Note that system collects 20 frames per second hence active only for first 20 seconds, later it will continue to display the captured contour shape.
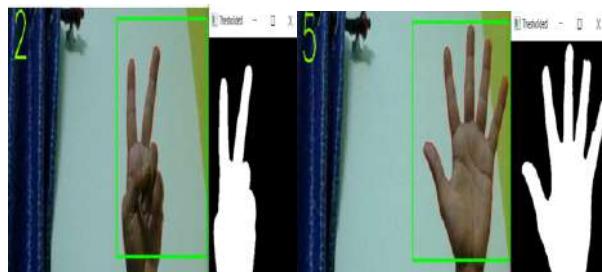
Figure 2. Detection of count of fingers

Capturing gesture is just not enough but understanding them adds value. The shapes are finalized with the help of contours and thresholding but recognization of the gestures require a separate approach. After successfully thresholding the hand gesture we could now count the number indicated by the fingers in the hand gesture. A convex hull is formed around the contour in the Region of interest (ROI). Basically it tracks the movement of hands to form the convex hull. Next step is to find the exact distance between the fingers for which Euclidean distance is a standard solution in computer vision. It is measured with the help of Sci-kit learn library in built function which performs bitwise AND operation to display the numbers as shown in figure 2. This model conveniently segments the background, no matter how complex it is.

5. Conclusion

There are various datasets and techniques available for evaluation of Gestures. Majority of work done is on the sign language detection. Only capturing the gesture is not enough but to gain understanding of the same is more important. Depth and geometry provide insight in the real time gesture detection thus adding extra variables to be worked upon. Accurate depth and geometry of the frame is obtained by using canny edge detection. In Python programming libraries like Open CV, Numpy and Sci-Kit Learn are used to recognize hand gestures and numbers associated with it. A specific technique needs to be opted for effective evaluation of gestures which will suit the applications of the design. The developed framework needs to be high efficiency, economical Model for Human Gesture Detection and can find its application in Human-Machine Interaction, Virtual Reality, Augmented Reality, Sport Videos, Airport Security and other public transportation hubs.

References

[1]   M. P. Pacaux-Lemoine et. Al, "Emulated haptic shared control for brain-computer interfaces improves human-robot cooperation", *IEEE International Conference on Human-Machine Systems (ICHMS)*, Rome, Italy, 2020.

[2]   R. Nair et. Al, "Hand Gesture Recognition system for physically challenged people using IoT", *6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2020.

[3]   Sanjay Koli, Rameez Shamalik, "Transformation of Video Signal Processing Techniques from 2D to 3D: A Survey", *Proceedings of the 2nd International Conference on Communications and Cyber Physical Engineering, springer lecture notes in electrical engineering(LNEE),* volume 570, pages 63-70, ICCCE 2019.

[4]   Rameez Shamalik, Sanjay Koli, "Emergence and Functionality of 3D Videos", *International Journal of Engineering and Advanced Technology (IJEAT)*, ISSN: 2249 – 8958, Volume-9 Issue-3, February, 2020

[5]   O. Saman and L. Stanciu, "Image Processing Algorithm for Appearance-Based Gestur Recognition", *23rd International Conference on System Theory, Control and Computing (ICSTCC)*, 2019, pp. 681-684, doi: 10.1109/ICSTCC.2019.8885888.

[6]   S. Sharma, S. Jain and Khushboo, "A Static Hand Gesture and Face Recognition System for Blind People", *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*, 2019, pp. 534-539, doi: 10.1109/SPIN.2019.8711706.

[7]   R. M. Gurav and P. K. Kadbe, "Real time finger tracking and contour detection for gesture recognition using OpenCV", *International Conference on Industrial Instrumentation and Control (ICIC)*, 2015, pp. 974-977, doi: 10.1109/IIC.2015.7150886.

[8]   G. Dai and P. Wang, "Design of intelligent car based on WiFi video capture and OpenCV gesture control", *2017 Chinese Automation Congress (CAC)*, 2017, pp. 4103-4107, doi: 10.1109/CAC.2017.8243499.

[9]   Manisha Talware, Sanjay Koli, "Video-Based Person Re-Identification: Methods, Datasets, and Deep Learning", *International Journal of Engineering and Advanced Technology (IJEAT)* ISSN: 2249 – 8958, Volume-9 Issue-3, February, 2020.

Using Artificial Intelligence in Source Code Summarization: A Review

Shraddha Birari [a,1], Sukhada Bhingarkar [a]

[a] *Computer Science and Engineering, MIT World Peace University*
Pune, India

Abstract. Source code summarization is the methodology of generating the description from the source code. The summary of the source code gives the brief idea of the functionality performed by the source code. Summary of the code is always necessary for software maintenance. Summaries are not only beneficial for software maintenance but also for code categorization and retrieval. Generation of summary in an automated fashion instead of manual intervention can save the time and efforts. Artificial Intelligence is a very popular branch in the field of computer science that demonstrates machine intelligence and covers a wide range of applications. This paper focuses on the use of Artificial Intelligence for source code summarization. Natural Language Processing (NLP) and Machine Learning (ML) are considered to be the subsets of Artificial Intelligence. Thus, this paper presents a critical review of various NLP and ML techniques implemented so far for generating summaries from the source code and points out research challenges in this field.

Keywords. Deep Learning, Neural Networks, Software maintenance, Source code, Source code summarization

1. Introduction

Summarization can be viewed as transformation of data into a concise yet meaningful representation which could be used further for analysis or storage. There are different use cases of the summarization like Document summarization [1][2], Newsletter summarization [3], Summarization of information over social media, Generating summaries over videos [4], Summaries of the content within email [5]and Summaries over the programming language code.

Source code summarization is a methodology of understanding the functionality performed by the source code and automatically generating the descriptions out of the source code. Proper documentation makes the software system maintainable. Correct, complete, and consistent documentation leads to successful maintenance of the software system [6]. The comment written with respect to source code should exactly specify the functionality performed by the code. The comment should include all the necessary information which is performed by the source code. Also, the comment

---

[1] Shraddha Birari, Computer Science and Engineering, MIT World Peace University, Pune, India
Email: bshraddha30@gmail.com.

written should follow the same format throughout the whole document.

In this paper we present a review of various source code summarization methodologies implemented so far using different natural language and machine learning techniques.

2. Need of Source Code Summarization

This section gives brief idea about the different use cases where source code summaries or comments written are useful.

2.1. Software Maintenance

The software document is an artifact which communicates the information about the software system to the people implicated in the production of that software. The people involved here are the customers, developers, project leaders as well as managers [7]. According to the survey conducted regarding the problems faced by developers in [8], 66 % developers faced challenges while understanding motive or purpose of the piece of code. Understanding code written by someone else is a serious problem rated by 56% developers. 17% developers find it difficult to understand their own code written just a while ago. This shows properly written documentation regarding the source code is important in the software maintenance.

2.2. Code Categorization

In software development finding relevant application in similar category is important to understand the functionality and useful for reusing the common functionality [9]. Programmers can save much time by finding the functionality in similar applications. For finding some application, one can search according to categories such as "Business", "Communication", "Audio/Video", "Games" etc [10]. The problem to find relevant functionality in similar category is due to the mismatch between descriptions written and implementation done. Thus, well written description written in natural language will be helpful in the code categorization.

2.3. Code Search

A code search is general activity while developing a source code. According to the survey conducted among developers regarding the code search [11], 33.5 % people search to refer the example code. 26 % people read or explore the code. 16% search for code localization like "*where class is instantiated*". 16% refer the code to determine the impact like "*understanding the dependencies*" or "*find side effect of proposed changes*".8.5% refer for metadata like "*who recently touched the code*".   Also, according to the same survey, on an average developer creates 12 search queries as per each working day. This show searching a code is a frequent and highly important activity. The performance of the code search is highly depending on the text involved in the search term and the code snippets [12]. Code search is difficult when search term

specified as input do not have the same words as the corresponding source code. Thus, well written comments will lead to effective code search.

3. Source Code Summarization Methodologies

As discussed earlier, source code summarization is task which generates the comments from the given code snippet. In this section, we present some existing techniques implemented for source code summarization which includes encoder-decoder model, language model, and reinforcement learning etc.

3.1. Deep reinforcement learning based code summarization.

Reinforcement learning is paradigm in Machine Learning which takes suitable action to maximize the reward for given situation. In [13], for generating relevant summary from the source code, deep reinforcement learning methodology is used. This framework is also known as Actor-Critic network which is used to generate the comments. The confidence to predict the next word is calculated by Actor network whereas Critic network evaluates the value calculated by the actor network. Figure. 1 shows the overview of this framework.
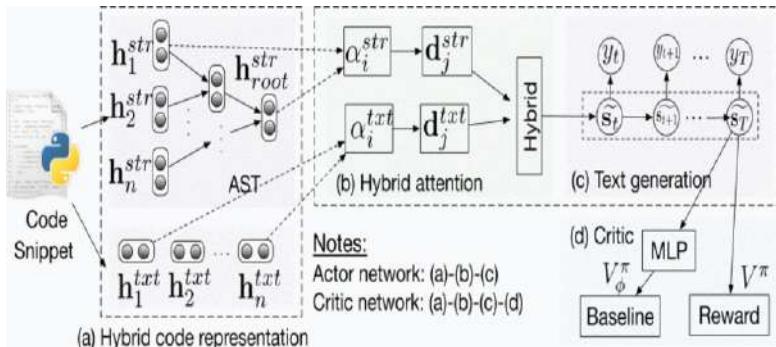
Figure 1. Overview of proposed code summarization framework [13].

This proposed architecture is based on encoder-decoder framework. Hybrid code representation is "Encoder" in this encoder-decoder framework.   In this work, sequential tokens as well as structural information is utilized for the code representation. The sequential tokens of the code snippet are denoted by Long Short-Term Network (LSTM) whereas structural information of the source code called as semantics is preserved by Abstract Syntax Tree (AST). AST also maintains the hierarchical representation of the source code. LSTM based on AST represents the structural contents of the code snippet. Thus, sequential as well as structural contents of the code snippet are represented in the "Hybrid code representation".

   Actor-critic network acts as a "Decoder" in this encoder-decoder framework. Actor network calculates next word's probability distribution with respect to the current state. Softmax function is used to predict the next word. Critic network approximates the

value generated by the actor network at each time step. In this work, the evaluation score i.e., Bilingual Evaluation Understudy (BLEU) is defined as reward.

The proposed work effectively captures the semantics of the code snippet due to use of AST. Also, actor-critic network predicts the summary, which resolves exposure bias issue. But this model is evaluated only on certain Python code and comment pairs, which may not represent all the types of comments and not generalized for other programming language. Also, only BLEU score is utilized in order to calculate the reward, which may not satisfy human evaluation criteria.

3.2. Statistical Machine Translation based code summarization.

Statistical Machine Translation (SMT) is a methodology for translation among two natural languages, which is adopted in [14]. SMT is basically Natural Language Processing technique (NLP) which identifies grammatical rules from the two languages such as English, Hindi and translates into corresponding language. Phrase Based Machine Translation (PBMT) is utilized which identifies phrase to phrase relationship among source and destination sentence. In PBMT modeling, set of phrase pairs introduced, which include source sentence and target subsequence. Consider an example for the statement: if x %5 == 0, Its phrase pair is: "if" → "if", "x" → "x", "% 5" → "by 5", "== 0:" → "is divisible". "Phrase table" is used to generate phrase pair, which contains various phrase-to-phrase relationships with probabilities. Here to maintain the grammatical structure of the natural language, reordering is performed.

Figure 2 shows the overview of PBMT technique for pseudo code generation. As shown in the figure, Phrase translation model calculates the probability Pr with respect to Target t and Source s. Reordering model calculates Probability Pr for arrangement of each phrase with respect to given Phrase pair ϕ and Alignment a. Language model finds the fluency of target sentence t.
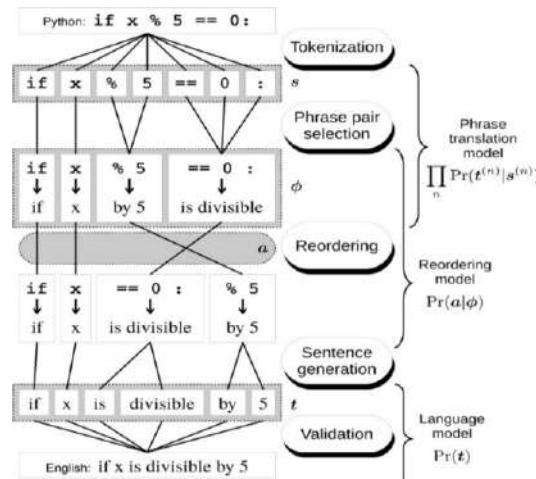
Figure 2. Python to English pseudo-code generation using PBMT technique [14]

PBMT is not able handle hierarchical correspondence of the source code and thus Tree to String Machine Translation (T2SMT) is useful. It makes use of parse tree of the input sentence. In this technique instead of phrase pair, "derivation" is introduced, where it represents relationship between source subtree and target phrase.

The above proposed approach is based on rule-based machine translation; thus, it can be generalized for variety of languages by applying corresponding rule. But this methodology does sentence wise translation due to which it is unable to handle multiple statements.

3.3. Code summarization based on Natural Language Generation

Natural language generation is a subfield of AI which translates given data into natural language such as English. In [15], source code summarization technique is utilized which generates description in English for given Java code. This approach is works by analyzing how method is invoked. In this approach, PageRank algorithm is used to find most important method in the most important context. Then Software Word Usage Model (SWUM) identifies the keywords from the action performed the important methods.

Finally, the custom Natural Language Generation (NLG) technique to generate the English descriptions which describes what methods actually do. In this NLG technique, first 6 different types of messages are created which represents different contexts of the methods. Table 1 shows different types of messages and its corresponding explanation.

Table 1. Types of messages

Message Type	Explanation
Quick Summary Message	This is short sentence which gives description of the function.
Return Message	Return type of the method is given by this type of message.
Importance Message	It shows the importance of the method based on PageRank.
Output Used Message	This is to describe maximum 2 methods which calls this method.
Call Message	This describes maximum 2 methods which is called by this method.

Next step is lexicalization, in which according to above message type phrases are used to describe it. After lexialization, more readable phrases are generated in the aggregation. Finally, senteneces are generated from the phrases generated from above step.

The above proposed approach makes use of context specific information, due to which it provides meaningful situation based or contextual output.

3.4. Tree Convolutional Neural Network (Tree CNN) based source code summarization.

Convolutional layer applied to neural network helps to extract important features from the input. In [16], Tree CNN is utilized in which program's structure is captured using

Abstract Syntax Tree (AST). AST generally captures syntactic structure of the language using hierarchical representation. Each component in the program is represented as AST's node. AST's node is denoted as a vector based on coding criterion. Then convolutional layer detects the structural features of the program. The new tree generated after convolution, has same size and shape as of original one. To solve this issue, dynamic pooling is utilized. This proposed approach basically classifies program according to corresponding functionality.

Above proposed work captures the meaning of the code snippet due to its hierarchical representation by using AST. But it exposed to training data which may cause it to suffer from the exposure bias.

3.5. Source code Markup Language (SrcML) based code summarization.

Source code Markup Language (SrcML) converts the source code of the various programming languages like Java, C, and C++ to XML file [17]. In the proposed work [18], input is the XML file for the given code snippet and output is the document file. This target document file is a combination of various parts in which every part gives description of the important part of the code snippet. SrcML considers numerous factors like white spaces, classes, parameters, and conditions and accordingly XML files is generated. The main components of the source code are represented with the help of the tags in XML like <class>, <function>, <loop> etc. Feature extractor fetches the data generated from XML file. Using XPath, queries are performed to identify each object from the code snippet that extracts four features: attributes, conditions, calls, functions. The variables or parameters included in the source code are identified as attributes. The tag<decl> is used to fetch these attributes. Conditions include "if" conditions as well as several types of loops. Number of calls performed by the source code are grouped into calls and are presented with the tag <call>. The functions with the name and the data type of the value returned by the code snippet are represented with the tag <function>. The feature extractor generates a program structure information file that can be used by code description generator. Code descriptor generator reads two files one is source code, and another is program structure information file then generates the comments based on the source code and related information.

The above proposed model makes use of tags which focuses on core components of the code, due to which complete and clear comments can be generated. Although it focuses on main concepts of code but does not show the inheritance relationships among classes. Also, proposed model only considers Object oriented aspects of the programming language and thus its useful only in case of object-oriented language.

4. Conclusion

Source code summarization technique generates the descriptions from the source code, which describes what source code intend to perform. Summary of the source code is useful for software maintenance, code search as well as code categorization.

Most of the existing source code summarization techniques were unable to capture

the structure of the source code and facing the exposure bias issue. Also, it is necessary to build the source code summarization technique that will have human evaluation criteria as well. In future, Generative Adversarial Networks (GAN) in which generator and discriminator can be combinedly designed to generate the summary which can be helpful to deal with the exposure bias issue.

References

[1]   Moratanch N, Chitrakala S. A survey on extractive text summarization. International Conference on Computer, Communication and Signal Processing (ICCCSP).2017.
[2]   Zhuang H, Zhang W. Generating Semantically Similar and Human-Readable Summaries With Generative Adversarial Networks.2019. IEEE Access 7:169426–169433.
[3]   Alwis V . Intelligent E-news Summarization. 18th International Conference on Advances in ICT for Emerging Regions (ICTer).2018.
[4]   Kini M. M, Pai K. A Survey on Video Summarization Techniques. Innovations in Power and Advanced Computing Technologies (i-PACT).2019.
[5]   Ayodele T, Shoniregun CA, Akmayeva GA. Security review of email summarization systems. World Congress on Internet Security (WorldCIS-2011).2011.
[6]   Kajko-Mattsson M. A Survey of Documentation Practice within Corrective Maintenance. Empirical Software Engineering. 2005; 10:31–55.
[7]   Forward A, Lethbridge TC. The relevance of software documentation, tools and technologies. Proceedings of the 2002 ACM symposium on Document engineering (DocEng '02).2002.
[8]   LaToza TD, Venolia G, DeLine R. Maintaining mental models. Proceeding of the 28th international conference on Software engineering (ICSE '06).2006.
[9]   McMillan C, Grechanik M, Poshyvanyk D. Detecting similar software applications. 34th International Conference on Software Engineering (ICSE).2012.
[10] Anh Tuan Nguyen, & Nguyen, T. N. Automatic Categorization with Deep Neural Network for Open-Source Java Projects. IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C).2017.
[11] Sadowski, C., Stolee, K. T., & Elbaum, S.How developers search for code: A case study. Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering.2015.
[12] Nie, L., Jiang, H., Ren, Z., Sun, Z., & Li, X. Query Expansion Based on Crowd Knowledge for Code Search. IEEE Transactions on Services Computing. 2016; 9(5): 771-783.
[13] Wan, Y., Zhao, Z., Yang, M., Xu, G., Ying, H., Wu, J., & Yu, P. S. Improving automatic source code summarization via deep reinforcement learning. Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering.2018.
[14] Oda, Y., Fudaba, H., Neubig, G., Hata, H., Sakti, S., Toda, T., & Nakamura, S. Learning to generate pseudo-code from source code using statistical machine translation. 2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE). 2015.
[15] McBurney, P. W., & McMillan, C. Automatic source code summarization of context for java methods. IEEE Transactions on Software Engineering. 2016. 42(2): 103-119.
[16] Mou, L., Li, G., Zhang, L., Wang, T., & Jin, Z. Convolutional Neural Networks over Tree Structures for Programming Language Processing. Proceedings of the AAAI Conference on Artificial Intelligence. 2016; 30.
[17] Mohsin, A. H., & Hammad, M. A code summarization approach for object oriented programs. International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT).2019.
[18] Collard, M. L., Decker, M. J., & Maletic, J. I. SrcML: An infrastructure forthe exploration, analysis, and manipulation of source Code: A Tool demonstration. IEEE International Conference on Software Maintenance.2013.

VirtualEye: Android Application for the Visually Impaired

Jay Bagrecha [a, 1], Tanay Shah [a], Karan Shah [a], Tanvi Gandhi [a], and Sushila Palwe [a]

[a] *School of Computer Engineering & Technology, MIT World University, Pune, Maharashtra, India*

Abstract. In India, almost 18 million visually impaired people have difficulties in managing their day-to-day activities. Hence, there is a need to develop an application that can assist them every time and give vocal instructions in both English and Hindi. In this paper, we introduced a robust lightweight Android application that facilitates visually impaired individuals by providing a variety of essential features such as object and distance detection, Indian currency note detection, and optical character recognition that can enhance their quality of life. This application aims to have a user-friendly GUI well suited to the needs of the blind user and modules like Object Recognition with Image Captioning so that the visually challenged user can gain a better understanding of their surroundings.

Keywords. Mobile Application; Visually Impaired; Object Detection; Image Captioning; Indian Currency Recognition; Optical Character Recognition; Deep learning; text-to-speech.

1. Introduction

Globally there are around 280 million people, who are visually impaired, of whom about 40 million people suffer from complete blindness. Considering India's vast population, it alone has 12 million people who live from blindness. With the rapid advancement in technology, many mobile devices include applications and features to help visually impaired people but most of the devices are designed for people with vision. One of the most challenging tasks for a visually impaired person is to identify the day-to-day things around them. With the availability of mobile devices and the rising computational capabilities, these people can be assisted using artificial intelligence and computer vision techniques. Imagine being a visually impaired person, able to locate and track the simplest of things like chairs, bookshelves, cupboards, and all sorts of daily items with ease. Today's technology has made life better for all. Artificial Intelligence has become entrenched in our everyday routines. As a result,there was a need to build an application that could assist visually disabled people with day-to-day activities using AI.

---

[1] Jay Bagrecha, School of Computer Engineering & Technology, MIT World University
Email: jaybagrecha1234@gmail.com.

2.   Literature Review

2.1. Object detection with Image captioning:

According to Faizad Amin in [1], If we want to understand the geometry of a scene, it is important to identify the obstacle present in our way as well as estimate the depth. There are two types of Object Detection: they can be static or dynamic. It is easy to locate static objects as they are fixed but it is a bit difficult for dynamic objects as they are constantly moving. In this approach, stereo images are used for object recognition. We calculate the unevenness of the detected objects and then calculate the distance of these objects from the camera which we call "depth". In Real-Time Object Detection Application by Selman Tosun [2], the visually weakened people will be able to recognize the obstacles while they are walking on the road using the feedback which they will get in form of audio and this will help them prevent possible accidents. The operations are performed using the inbuilt audio and the camera modules. This application has different modes for both indoor and outdoor transportation, voice feedback is a plus. According to Xiaofei Fu [3] in his Mobile Application for visually impaired people, they have made it possible to create most of the functionalities of the app offline. The functionalities include face detection, gender classification. They have made a system where the contents of the picture are given as output through voice. In the Intelligent Eye application by M Awad [11], there are features like the detection of objects, banknote, light, as well as color. All these features work completely fine even when the device is offline. The accuracy is also very good in this application and here the focus is on some different features like light and color detection, which are useful from the point of view of blind people. As we saw in Xiaofei Fu's Mobile Application [3], there is a provision of getting output through voice, so Quan Zheng You's [4] approach gives a perfect explanation of the current scene, instead of just naming the objects. Here an automatic generated natural language description of an image is given as output. When an image is fed to the CNN, then first the extraction of the top-down visual features are done and at the same time the visual concepts like the attributes, regions are detected, and proper structuring of all these words are done and the whole sentence is given as output. According to O Vinyals [12] for Image Captioning, A RNN is used for the generation of sentences that follows CNN for encoding the image into a proper representation, and then the Probabilistic Neural network approach is used, and the softmax algorithm is used for word prediction, which is very light.

2.2. Indian Currency Recognition:

The Recognition of the different denominations of Indian banknotes has been tackled using several methods and a considerable amount of research has been done in the field of currency note detection. Over the years the researchers presented their work based on characteristics like color, texture, etc., and have used many Deep Neural Networks and Machine Learning algorithms like CNN, ANN, RCNN, Masked RCNN, PCA, Naive Bayes classifiers, Random Forest, etc. for their research. According to [5] paper, a simple Convolution Neural Network architecture-based method was used to train the model and the implementation of currency recognition was executed in both web and android applications. Some of the frameworks used were TensorFlow, TensorFlowLite for android, and at last, enhanced their model by performing hyper-parameter tuning. New Dataset was created and then Data Augmentation was applied

to get 11657 images. A detailed tabular comparison based on Training and Testing accuracies, Computation Time, and overall performance of the models was done by the authors with many popular pre-trained models like VGG19, Xception, Resnet50, Alexnet, InceptionV3, and with a simple CNN model proposed in [6]. The results were quite fascinating and the proposed model outperformed all other models by giving the best training and testing accuracy of 100% and 87.5%. According to [6] paper, a DL model was used to detect the different denominations of Indian Banknotes. A Pre-trained model MobileNet was used that is a transfer learning method available in Keras Applications. The creation of a new dataset for four different denominations of the Indian currency and performing data augmentation on the dataset was done to get 12160 images. Results of their classification framework were good enough with a Training and Testing accuracy of 100% and 96.6%. The authors of this paper say that their approach requires very little data preprocessing and will perform great even if the input images have some disturbances or if the images are unclear. According to [7] paper, an android application was developed by the authors especially for blind people so that they can easily know the denominations of the Indian Banknotes. The authors of this paper have implemented a basic Deep learning model, which scans an image from your smartphone's camera and then gives an output based on some probabilities in the form of voice so that a blind person can hear it easily. The authors have gathered a dataset of all the valid Indian currencies and have performed data labeling using a labeling tool to get 2536 images. They have used another transfer learning method - Faster RCNN with Resnet v2 to get 87% accuracy and loss of 0.201.

2.3. *Optical Character Recognition:*

According to [8] paper, Tesseract is a perfect engine for OCR. HP has already developed page layout analysis technology that itself was used in the products. And that's the reason why there was no need for Tesseract to have its page layout analysis. After the page layout analysis is done, we use the line-finding algorithm. The line-finding algorithm is used to save loss of image quality without eschewing the images. The major parts of the process include filtering blobs and construction of lines. In the first step, components are outlined and stored, this is called connected component analysis. It is computationally expensive but easier to handle white text with a black color background. Now, outlines are gathered by the process of nesting them into Blobs. Then the Blobs are simplified into text lines. Text lines are broken into words. The next stage is Recognition, which is a two-step process. In the first pass, the algorithm tries to recognize each word which is then passed to an adaptive classifier as input. Then in the second pass, unrecognized words of the page are tried to recognize again. The unique thing about Tesseract is that it handles white-on-black texts in a better way. Paper [9] talks about all the methods by which OCR is done and its major challenges. According to the [9] paper, the main phases of optical character recognition include preprocessing. Over here, systems utilize binary or grey images as processing color images is computationally expensive. After this, we do Segmentation. In segmentation, the system separates the text part from the input image. There are three kinds of algorithms for document segmentation: Top-down method, Bottom-up method, Hybrid method. This gives us about 98% accuracy. The next step is Normalization where the characters which were separated are reduced in size depending on the algorithm used. The image is converted in the form of m*n matrix. After this, feature extraction is performed which can be time-consuming and complex

too. OCR systems use a lot of methodologies of pattern recognition. In this, an example is assigned to each predefined class. Techniques of OCR classification are: Statistical Techniques (The main statistical methods that are performed in the area of OCR [19] are Likelihood or Bayes classifier, Nearest Neighbor (NN), Clustering Analysis, Fuzzy Set Reasoning, and Quadratic classifier, Hidden Markov Modelling (HMM)), Neural Networks, Template Matching (least complex method) and Support Vector Machine (SVM) algorithms, and Combination of the classifier. The final step is post-processing. The main objective of OCR is to decide the context of the image. OCR systems make use of a dictionary to make minor changes in the errors that the system produces.

Paper [10] discusses Cloud Vision API and CNN. The first step is to do a layout analysis to locate the text on the image. The next step is to perform a text recognition analysis to produce the text. The step is performed through a convolutional neural network. Convolutional neural networks are a subset of neural networks. CNN follows the complex structure of the Human's visual cortex that is present in the brain through which we identify objects around us. The accuracy we achieve is about 80 percent.

3.   Research Gaps

On reviewing [1], [2], [3], [4] papers we figured out a few extremely important things, which are not present very widely, and we would be planning to implement and emphasize more on it while creating our project. First thing is that along with object detection it is very important to also, find the depth and the distance between the object and the user because we are trying to create an application that can be used in daily commute, so according to the depth of the object, priorities are set, for example, if in the scene if a dog and a bike both are present, but the dog is at 100 cm and the bike is at 1 m, so object detection will only mention the name of both objects, but we want that user should get first output as the dog is present at 50 cm and then about the bike. So to set priorities and to know the rough distance between the objects is very important. After reviewing [11],[12] papers we figured out that only giving the output of the name of the object and also the distance isn't sufficient, if the user is moving on the road, he would need to know about the whole scene present in front of him so that he confidently moves forward. In addition, explaining the whole scene can be done using Image Captioning, because if a person is driving a bike object detection will only give output as Bike and Person present at 200 cm, but with Image Captioning, it will give output as a person is driving a bike at 200 cm. We would also be making a system where the user has the option to get all outputs in Hindi as well because it should not happen that due to the language barrier, the user is unable to use our application. After reviewing [5], [6], [7] papers we established that the main problem of all the above-mentioned approaches is that they employ conventional pre-trained models, such as VGG16/19, AlexNet, Resnet v2, MobileNet, etc., which require a large number of annotated data. The datasets used in the papers do not have all the denominations of currency accepted in India. No previously mentioned papers incorporate the identification mechanism for counterfeit currency notes. Therefore, there is a need for another model, which is capable of extracting more deep features so that visually weak people can depend less on normal people and lead a better life. After reviewing [8], [9], [10] we established that the entire process of OCR is complex and more prone to errors if we are doing it from scratch. The software used in [9] expects us to give a processed image. So overall, there are various approaches to perform OCR. Each has its pros and

cons. Maintaining accuracy along with fast response is a challenge. There is a need to develop a system that does OCR along with Text to Speech Conversion in whichever language the user wants.

4.   Proposed Methodology

4.1. *Objectives:*

❖   To assist a visually impaired person about surroundings in real-time.

❖   To integrate Depth detection and Image Captioning with Object Detection so that they can understand what exactly is happening around them.

❖   To detect both static and dynamic objects present in the surrounding, and to get a proper output of those objects and the scene using NLP.

❖   To describe the content of an image using properly formed English sentences using Image Captioning.

❖   To convert formed sentences to the regional language.

❖   To allow the visually impaired to autonomously deal with Indian banknotes, particularly while accepting their money back during their day-to-day activities.

❖   To help visually impaired people know what is written on a piece of paper, or in an image, or anywhere around them, whenever needed.

4.2. *Modules:*

This application helps visually impaired people to visualize and navigate through their surroundings. All they have to do is to launch the app using Google's Talkback feature, and then onwards, the app will take care of the rest. The android application will be built using Flutter Framework and Google Talkback utilization. User Interface will be created in Dart Language. UI will consist of Home Page, which cordially invites blind users to start the app through audio. The proposed app uses deep learning for object detection and provides the name of the object and its position relative to the user (either left side or right side) as an audio output. It promptly detects Indian Banknotes for the impaired user and also uses Optical Character Recognition to scan text and provide the price and expiry date of the product. Flutter framework was used to develop this application; thus, this app will run on Android as well as on iOS platforms. TensorFlow Lite and MobileNet provide the necessary dependencies and will do the required object detection.
A use case diagram shown in Figure 1, explains all the scenarios when a user interacts with the application.

Module 1: Object Detection with Image captioning and distance determination

In the Object Detection module, the user will just have to point the mobile's camera in the surrounding, and the app will start detecting objects in real-time. When an image or video is fed, the object recognition model can detect any set of objects in the surroundings and gives information about the position of the objects in the image. Object detection models are trained in such a way that they can analyze the location of multiple classes of objects. When we provide an image to the next model, it produces a list of identified items, the location of the bounding box containing each item, and a score indicating the correct confidence.

The model returns an array of four numbers representing the bounding rectangle that surrounds its location for each object found. Objects with maximum accuracy will be labeled and their speech output is given.
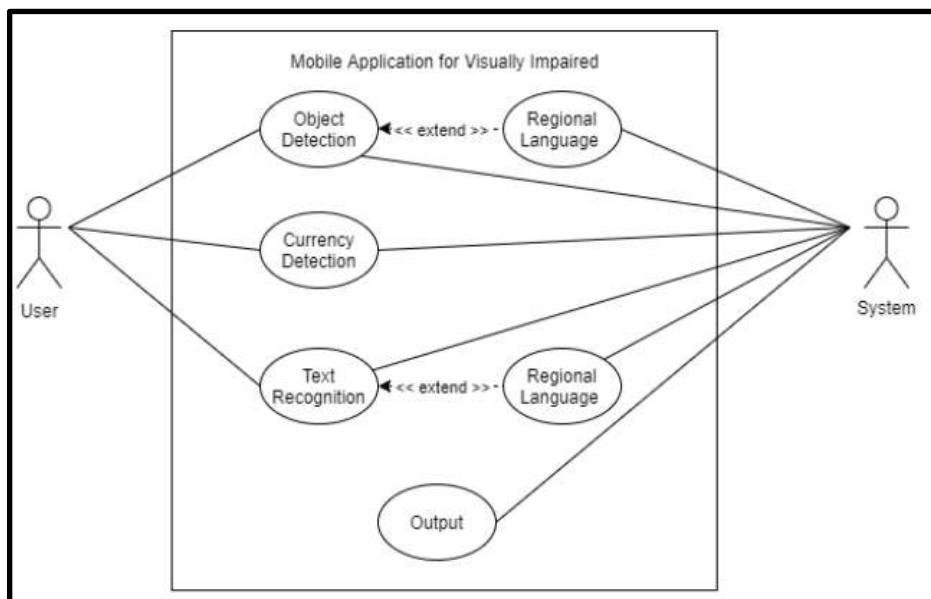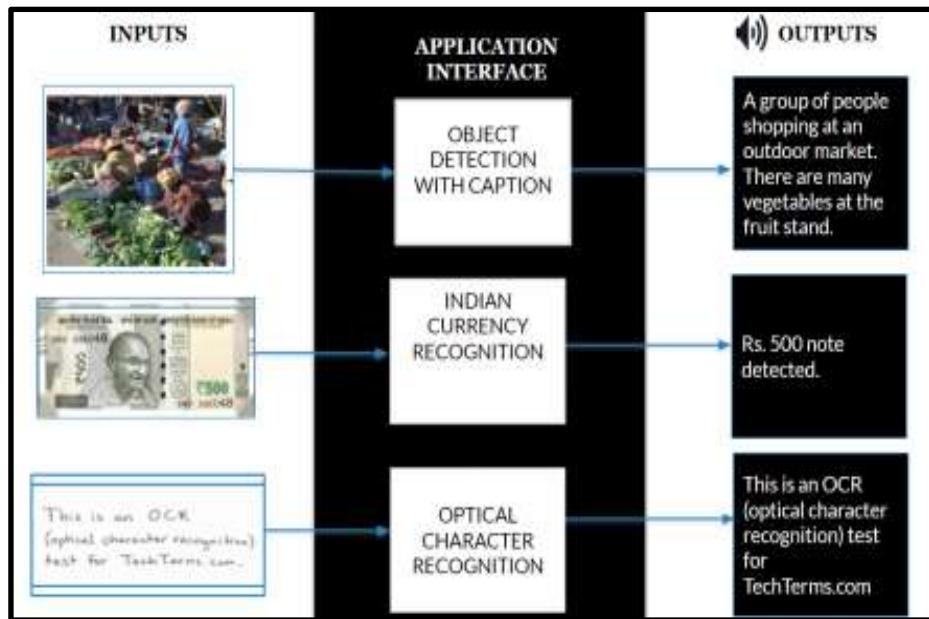
Figure 1. Use case diagram

Module 2: Indian Currency Recognition

In the Indian Banknotes recognition module, the training of a lightweight CNN model is required or any Pretrained model available in Keras Applications which not only accurately recognizes all the denomination of Indian banknotes in real-time with good accuracy but also determines that whether the banknotes given to a visually impaired individual is real or fake. This module will also concentrate on voice commands to guide the user through each move, reducing their reliance on others, particularly during outdoor activities.

Module 3: Object Character Recognition

In the OCR, the user just has to point the mobile's camera at a price tag of a product. After that, the app will tell the user about the price, expiry date, etc. of that product. OCR module is used to identify the text inside an image and separate it into actual text. Our app needs this to recognize the content on the price tags, bills, flex boards and to extract important information from various other sources. To do this we have to divide the photo into two halves. This saves extra computations. Figure 2 the lower half of a tag is where the price is written most of the time. Further, in the lower half, we then

locate the bold or the bigger size text.

Figure 2. Workflow Diagram (I/O Block Diagram)

5.   Requirements

5.1. *Hardware Requirements:*

Android Smartphone with a well-functioning camera with at least 2 GB of RAM.

5.2. *Software Requirements:*

❖ TensorFlow- It is an open-source library used to implement different deep learning algorithms. With the help of TensorFlow API, a lot of mathematical and numerical computations can be done with ease. It was developed by Google such that any model can be trained and implemented easily.

❖ OpenCV- It's an open-source library for machine learning and computer vision applications. It was built in such a way that the use of machine perception in commercial products and provision is present for common infrastructure for computer vision applications.

❖ Google Cloud Vision - It is an application programming interface that enables the developer to use powerful pre-trained machine learning models including image labeling, face and landmark detection, Optical Character Recognition (OCR) through REST APIs. Using Google Cloud vision API, we can easily integrate CV with different technologies.

❖ Android Studio- It is an IDE for the android operating systems, built and designed specifically for Android development. It has a proper system based on Gradle.

❖ Python- It is an open-source general-purpose interpreted, high-level programming language used to develop web applications, software applications, games, data-science applications, and many other things.

❖ Flutter- It is an open-source UI software development kit developed by Google to create various applications for Windows, Android, iOS, etc from a single codebase.

❖ Google TalkBack - Google TalkBack is a service that gives vocal instructions to its users and allows them to access Android applications with ease, and can interact better with the device and whole Android ecosystem.

❖ Google Colab- Colab is a Jupyter notebook-based environment that runs purely on the cloud. It doesn't require any proper setup and different people can simultaneously work on Colab notebooks that you create the same way as we work on Google docs.

6.   Conclusion

Visual impairment is one of the most debilitating disorders a man can have. It affects an individual's overall well-being as well as their emotional and social relationships. In this paper, an Android application is introduced for visually impaired individuals that assist them in visualizing and navigating their surroundings, thereby reducing their reliance on others, especially during outings. The application includes modules like object detection with image captioning and object distance from the user, Indian banknote detection that also alerts the user if the note is fake, and optical character recognition (OCR) that helps the user know what is written on a piece of paper, or in an image, or anywhere around them, whenever needed. Many different functionalities, such as barcode scanning, card readers, light identification, color recognition, road assistance, voice-based SMS and Email, Emotion Recognition, and location sharing, can be built in the future to benefit visually impaired people.

References

[1]   Amin, F., Mehdi, S. A., & Khan, A. D. (2019). Distance Estimation using Tensorflow Object Detection. 2019 International Conference on Electrical, Electronics and Computer Engineering (UPCON). doi:10.1109/upcon47278.2019.8980216

[2]   TOSUN, S., & KARAARSLAN, E. (2018). Real-time object detection application for visually impaired people: Third eye. 2018 International Conference on Artificial Intelligence and Data Processing (IDAP). doi:10.1109/idap.2018.8620773

[3]   Fu, X. (n.d.). Mobile assistant app for visually impaired people, with face detection, gender classification and sound representation of image. Electrical Engineering Department, Stanford University.

[4]   You, Q., Jin, H., Wang, Z., Fang, C., & Luo, J. (2016). Image captioning with semantic attention. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2016.503

[5]   Veeramsetty, V., Singal, G., & Badal, T. (2020). Coinnet: Platform independent application to recognize Indian currency notes using deep learning techniques. Multimedia Tools and Applications, 79(31-32), 22569-22594. doi:10.1007/s11042-020-09031-0

[6]   Mittal, S., & Mittal, S. (2018). Indian banknote recognition using convolutional neural network. 2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU). doi:10.1109/iot-siu.2018.8519888

[7]   Bhavsar, K., Jani, K., & Vanzara, R. (2020). Indian currency recognition from live video using deep learning. Communications in Computer and Information Science, 70-81. doi:10.1007/978-981-15-6648-6_6

[8]   Smith, R. (2007). An overview of the tesseract ocr engine. Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) Vol 2. doi:10.1109/icdar.2007.4376991

[9]   Hamad, K., & Kaya, M. (2016). A detailed analysis of optical character recognition technology. International Journal of Applied Mathematics, Electronics and Computers, 4(Special Issue-1), 244-244. doi:10.18100/ijamec.270374

[10] Neves, António & Lopes, Daniel. (2016). A practical study about the Google Vision API

[11] Awad, M., Haddad, J. E., Khneisser, E., Mahmoud, T., Yaacoub, E., & Malli, M. (2018). Intelligent eye: A mobile application for assisting blind people. *2018 IEEE Middle East and North Africa Communications Conference.*doi:10.1109/menacomm.2018.8371005

[12] Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/cvpr.2015.7298935

A Novel Technique for Handling Small File Problem of HDFS: Hash Based Archive File (HBAF)

Vijay Shankar Sharma[a, 1,] and N.C Barwar[b]

[a] *Research Scholar, Dept. of CSE, M.B.M Engineering College, Jodhpur, India*
[b] *Professor & Head, Dept. of CSE, M.B.M Engineering College, Jodhpur, India*

Abstract. Now a day's, Data is exponentially increasing with the advancement in the data science. Each and every digital footprint is generating enormous amount of data, which is further used for processing various tasks to generate important information for different end user applications. To handle such enormous amount of data, there are number of technologies available, Hadoop/HDFS is one of the big data handling technology. HDFS can easily handle the large files but when there is the case to deal with massive number of small files, the performance of the HDFS degrades. In this paper we have proposed a novel technique Hash Based Archive File (HBAF) that can solve the small file problem of the HDFS. The proposed technique is capable to read the final index files partly, that will reduce the memory load on the Name Node and offer the file appending capability after creation of the archiv.

Keywords. HDFS, Small File Problem, Meta Data Management, Hash Function, HAR, Map File, SSHF, HT-MMPHF, Merging Technique.

1. Introduction

Hadoop is open-source technology to handle the vast amount of unstructured and big data, which offers the wide range functionality in comparison to the traditional relational data bases. The file system of Hadoop is known as the Hadoop Distributed File System (HDFS) that is based on the master slave architecture. In this architecture there is a Name Node that acts as a master with processing capabilities and stores the meta-data information of the files stored in the file system. There are number of Data Node's that act as the slave means these Data Node's are only used to store the data, no processing is required at the Data Node's. Once a file is stored on the HDFS it is divided in the 128 MB size blocks and then these blocks are stored on the HDFS. The size of the HDFS block is variable means client can configure the size of the HDFS block as per the requirement, by default it is 128 MB. To ensure the availability of the data, HDFS replicate the data blocks on the Data Node's and it will be decided by the replication factor that is by default 3, means each data block is written on the three Data Node's, in case if any one of the Data_Node's gets down then data block can be

---

[1] Vijay Shankar Sharma, Research Scholar, Dept. of CSE, M.B.M Engineering College, India
Email:vijay.mbmit09@gmail.com.

recovered from the other Data Node's that is available as a replica. Handling of large files in HDFS is done efficiently as it is designed according to the application of the large files. There are number of major platforms which generate the small files i.e., Facebook, Twitter, Instagram, LinkedIn, Amazon, Flip cart, snap deal etc. This list of platforms is very long therefore it is easy to understand small files generation platforms by the application areas i.e., social networking sites, e-commerce websites, educational websites, research and analysis websites, weather forecast websites, entertainment websites, log files generated by the severs, health care data etc. a file is termed as the small file, if it is less than the size of the default HDFS block size. The application area of the small files is very vast therefore the importance of small files in analytics and in technology is very crucial and important. Unfortunately, majority of the distributed file systems are not designed to deal with the problem of massive small files. Massive small files generate the large amount of the meta-data at the central node in the distributed system that will degrade the overall performance of the distributed file system. Unfortunately, HDFS is also not capable to deal with the massive number of small files; Name Node in the HDFS will be overloaded due to excessive meta-data generation while dealing with the massive number of the small files.

In this paper we have proposed a novel technique to handle small file problem of the HDFS called as "Hash Based Index File (HBAF Archive)". The major contribution of our technique is that one can directly access the small files meta-data without use of any caching mechanism. Now, there is no need to read the index file entirely in the memory, only required part of the index file will be read and loaded to the memory. To read the index file partly we have used the special-order preserving hash function: Hollow Trie Monotone Minimal Perfect Hash Function (HT-MMPHF) [1] [2] with index file. This function identifies the location of the searched file meta-data in the index file and calculates the limit (how much index file is to be read) and offset of the index file. With help of limit and offset our technique seeks the index file and loads the required meta-data to the memory. To access index file randomly may be an expensive operation in case of the large index files therefore to limit the size of index files another special hash function: Scalable-Spittable Hash Function (SSHF) [3] [4] [5] is used that will dynamically distribute the meta-data of the massive number of small files to the various index file in place of the single index file. The remaining section of the paper is as follows; Section 2 presents the literature review on the existing techniques to deal with small file problem. Section 3 presents the proposed technique in detail with explanation of the HBAF creation algorithm and appending files after creation of the archive. Section 4 presents details of experimental setup and analysis of the result. At last Section 5 briefs the conclusion and future work.

2. Related Work: A Brief Survey

Jude Tchaye-Kondi et al. [6] proposed a archive file system, known as the Hadoop Perfect File. To access and distribute the meta-data of a particular file, special hash functions with order preserving capacity are used. Jian-feng Peng et al. [7] proposed a new variant to the HDFS with caching and merging module. The working of these modules are interrelated, to utilize the memory space efficiently the co-related files are merged and a special cache is designed for the fast access of the frequently accessed data. To solve the problem of the small files Xun Cai et al. [8] proposed the optimized

merging algorithm that is based on the correlation and distribution of the files. Hwajung Kim et al. [9] proposed a digital archive the will significantly reduce the storage I/O operation by modifying the inode structure of existing file system at both primary and secondary memory level. Yanfeng Lyu et al. [10] presented an optimized approach that will reduce the name node memory usage and access time, while handling massive number of small files. To improve the efficiency of read/write operation for the small files Xiong Fu et al. [11] proposed block replica placement algorithm. This algorithm also suitable for the cloud environment. Qi Mu et al. [12] proposed improved storage architecture for the massive small files. This architecture is based on the use of the secondary indexes. Tao Wang et al. [13] proposed a technique called as "Modified PLSA" that handles the massive small files by establishing the association among the application, access file and access tasks. The balancing of the data blocks can also be a measure of handling massive small files, Hui He et al. [14] proposed a unique algorithm that will consider the even utilization of the data blocks while merging the small files. Songling Fu et al. [15] proposed a technique that will reduce the memory required for the meta-data management while dealing with millions of small files. The proposed technique is called as the "iFlatLFS", which is based on the concept of the flat storage architecture. Yingchi Mao et al. [16] proposed SIFM technique that will use the multi level indexing for handling million of small files efficiently. Bo. Dong et al. [17] categorize the small files logically and structurally. On the basis of this division prefetching and merging of small files is applied to the structurally oriented small files and prefetching and file groping concept is used for the logically oriented small files. Ahad M. A et al. [18] proposed a dynamic merging technique. This technique identifies the small files by their size and type and Two-Fish cryptographic technique is used to secure the data in the file system.

3. Proposed Architecture

It is obvious that accessing of small files in HDFS is a complex and time-consuming task; therefore, to achieve fast meta-data access for small files we have proposed a Hash Based Archive File (HBAF) method. "Write-Once, Read-Many" is the prime property of the Hadoop Distributed File System, to keep this property in mind we designed our proposed technique in such a way that we will be able to append new files after creation of the archive. Our proposed technique will provide better processing and accessing performance in comparison to the Hadoop Archive (HAR). As the Figure.1 depicts that our HBAF Archive consist several slave indices files that will be generated from the temporary master index file using SSHF. Apart from the several index files HBAF also consists part file (file created after merging small files) and master name file which consists name of all the small files that be appended to the part file. Index files are responsible to store the meta-data of the small files; the selection of the particular index file will be done by the special hash function. In our technique two-level hashing will be used, at level-1 particular index file is identified by the SSHF and at level-2 HT-MMPHF order preserving hash function is used to locate the particular files meta-data location in the index file. The proposed approach HBAF will improve the performance in two ways, one is by the concept of merging, all the small files are merged therefore memory utilization will be improved and overall performance of Name Node will be better, due to the reduced memory load. Another way is the use of the two-level hash functions to build the index files for small files meta-data that will

provide fast access to the small files. While merging the small files, parallel multiple part files are created, this parallelism will merge the small files comparatively faster than the HAR. The level of parallelism can be increased or decreased, by default it set as 'two'. The whole process can be summarized by Algorithm 1.

Algorithm 1: HBAF Creation & Updation

Step-1: **Initial Variable Declaration and their Initialization**
1.1 small files // a set of small files;
1.2 slave index file // temporary index files generated by SSHF;
1.3-part file; // creation of initial part file
1.4 temporary master index file; // creation of the initial temp master index file
1.5 master name file; // creation of the master's name file
1.6 meta-data; // creation of string variable for storing meta-data of small file
1.7 small file name; // creation of string variable for storing name of small file
1.8 final index files; // creation of initial empty final index file
Step-2: Process of Merging Small Files and Building Client-Side slave_index_files
2.1 start of loop-1; // for each small file from small files
2.2 merge each small file to the part file;
2.3 copy the small file meta-data to the meta-data variable;
2.4 copy the name of small file to the small file name variable;
2.5 append the meta-data of the small file to the temporary master index file;
2.6 append the name of the small file to the master name file;
2.7 provide unique id to the slave index file and final index file by using SSHF;
2.8 append the value of meta-data to the unique slave index file created in previous step;
2.9 check the threshold limit of the slave_index_file, if limit reaches its maximum, then create another unique slave index -file and final index file using SSHF and then continue with the append operation of meta-data
2.10 end of loop-1;
Step-3: Sorting slave_index_files and building final_index_files
3.1 start of loop-2 // for each slave_index_files with unique id
3.2 sort the slave index file's meta-data
3.3 implement the HT-MMPHF for all the all-slave index files
3.4 associate HT-MMPHF to the respective final index files according to their mirror slave index files
3.5 copy all the meta-data entries from slave index files to the corresponding final index files along with order preserving mechanism of HT-MMPHF;
3.6 end of loop-2;

Initially temporary master index file and master name file are created, temporary master index file is used for the purpose of backup once the slave index files are created finally this temporary master index file will be deleted, master name file is a file which reside permanently with the HBAF archive and hold the names of all the small files to process. Before appending to the part files, small files can be compressed at client side and can take advantage of fast processing at client side in comparison to the HDFS. A threshold limit on the capacity of part files is fixed and checked regularly while appending/adding the files to the part file. Once the threshold limit reached to its

maximum, the new part file will be created and rests of the small files are appended to the newly created part file. There is also requirement for imposing the limit on the size of index file because when each time a random seek operations is performed a new connection is established to read a file from various data node blocks therefore it is desirable that index file should be less than the size of the HDFS block. One of the important concerns in our approach is the dynamic distribution of the small files metadata to the various slave index files, to implement this dynamic distribution we are using SSHF, later on these slave index files will be converted to the final index files. The process of building final index files is accomplished in two phases, the first phase of building final index files starts along with the merging process of small files, when a small file is added/appended to the part file, simultaneously its meta-data and name of the file will be added to the temporary master index file and master name file respectively after that with the help of SSHF its meta-data will be added to the corresponding slave index file.
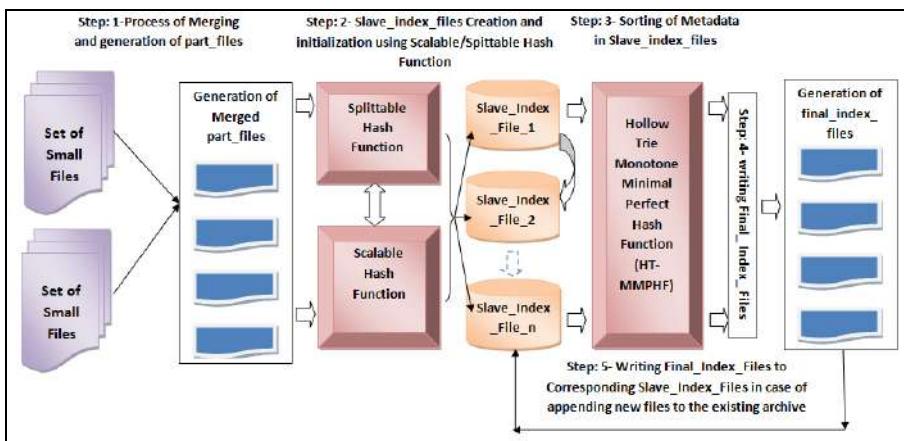
Figure 1. Proposed Architecture of HBAF Archive

SSHF belongs to the class of extendible hashing Zhang D. et al. [5] that uses dynamic hashing technique to allocate meta-data of the small files to the slave index files. In this technique hash is considered as the bit string and uses an ordered tree data structure Tarjan R.E. et al. [19] for lookup purpose. Figure 1 the decision of choosing particular slave index file for the entry of file's meta-data will be done by the hash function that will determined by the last two bits of the bit string of a file name hash value. The entries that have same pattern at last bits will belong to the same slave index file. While addition and deletion operations are performed on slave index files they grow and shrink dynamically with help of Scalable-Spittable Hashing. If a slave index file reaches its threshold limit it spilt and generate a new slave index file. The spilt hash operation is responsible for the dynamic creation of the slave index files at the same time slave index files can be directly accessed during look-up. Creation of slave index files and corresponding final index files is a parallel process, both files are created simultaneously. Re-arrangement of the meta-data entries is done in newly created slave index file during the spilt hash operation as it is highly essential for the synchronization of old and newly spitted slave index file.

HT-MMPHF is a perfect hash function, in this function a set of 'p' key that are static type are mapped with the 'q' numbers of integer type without any collision and there should be the value of integer number is always greater than or equal to the value of static key (q>=P). When value of 'q' and 'p' is equal, the hash function satisfies the 'minimal' property and hash function is called as the minimal perfect hash function. To preserve the order of keys we are supposed to use order preserving minimal perfect hash function, this function return the integer values strictly in the order of the static key, hence using this function the lexicographic order of meta-data entries in final index files are kept in order. Meta-data entries in slave index file are sorted in lexicographic order on the basis of file name hash values, these hash values act as the keys of the hash function and finally the minimal perfect hash function is created and added in the beginning of the final index files. At last, after writing the entire slave index files to the corresponding final index files, temporary master index file will be deleted. The main advantage of this minimal function is that its time and space complexity (logarithmic) is much lesser in comparison to the other comparative hash functions. As the meta-data entries in final_index_files are sorted and can be accessed directly therefore access time of particular record is minimum (Big O (1)) [2].

4. Experimental Setup & Result Analysis

To test the proposed HBAF Archive and other competitive archives, a cluster of 5 nodes is being setup. The configuration of the Name Node and Data Node are same, that is Intel® core™ i5-7500 CPU@3.40GHz, 64-bit Operating System with 4 GB Installed RAM. Ubuntu 18.04.1 LTS is used as the operating system with open JDK-11.0.4 in the system in the cluster. The latest version of Hadoop (3.1.3) is used in all the machines over the 1 GBPS (Backbone) / 100 MBPS Network. The replication factor and block size of the HDFS are set to its default values that is 3 and 128 MB respectively. For the purpose of the testing, we created five data sets with different number of files i.e., 10000, 20000, 30000, 60000, 120000. The size of files in theses data sets will ranges from 1 KB to 1 MB. To evaluate the performance of the archives with proposed technique we have analyzed the few parameters while creation of archives and few parameters after creation of the archives i.e., Time To Create Archive (Milli-Seconds), Meta Data Usage (Bytes) are the parameters that will be analyzed while creating the archives and Time Required to Randomly Accessing 10, 50 and 100 Files from different archives with caching (Milli-Seconds), Time required to randomly accessing 10, 50 and 100 files from different archives without caching (Milli-Seconds) are the parameters that will be analyzed after creating the archives. The concept of caching is all about the using client's memory while accessing the files from the archives therefore resultant access time can be reduced. HAR Archive and Map File Archive supports the caching means corresponding index files are loaded into the client's memory when accessed first time after that corresponding files meta-data will be perfetched on the basis of LRU Bok K. et al. [20] Dong Bo. Et al. [21]. This caching mechanism put the burden on the client's memory and will be problematic where memory is limited at the client side therefore in our proposed design HBAF Archive; we implemented the concept of the centralized caching of HDFS [22] and use the memory of the Data Node's for managing the caching operation

4.1. Time to Create Archive (Milli-Seconds)

Map File Archive will take the minimum time for creation and HAR Archive will take the highest time for creation. Our proposed approach HBAF Archive lies in between Map File Archive and HAR Archive. Experimental result shows that proposed HBAF Archive is 28% to 34% (result varies for different data sets) faster than the HAR Archive. The Archive creation time of Map File Archive is 56% to 58% faster than the Hadoop Archive creation time and 24% to 40% faster than proposed HBAF Archive. Figure.2 depicts that although Map File Archive is taking minimum time for creation as it is based on the sequential file approach but this cannot be a performance measuring parameter of the archive, we have considered this parameter only to show that our proposed approach 'HBAF Archive' is taking moderate time for creation of the archive that is acceptable.
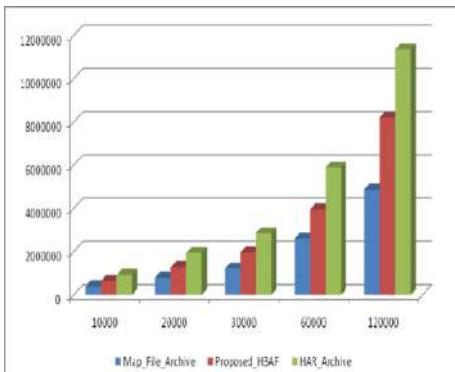
Figure 2. Time to Create Archive (Milli-Seconds)   **Figure 3**. Meta-data Usage (Bytes)
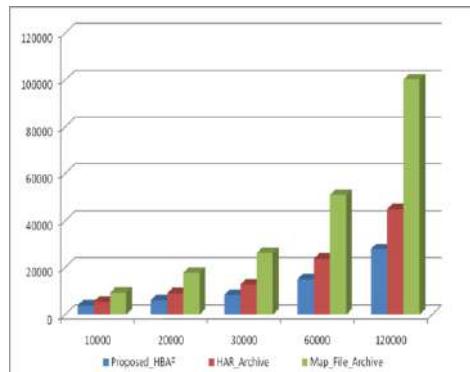
4.2. Meta Data Usage (Bytes)

Meta Data Usage of the proposed HBAF Archive is minimum in comparison with the HAR Archive and Map File Archive. There is one important point to note that when we increase the size of our datasets (i.e., 10000, 20000, 30000, 60000, 120000), we are continuously getting the better results means there is need of lesser space to manage the meta-data for larger datasets. This phenomenon proves our theory that proposed HBAF Archive will be able to handle millions of small files while consuming the minimum space for storing the meta-data of small files. Experimental result shows that in terms of Meta Data Usage proposed HBAF Archive perform 28% to 38% (result varies for different data sets) better than the HAR Archive and 58% to 72% better than the Map File Archive. These results can be clearly visualized in the Figure.3, which will show the strength of our proposed approach.

4.3. Time Required to Randomly Accessing 10, 50 and 100 Files from different Archives with Caching (Milli-Second)

When randomly accessing 10 files, the access time of proposed HBAF Archive is 23% to 73% (result varies for different data sets) faster than the HAR Archive. Experimental result shows that Map File Archive will take the highest access time, HBAF Archive is 14 to 18 time faster than the Map File Archive. The minimum access time will be taken

by the native HDFS but it can clearly see that proposed HBAF Archive is very close to the native HDFS. Native HDFS will perform 8% to 21% faster than HBAF Archive. As the Figure.4 depicts that HBAF Archive will continuously improve the access time with the increase of the number of files in the data set. To ensure the correctness and preciseness of the results we also accessed 50 and 100 files from the HBAF Archive and found the approximately same pattern results. In case of accessing 50 files, HBAF Archive will be 4 to 34% faster than the HAR Archive. In this case it has been noted that for lower data sets there is much lesser difference in the performance of the HAR Archive and HBAF Archive. Experimental result shows that Map File Archive will take the highest access time; HBAF Archive is 18 to 20 times faster than the Map File Archive. As the Figure.5 depicts that Native HDFS will perform 10% to 21% faster than HBAF Archive and HBAF Archive results improve for the larger data sets. If we calculate the average performance of the native HDFS in comparison with our proposed HBAF then it is overall 15% faster. In future the work can be carried out to improve our proposed technique in this direction. In case of accessing 100 files from the archives, HBAF Archive is 1% to 18% faster than the HAR Archive.
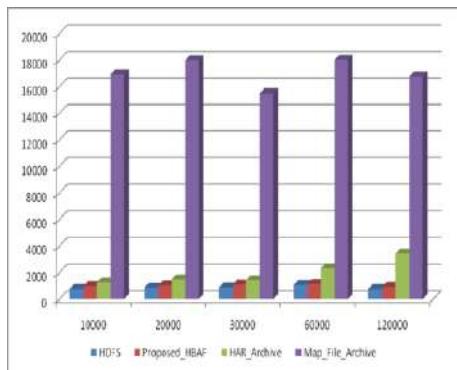
Figure 4. Access Time for 10 Files with Caching
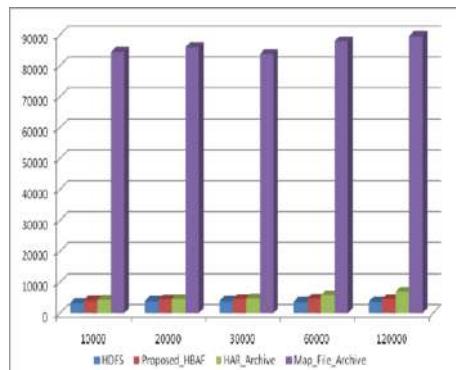
Figure 5. Access Time for 50 Files with Caching

Experimental result shows that Map File Archive will take the maximum access time and HBAF Archive is 18 times to 22 times faster than the Map File Archive. Figure.6 depicts that there is no impact on the performance of the Map File Archive by varying the number of files in the data sets. Native HDFS will perform 12% to 22% faster than HBAF Archive, in terms of average results Native HDFS is 15 to 16% faster than our HBAF Archive.

4.4. Time Required to Randomly Accessing 10, 50 and 100 Files from different Archives without Caching (Milli-Second)

As the Figure.7 depicts that the minimum access time will be taken by the native HDFS and there is very minor and negligible difference between the performance of the HBAF_Archive and Native HDFS. Native HDFS will perform 5% to 18% faster than HBAF Archive. The pattern of performance improvement of HBAF Archive is same as with caching enabled, HBAF Archive will perform better for the larger data sets. To analyze the pattern, correctness and preciseness of the results, we also accessed 50 and 100 random files from the archives and found the approximately same pattern results while caching is disabled. Experimental result shows that in case of accessing 50 files,
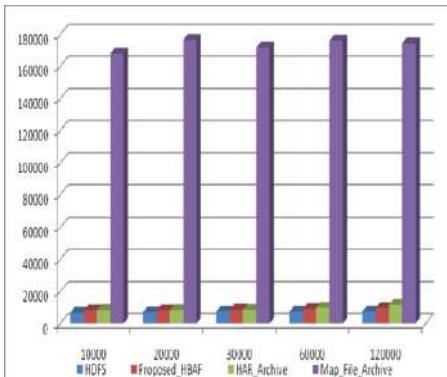
Figure 6. Access Time for 100 Files with Caching
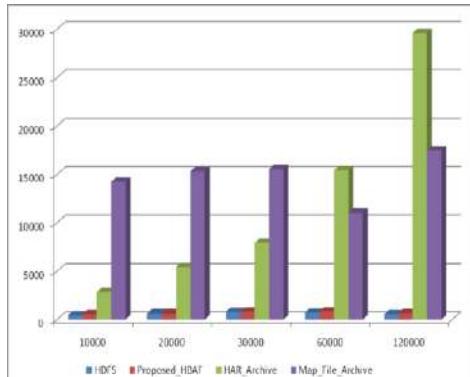
Figure 7. Access Time for 10 Files without Caching

HBAF_Archive will be 5 to 45% faster than the HAR_Archive that will be same as the accessing 10 files, means the impact of increasing number of files for accessing is negligible or is very less. As the Figure. 8 depicts that Map_File_Archive will take the highest access time; HBAF_Archive is 26 to 28 times faster than the Map_File_Archive, this will prove that when cache is disabled our HBAF_Archive will perform better than cache enabled environment.
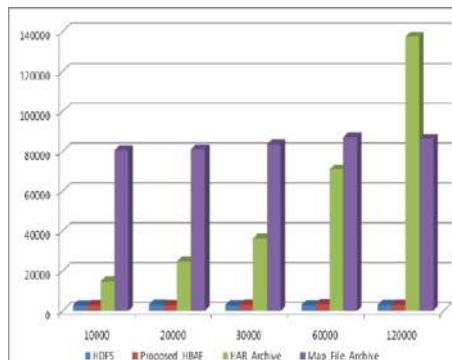
Figure 8. Access Time for 50 Files without Caching
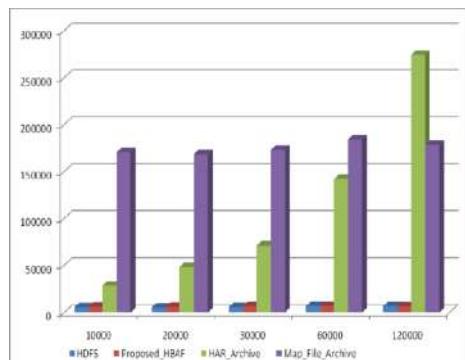
Figure 9. Access Time for 100 Files without Caching

Native HDFS will perform 8% to 14% faster than HBAF_Archive but the performance of Native HDFS degrades in comparison to the cache enabled environment. As the Figure.9 depicts that in case of accessing 100 files from the archives, HBAF_Archive is 5% to 43% faster than the HAR_Archive that is more or less equal to the previous reading while accessing 10 and 50 files. Experimental result shows that as usual Map_File_Archive will take the maximum access time and HBAF_Archive is 26 times to 28 times faster than the Map_File_Archive, these results are same as it was with accessing 50 files. Native HDFS will perform 8% to 11% faster than HBAF_Archive. There are also few cases where proposed HBAF_Archive will perform better than the Native HDFS but the difference in the performance is minor therefore it can be neglected.

5. Conclusion and Future Work

The purpose behind the design of the HDFS is to manage the daily growing big data/large files efficiently and ensure the data availability at all time along with fast access of the data. The design of HDFS is not compatible with the small files means handling small files is quite complex in the HDFS in terms of accessing of small files and their meta-data management. There is requirement of the mechanism that will handle the small files efficiently as well as reduce the Name_Node memory usage and access time for the small files. Number of researchers worked in this field and proposed various solutions to efficiently handle the small files. Most of the solutions provided reduce the meta-data usage of the Name_Node by shifting the process of indexing at the client side but these approaches are lagging behind while analyzed in terms access time. There is a requirement of such a method that will reduce the Name_Node memory usage as well as provide fast access to the small files. This paper presents Hash Based Archive File (HBAF) that will provide reasonable fast meta-data access for the small files along with the appending facility after creation of the archive. Data Node's are used for the purpose of the caching; this concept will reduce the memory pressure from the client side that results in the reduced access time for the small files. Small file's meta-data will be placed to the particular slave index file with help of the special hash function (SSHF). The use of this hash function for placement of meta-data will lead to the efficient seek operation for accessing the content of the small files. To preserve the order of the meta-data stored in the final index files a order preserving hash function (HT-MMPHF) is used. With help of this hash function, we will be able to read the final index files partially means when there is a access request for a particular small file's meta-data, final index files are read partly (only the part which contain the accessed file's meta-data), there is no need to read the entire index file hence this will result in the faster meta-data access form the final index files.

Experimental result shows that Proposed HBAF Archive performing better than the HAR Archive and Map File Archive. It is clear that when caching is disabled the access time will be very higher in case of the HAR Archive, its due to the multi level index files in the HAR Archive, but our approach is not affected form the impact of caching enable or disabled as our approach is not dependent on the client's memory. There is little limitation in our approach that cannot be addressed in this paper; these limitations will be resolved in future. The following are the key points for the future work on our approach.

- Experimental result shows that in terms *of* access time our approach performing better than the HAR Archive and Map File Archive but when it compared to the native HDFS, the results are not satisfactory; the further work can be carried out to make our approach better than the native HDFS.
- A number of other hash function combination can be used to further improve the performance of the proposed HBAF Archive.
- In this paper text files are considered as data sets, proposed HBAF Archive can also be modified for other file formats and results can be compared with original one.
- The minimization of client memory usage can be carried out, there is need to identify the various factors in our approach that are still using the client memory i.e., client memory used by the hash functions.

- We have proposed the appending facility to proposed HBAF Archive; deletion facility is still a future work in our approach.
- While merging small files, we have not applied any compression technique, implementing a compression technique at HDFS block level will be a future work for our approach.

References

[1]   Belazzougui, Djamal & Boldi, Paolo & Pagh, Rasmus & Vigna, Sebastiano. (2009), Theory and Practice of Monotone Minimal Perfect Hashing. 2009 Proceedings of the 11th Workshop on Algorithm Engineering and Experiments, *ALENEX 2009. 132-144. 10.1137/1.9781611972894.13.*

[2]   Belazzougui, D., Boldi, P., Pagh, R., & Vigna, S. (2009). Monotone minimal perfect hashing: searching a sorted table with O(1) accesses. *SODA.*

[3]   Mendelson, G., Vargaftik, S., Barabash, K., Lorenz, D.H., Keslassy, I., & Orda, A. (2018). AnchorHash: A Scalable Consistent Hash. *ArXiv, abs/1812.09674.*

[4]   Claessen, K., & Palka, M. (2013). Splittable pseudorandom number generators using cryptographic hashing. *Haskell '13.*

[5]   Zhang D., Manolopoulos Y., Theodoridis Y., Tsotras V.J. (2018) Extendible Hashing. In: Liu L., Özsu M.T. (eds) Encyclopedia of Database Systems. Springer, New York, NY. https://doi.org/10.1007/ 978-1- 4614-8265-9_741

[6]   Tchaye-Kondi, et al. (2019) Hadoop Perfect File: A Fast Access Container for Small Files with Direct in Disc Metadata Access. ArXiv.org, 26 Apr. 2019, arxiv.org/abs/1903.05838.

[7]   Peng, J., Wei, W., Zhao, H., Dai, Q., Xie, G., Cai, J., & He, K. (2018). Hadoop Massive Small File Merging Technology Based on Visiting Hot-Spot and Associated File Optimization: 9th International Conference, BICS 2018, Xi'an, China, July 7-8, 2018, Proceedings. 10.1007/978-3-030-00563-4_50.

[8]   Cai, Xun, et al. (2018) An Optimization Strategy of Massive Small Files Storage Based on HDFS. Proceedings of the 2018 Joint International Advanced Engineering and Technology Research Conference (JIAET 2018), 2018, doi:10.2991/jiaet-18.2018.40.

[9]   Kim, H., & Yeom, H. (2017). Improving Small File I/O Performance for Massive Digital Archives. 2017 IEEE 13th International Conference on e-Science (e-Science). doi:10.1109/escience.2017.39.

[10] Lyu, Y., Fan, X., & Liu, K. (2017). An Optimized Strategy for Small Files Storing and Accessing in HDFS. 22017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC). doi:10.1109/cse-euc.2017.112.

[11] Fu, X., Liu, W., Cang, Y., Gong, X., & Deng, S. (2016). Optimized Data Replication for Small Files in Cloud Storage Systems. Mathematical Problems in Engineering, 2016, 1–8. doi:10.1155/2016 /4837894.

[12] Mu, Q., Jia, Y., & Luo, B. (2015). The Optimization Scheme Research of Small Files Storage Based on HDFS. 2015 8th International Symposium on Computational Intelligence and Design (ISCID). doi:10.1109/iscid.2015.285.

[13] Wang, T., Yao, S., Xu, Z., Xiong, L., Gu, X., & Yang, X. (2015). An Effective Strategy for Improving Small File Problem in Distributed File System. 2015 2nd International Conference on Information Science and Control Engineering. doi:10.1109/icisce.2015.35.

[14] He, H., Du, Z., Zhang, W., & Chen, A. (2015). Optimization strategy of Hadoop small file storage for big data in healthcare. The Journal of Supercomputing, 72(10), 3696–3707. doi:10.1007/ s11227-015-1462-4.

[15] Fu, S., He, L., Huang, C., Liao, X., & Li, K. (2015). Performance Optimization for Managing Massive Numbers of Small Files in Distributed File Systems. IEEE Transactions on Parallel and Distributed Systems, 26(12), 3433– 3448. doi: 10.1109/tpds. 2014.2377720.

[16] Mao, Yingchi, et al. (2015), Optimization Scheme for Small Files Storage Based on Hadoop Distributed File System. International Journal of Database Theory and Application, vol. 8, no. 5, 2015, pp. 241–254., doi:10.14257/ijdta.2015.8.5.21.

[17] Dong, B., Zheng, Q., Tian, F., Chao, K.-M., Ma, R., & Anane, R. (2012). An optimized approach for storing and accessing small files on cloud storage. Journal of Network and Computer Applications, 35(6), 1847–1862. doi:10.1016/j. jnca.2012. 07.009.

[18] Ahad, M. A., & Biswas, R. (2018). Dynamic Merging based Small File Storage (DM-SFS) Architecture for Efficiently Storing Small Size Files in Hadoop. Procedia Computer Science, 132,

[19] Tarjan R.E., Werneck R.F. (2007) Dynamic Trees in Practice. In: Demetrescu C. (eds) Experimental Algorithms. WEA 2007. Lecture Notes in Computer Science, vol 4525. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-72845-0_7

[20] Bok, K., Oh, H., Lim, J. et al.(2017) An efficient distributed caching for accessing small files in HDFS. Cluster Compute 20, 3579–3592 (2017). https: //doi.org/10.1007/s10586-017-1147-2

[21] Dong, Bo & Zhong, Xiao & Zheng, Qinghua & Jian, Lirong & Liu, Jian & Qiu, Jie & Li, Ying. (2010). Correlation Based File Prefetching Approach for Hadoop. Proceedings - 2nd IEEE International Conference on Cloud Computing Technology and Science, CloudCom 2010. 41-48. 10.1109/CloudCom.2010.60.

[22] Zhang, Jing & Wu, Gongqing & Xuegang, Hu & Wu, Xindong. (2012). A Distributed Cache for Hadoop Distributed File System in Real-Time Cloud Services. Proceedings - IEEE/ACM International Workshop on Grid Computing. 12-21. 10.1109/Grid.2012.17.

Study on Product Opinion Analysis for Customer Satisfaction on E-Commerce Websites

Ms. Vijaya Sagvekar [a, 1], Dr. Prashant Sharma [b]

[a] *Research Scholar, Department of Computer Engineering, PAHER, Udaipur, India*

[b] *Associate Professor & HOD Computer Science & Engineering Department PAHER, Udaipur (Rajasthan),* India

Abstract---- The E-commerce websites have been emerged in a high range of marketing benefits for the users to publish or share the experience of the received product by posting review that contain useful comments, opinions and feedback on the product. These days, a large number of clients acquire freedoms to look at comparative items in online sites and pick their top choices in computerized retailers, like Amazon.com and Taobao.com. Client audits in online media and electronic trade Websites contain important electronic word data of items. Sentiment Analysis is broadly applied as voice of clients for applications that target showcasing and client care. Sentiment extractors in their most essential structure classify messages as either having a good or negative or once in a while neutral supposition. A typical application of sentiment investigation is the programmed assurance of whether an online review contains a positive or negative review. Subsequently, in this paper, with the use of the strategies on sentiment analysis, obstinate sentences alluding to a particular element are first recognized from item online audits. We have proposed deep learning strategy as a classification model for discovering the condition of review. The outcomes showed suggested site for the client dependent on the early audits, past reviews and answer given to inquiry audit for the client. Additionally, it is seen that the proposed strategy can ready to answer every one of the reviews with a superior closeness like a human reaction to the client.

Keywords--- E-commerce, opinion analysis, sentiment analysis, deep learning

1. Introduction

The marketers and manufacturers have been focused on the market performance from long time. The development of product marketing strategy for managing the product quality is helpful in making a better decision shows improvement in the performance [1]. The manufacturers and advertisers continuously gather the product information that useful in analyzing the performance in the market. In the traditional methods, the main data sources for analysis of main data sources were collected from the manufacturers" internal data, offline customer reviews by the surveys, review forms should be in handwritten format etc [2].

---

[1] Ms. Vijaya Sagvekar, Research Scholar, Department of Computer Engineering, PAHER, Udaipur
Email: vijayasagvekar@gmail.com.

As the technology became advanced, the e-commerce websites enabled to publish the opinion of users on product and the users got a public platform to share their useful comments and opinions towards the product they purchased. According to the survey, 97 % of consumers will be influenced for purchasing the product after going through the reviews posted by other customers. So, a most of clients will see online reviews prior to settling purchase decision [3,4].

91 % of people regularly or occasionally read online reviews before purchasing a product and also the early reviews on product before purchasing has high impact on succeeding the product sales [5]. Even though, the early reviewers contribute small proportion of reviews, it is easy to determine the failure or success of new services and new products. Based on the early reviews from the early reviewers helps to adjust marketing strategies in product improvement on designs and helps in succeeding the new product [6]. Based on the early reviews, the companies will recognize the early reviews thereby improve the product designs, marketing strategies that lead to a success on newly launched products. Thus, early analysts become the accentuation to screen and pull in at the early advancement phase of an organization in the world [7-9]. The early reviews have pulled in showcasing experts broadly investigate the customer buy goals. For example one of the largest e-commerce companies is Amazon in the world, where it provides early reviewer program opportunities for making early reviews that helps the company to acquire those early reviews on products that have few reviews or no reviews. With the help of Amazon shoppers program, will provide information about buying a product by making smarter decision. Based on the above discussions, we can see that for product marketing the most important is early reviewers of product [10-14].The present research will take initiative for studying the behavior characteristics for the early reviewers and posting it in e-commerce platform such as Amazon, Yelp. The main aim of the research work is to analyze the consumer satisfaction on products bought from E-commerce sites and improved the performance. The overall characteristics of early reviews have to be analyzed from the early reviewers needed to be compared to majority and laggard reviewers. An early reviewer tends to post more helpful reviews and helps to gives a higher average rating score to products. The rating behaviours are characterized that helps to find the scores received from others that helps to determine the correlation of the reviews based on the product popularity [15]. The discoveries with the character factors are connected with the hypothesis as follows: higher normal rating scores can be considered as the ideal attitude towards the items, and higher support votes of early surveys given by others can be seen as an intermediary proportion of the opinion leadership.

2. Literature Review

Many researches have been developed for predicting the ranking based on the online reviews. Some of the studies are as follows. In the literature survey, a survey of recent techniques is highlighted with its advantage and limitations.

Table 1. Literature Review

Authors	Method	Advantage	Disadvantage
Ahani, A., et al [16]	Self-Organizing Map and Higher Order Singular Value decomposition clustering algorithm	The hybrid algorithm was used for assisting that overcame the data related complications for online reviews and presented spa hotel market segmentation for predicting the travel choice using machine learning algorithms.	The available customer data from TripAdvisor included only the general preferences of spa hotel customers degraded the performance rate

Rita, P et al [17]	Four-dimensions of e-service quality model	The four dimensions of e-service quality were considered such as the impact on customer trust, satisfaction, customer behaviour and building existed literature based on e-service quality during online shopping.	The quality of online stores, in general, was not based on the product segments sold in online stores and the measurements were not applicable for assessing product segments.
Lucini, F.R., et al [18]	Texting Mining approach Latent Dirichlet Allocation (LDA)	The developed model presented a novel framework for customer satisfaction measuring in the airline industry using Latent Dirichlet Allocation (LDA) that detected the popular topic using the natural language processing and machine learning process	The data restricted the diversity of opinions as the proficient in English were in more and likely provided airline experience information.
Bai, T. et al [19]	Margin based Embedding Ranking Model (MERM)	The characterized and predicted the early reviewers for E-commerce sites to present effective product marketing. The developed model used Margin based Embedding Ranking Model (MERM) that predicted the early reviewers in a cold-start setting	The developed model used Margin based Embedding Ranking Model (MERM) that predicted the early reviewers in a cold-start setting.
Zhao, Y.,et al [20]	Technical attributes and sentiment polarity	The review samples were taken from trip advisor that predicted overall customer satisfaction using technical attributes in online and textual reviews of customers.	The textual reviews are influenced by languages and are different cultures needed extension for examining different language reviews.
Liu, Y. et al [21]	Product Competitive and Quality Management and Marketing Strategy	The developed a product competitive advantage analysis for providing an essential basis for quality management and marketing strategy development on social media. The novel method provided an essential basis for managing the marketing strategy and quality using user generated content	The customer comments relied which comes after target product availability of the customers and was not applicable during the design and development of the product.
Jian Jin et al [22]	Product feature extraction and sentiment analysis	The developed model performed opinionated representative for specific product based on the features especially for the competitive products. The sentimental analysis was performed for opinionated sentences that refer specific features for online reviews.	During choosing comparative sentences in the review that were of many for different products resulted lower information comparativeness values.
Sun, Q., et al [23]	Sentiment analysis eWOM	The developed model extracted large volume of online customer reviews and performed sentiment analysis for eWOM products. The developed model used semi-supervised fuzzy product ontology mining algorithm for extraction of features with negative or positive labels.	The developed model required improvement in positive and negative opinion words extraction and also polarity computation.
Thompson, J.J., et al [24]	Lexicon-Based Sentiment Extraction Analysis	The developed model performed sentimental analysis for chat messaging by the player who was involved in video game StarCraft 2. The developed model performed sentimental analysis was applicable for toxicity detection and also identified the players and their messages that are threat for the player.	The human raters disagree not only the sentiment but also disagreed the toxicity that suggested that the task is not straight forward and thus the performance at high rate was impossible

Kumar, S., et al [25]	Sentimental Analysis For The Product Review For EEG Response	The developed model performed multimodal framework for the estimation of product rating on customer product and their brands. The reviews obtained from the global viewers were processed using Natural Language Processing (NLP) technique that computed the score for global rating.	The emotional state needed to be considered for improving the performance.

3. Motivation for Study

There are numerous quantities of studies that examine the monetary results or the drivers of online reviews and, simultaneously, propose suggestions for the plan of review frameworks, for example, giving reviewers a predefined review format. Albeit a significant number of review framework configuration highlights have been proposed throughout the long term, truth be told, not very many have really been dissected. The new online plans of action and conditions have arisen including two-sided stage organizations (e.g., Uber). These empower two-sided reviews and require adjusted plan highlights to, for example, alleviate correspondence in two-sided review frameworks. Finally, most plan highlights of review frameworks have been investigated for fixed gadgets like PCs. Notwithstanding, online reviews are progressively delivered and burned through cell phones, which require explicit plan highlights. These are the unsolved problems in the ranking analysis for online reviews and many research work based on this problem are still examining.

4. Motivation for Study

There are numerous quantities of studies that examine the monetary results or the drivers of online reviews and, simultaneously, propose suggestions for the plan of review frameworks, for example, giving reviewers a predefined review format. Albeit a significant number of review framework configuration highlights have been proposed throughout the long term, truth be told, not very many have really been dissected. The new online plans of action and conditions have arisen including two-sided stage organizations (e.g., Uber). These empower two-sided reviews and require adjusted plan highlights to, for example, alleviate correspondence in two-sided review frameworks. Finally, most plan highlights of review frameworks have been investigated for fixed gadgets like PCs. Notwithstanding, online reviews are progressively delivered and burned through cell phones, which require explicit plan highlights. These are the unsolved problems in the ranking analysis for online reviews and many research work based on this problem are still examining.

5. Proposed Methodology

In our paper, the proposed a application for predict online items in different E-commerce sites like Amazon, Paytm, Flipkart, ShopClues and Snapdeal in reviews of clients. Firstly, we processed the datasets of different items of particular organization from E-commerce sites. Each product has its own features for categorized depends on a particular feature. This features such as significantly positively or negatively influences customer ratings, readability, subjectivity, length–significantly, and sentiment polarity. The entire surveys of client that refreshed itemsbased are generated for collecting FCM model [27].
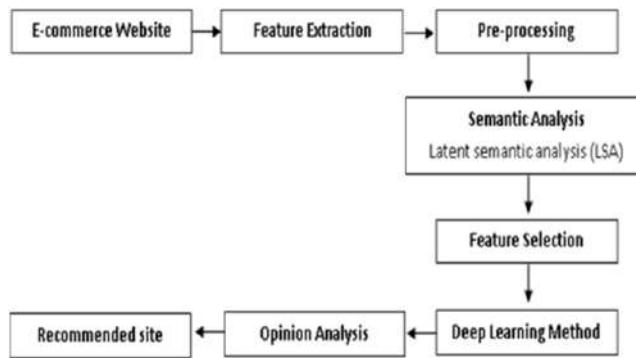
Figure 1. Flow graph of Proposed Method

Dataset

The proposed research uses a row dataset obtained from online E-shopping sites (ShopClues, Paytm, Flipkart, Amazon and Snapdeal). Each and every site is important in introducing a chance to buy the product by great population. The popularity of these sites is identified by various administrations that help to improve the distribution of various products belonging to various categories. Among the E-commerce database, Amazon dataset is used for the present research that has around142 million product reviews taken for the duration May 1996 to July 2014. Each of the review from the site consisted of textual comment that was posted by the product user that accompanied to publish time stamp accurate in the study. Usually the reviews are associated scale that is having up to five star of rating that is associated with textual description. A review an audit is related with a rating score in a five-star scale. Every item is related with a category label and a textual description.

Feature Extraction The reviews from the dataset are extracted using feature extraction technique that reviews of collected item from the E-commerce sites improved the quality of review during analysis. The feature extraction includes the items enhancement for performing audit rating of each item. The attributes here refers to the cost, quality, positive survey. The feature extraction mines the opinion of customer reviews, summarize the reviews, store and produce step preparation.

Pre-processing Transforming text into something an algorithm can process is a confounded process. In this part, the means associated with text processing are as per the following. (i) Tokenization: The superfluous tokens are not difficult to filter, where a document is changed into passages or sentences into words. In our work, the online review is tokenizing into words. (ii) Removal of Unnecessary labels and Punctuation: The following stage is to eliminate Punctuation, as the Punctuation doesn't do any additional data while treating text data. (iii) Removing stop words: Frequent words such as "the", "is", etc. that do not have explicit semantic. (iv) Lemmatization: Another way to deal with eliminate affectation by deciding the grammatical form and using point by point data set of the language.

Semantic Analysis The semantic analysis is performed for the preprocessed data where, the natural language content reads all the words captures the content by performing real meaning of any text. The text elements are assigned and are assigned based on the logical and grammatical role. The semantic analysis analyzes the surrounding context in the text to accurately disambiguate the exact meaning of words. The relationship

between the concepts in the text is also developed to identify the most relevant elements in text and understand the topic discussed. In the semantic analysis, Latent Semantic Analysis (LSA) was used in NLP that analyses the relationship among the set of documents and the terms they contain by creating a set of concepts related to the documents and terms.

Feature Selection From the extracted concepts, feature selection process is performed for selecting the subset terms that were occurred in training and these selected subsets were treated as features that performed text classification. Firstly, the training is performed and is applied for the classifier decrease the size of the vocabulary effectively. Secondly, feature selection improves the classification accuracy thereby eliminates the noise features in figure1.

Deep Learning the deep learning methods provide an opportunity that faces challenges in NLP problems such as sequence-to-sequence prediction. The developed model performed deep learning methods for learning the features based on the NL which is required by the model specifies the required features and were extracted. In natural language processing, the performance of deep learning is depending on genuine outcomes and that the enhancements give off an impression of being proceeding and maybe speeding up.

Opinion Analysis The optimal ranking opinion assesses the significance of each element relatively with respect to the sentiment score that utilized for rank highlights measuring. Based on the opinion analysis, the commentators rated the sites that are having vital data distinguished the untruthful opinions.

Recommended Site Based on these recommendations, a perspective positioning calculation were performed for ranking the vital angles resulted a viewpoint recurrence that impacted opinions for each perspective for general sentiments.

6. Results and Discussion

For accurate recommendation of products a new opinion analysis system is developed, by analyzing the reviews that users are posted for the products. The main aim of this research is to create a accurate keyword extraction technique and clustering approach for recommending the products using amazon customer review dataset interms of positive and negative form. In this paper, with GWO algorithm, a keyword extraction method (LDA) is used for selecting the proper key words. The acquired same keywords are clustered using PFCM algorithm. The developed recommendation system has main advantage is that, system has ability to find the fake products, keep track of clients satisfaction etc.

Table 2. Comparison between proposed systems and different classifiers.

	Precision	Recall	F1-Measure	AUC	Accuracy
Random Forest	69.003	73.25	71.132	48.263	73.187
Decision Tree	75.766	75.583	74.618	54.628	75.528
Proposed	76.106	76.781	74.167	57.404	77.236

The proposed system conveyed a powerful execution through quantitative analysis and comparative analysis. From the test analysis, the proposed system accomplished around 77.236% 77.236% of classification accuracy, but the existing techniques achieved

limited accuracy in amazon customer review dataset. In future work, an accurate system is implemented for further improvement in the classification accuracy for recommendation of product.

7. Conclusion

Now a days the developer community is more and more focusing on the user experience of browsing, because number of users browsing the internet are exponential increased. In this paper, proposed the approach to predict the most effective web-based shopping sites. By client survey determined whether poor or great is the product from business websites. It is more important for any business to have the knowledge about the reviews of customer regarding its any particular items.

In this paper, to analyze the characteristics and categorize early reviewers on an e-shopping sites and their effect on product popularity will use the Amazon datasets. This proposed study review posting process and develop a deep learning model for the prediction of reviewers. We also mine the summary of surveys, customer reviews. Based on the opinion analysis, a recommended perspective positioning computation to rank the fundamental points by contemplating both the viewpoint recurrence and the impact of opinions gives the recommended site.

References

[1]   Smith, A.N., Fischer, E. and Yongjian, C., 2012. How does brandrelated user-generated content differ across YouTube, Facebook, and Twitter?. Journal of interactive marketing, 26(2), pp.102-113.

[2]   Dong, J., Li, H. and Zhang, X., 2014, November. Classification of customer satisfaction attributes: an application of online hotel review analysis. In Conference on e-Business, e-Services and e-Society (pp. 238-250). Springer, Berlin, Heidelberg.

[3]   De Giovanni, P. and Zaccour, G., 2014. A two-period game of a closed-loop supply chain. European Journal of Operational Research, 232(1), pp.22-40.

[4]   Farhadloo, M., Patterson, R.A. and Rolland, E., 2016. Modeling customer satisfaction from unstructured data using a Bayesian approach. Decision Support Systems, 90, pp.1-11.

[5]   Elish, M.O., 2014. A comparative study of fault density prediction in aspect-oriented systems using MLP, RBF, KNN, RT, DENFIS and SVR models. Artificial Intelligence Review, 42(4), pp.695-703.

[6]   Clavel, C. and Callejas, Z., 2015. Sentiment analysis: from opinion mining to human-agent interaction. IEEE Transactions on affective computing, 7(1), pp.74-93.

[7]   Nahm, Y.E., 2013. A novel approach to prioritize customer requirements in QFD based on customer satisfaction function for customer-oriented product design. Journal of Mechanical Science and Technology, 27(12), pp.3765-3777. Yang, L., Geng, X. and Liao, H., 2016. A web sentiment analysis method on fuzzy clustering for mobile social media users. EURASIP Journal on Wireless Communications and Networking, 2016(1), p.128.

[8]   Yang, J.B., Wang, Y.M., Xu, D.L., Chin, K.S. and Chatton, L., 2012. Belief rule-based methodology for mapping consumer preferences and setting product targets. Expert Systems with Applications, 39(5), pp.4749-4759.

[9]   Decker, R. and Trusov, M., 2010. Estimating aggregate consumer preferences from online product reviews. International Journal of Research in Marketing, 27(4), pp.293-307. [11] Dehkharghani, R., Mercan, H., Javeed, A. and Saygin, Y., 2014. Sentimental causal rule discovery from Twitter. Expert Systems with Applications, 41(10), pp.4950-4958.

[10] Chen, A., Lu, Y. and Wang, B., 2017. Customers'' purchase decisionmaking process in social commerce: a social learning perspective. International Journal of Information Management, 37(6), pp.627-638.

[11] Uysal, A.K. and Gunal, S., 2014. The impact of preprocessing on text classification. Information Processing & Management, 50(1), pp.104- 112.

[12] Zhang, W., Yoshida, T. and Tang, X., 2011. A comparative study of TF* IDF, LSI and multi-words for text classification. Expert Systems with Applications, 38(3), pp.2758-2765.

[13] Zhao, T. and Eskenazi, M., 2016. Towards end-to-end learning for dialog state tracking and management using deep reinforcement learning. arXiv preprint arXiv:1606.02560.

[14] Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L. and Weaven, S., 2019. Market segmentation and travel choice prediction in Spa hotels through TripAdvisor"s online reviews. International Journal of Hospitality Management, 80, pp.52-77.

[15] Rita, P., Oliveira, T. and Farisa, A., 2019. The impact of e-service quality and customer satisfaction on customer behavior in online shopping. Heliyon, 5(10), p.e02690.

[16] Lucini, F.R., Tonetto, L.M., Fogliatto, F.S. and Anzanello, M.J., 2020. Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. Journal of Air Transport Management, 83, p.101760.

[17] Bai, T., Zhao, W.X., He, Y., Nie, J.Y. and Wen, J.R., 2018. Characterizing and predicting early reviewers for effective product marketing on e-commerce websites. IEEE Transactions on Knowledge and Data Engineering, 30(12), pp.2271-2284.

[18] Zhao, Y., Xu, X. and Wang, M., 2019. Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. International Journal of Hospitality Management, 76, pp.111-121. [

[19] Liu, Y., Jiang, C. and Zhao, H., 2019. Assessing product competitive advantages from the perspective of customers by mining usergenerated content on social media. Decision Support Systems, 123, p.113079.

[20] Jin, J., Ji, P. and Gu, R., 2016. Identifying comparative customer requirements from product online reviews for competitor analysis. Engineering Applications of Artificial Intelligence, 49, pp.61-73.

[21] Sun, Q., Niu, J., Yao, Z. and Yan, H., 2019. Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. Engineering Applications of Artificial Intelligence, 81, pp.68-78.

[22] Thompson, J.J., Leung, B.H., Blair, M.R. and Taboada, M., 2017. Sentiment analysis of player chat messaging in the video game StarCraft 2: Extending a lexicon-based model. Knowledge-Based Systems, 137, pp.149-162. [25] Kumar, S., Yadava, M. and Roy, P.P., 2019. Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. Information Fusion, 52, pp.41-52.

[23] Pradip Suresh Mane, Ashok Kumar Jetawat, Pravin Jagannath Nikumbh", 2020, "Web Page Recommendation using Random Forest with Fire Fly Algorithm in Web Mining", International Journal of Engineering and Advanced Technology (IJEAT), Vol. 9, Issue-3, February 2020, pp. 499-505.

[24] Kolhe L.N., Khairnar V., Jetawat A.K. :Prediction-Based Parallel Clustering Algorithm for M-Commerce. In: Fong S., Akashe S., Mahalle P. (eds) Information and Communication Technology for Competitive Strategies. Lecture Notes in Networks and Systems, vol 40. Springer, Singapore, pp. 31-39

A Survey on Different Techniques for Product Fake Review Detection and Product Rating

Adnan Telwala [a,1], Ayush Pratap [a], Ketan Gaikwad [a], Tushar Chaudhari,Sukhada Bhingarkar [a]

[a] School of Computer Engineering and Technology, MIT World Peace University, Kothrud, Pune 411 038, Maharashtra – India

Abstract. Numerous online business sites empower the customers to create a product reviews along with feedback in the shape of ratings. This gives the organization work force a sign about their items' remaining on the lookout, while likewise empowering individual customer to frame an assessment and help buy an item. As of late, Sentiment Analysis (SA) has gotten quite possibly interesting due to the potential business advantages of text analysis. One of the most important problems in confronting SA is the manner by which to remove feelings in the assessment, as well as how to identify counterfeit good reviews and negative surveys derived from assessment surveys. Besides, the assessment surveys acquired from clients can divided into two categories: positive and negative, which can be utilized by a shopper to choose an item. In this survey, we have thoroughly discussed about fake review detection of products as well as product rating by different SA techniques. Further, we have discussed the research direction in fake review detection and product rating.

Keywords- Commercial benefits, Fake reviews, Opinion reviews, Opinion mining, Product rating, Sentiment analysis, Text analysis.

1. Introduction

As web-based business keeps on developing, the opposition among vendors for customers and deals has expanded essentially. A few traders decide to acquire an upper hand by utilizing misleading strategies to control online business stages, like the Amazon Marketplace. These beguiling strategies can incorporate controlling calculations, counterfeit item audits and evaluations just as snap ranches. Traders utilize these and different strategies to build their deals and gain piece of the overall industry. [1]

Ordinarily, a shipper will either boost its clients with blessings, money related remuneration or potentially guarantees to create and to compose an item survey or enroll the assistance of an outsider to creator an item audit. These reviews assist with driving buyer conduct; regardless of whether they are shopping through internet business stages or as our forefathers would have done it in retail outlets the nation over

---

[1] Adnan Telwala, School of Computer Engineering and Technology, MIT World Peace University, Email: adnantelwala31@gmail.com.

Individuals depend on the audits and evaluations on Amazon when settling on buying choices [1]. The Washington Post nitty gritty in what way sellers make use of Facebook to help their Amazon assessments. The assessment questionnaire in December 2018 has discovered that 61% reviews for contraptions are fake, and in the meantime, overviews for supplements are 64% phony. One more assessment in 2019 discovered that 82% of reviews read by the buyers are fake. [2] The fictitious product reviews are made and recorded either based on the dealer's item or on its rivals' items. On account of the contender, they may have a case identified with slander since the surveys will adversely misrepresent their item and its highlights. Then again, when the fake reviews are recorded on the trader's item, the rival may have a case identified with fake promoting and the reviews would be constructive misrepresent the dealer's own item and its highlights, accordingly arrogate deals from the contender. [2]

Existing methodologies for fake review detection and product rating are as follows:

Opinion Mining (OM) otherwise called Sentiment Analysis (SA) is the area of research that analyze individual's opinion, assessments, conclusions, sentiments, evaluations, and feelings towards substances like services people, issues, subjects, and their properties. There are various automated techniques for Sentiment Analysis. Machine learning techniques in addition with SA techniques are relied upon to own a significant beneficial outcome, particularly for fake review detection in product reviews, social networking sites, and different domains. There are diverse AI based strategies accessible like Support Vector Machine (SVM), Random Forest, and K-Nearest Neighbors (KNN) that are applied for the classification purposes. [3]

The remaining of the paper is organized as follows: Section II discusses about Fake Review Detection based on Sentimental Analysis. Section III throws a light on fake review detection features Section IV concludes the paper.

2. Fake Review Detection based on Sentimental Analysis

2.1. Iterative Computer Framework (ICF++)

Iterative Computer Framework (ICF++) [4] starts with a direct execution of a little arrangement of the product prerequisites and iteratively improves the propelling structures until the absolute system is done and fit to be sent.

ICF++ measures the trustworthiness estimation of a review by using the content mining and assessment mining methods. The name of an item, the name of an observer, the content of a survey, the amount of focus, and minsup as a boundary are all used in the ICF++ methodology. This technique adds the counter to nothing, genuineness value, and dependability worth to one after taking certain attributes. From that point onward, the following cycles are Parts of Speech (POS) labeling, production of transaction file, frequent pattern (FP) growth, generation of polarity, agreement value calculation.

Iteration is the next step. In each iteration, the system calculates the fairness, trustworthiness, and reliability values, as well as updating the agreement attribute with the updated integrity, reliability, and increment counter. A major advantage of the ICF++ is that the accuracy is enhanced. The disadvantage of this approach is that

certain processes must be streamlined in order for it to detect a fake review in a reasonable period of time.

Further steps used in ICF++ are as follows:

- Parts Of Speech (POS) tagging**:** The Stanford Part-Of-Speech (POS) tagger from stanfordcorenlp-3.5.1 [5] is used for POS tagging. POS Tagger is a product that receives messages and selects the appropriate linguistic structure for each token. For English taggers, this connection transmitted a tag as shortenings, and the truncation's Penn Treebank standard is used. A Java-based work territory framework is created for this communication. The setup of the tagger model was the first step in this process. This connection's commitment is a sentence taken from the analysis and development of a POS tag for each component that is inserted into the database.

- Creation of transaction file: All tokens that are stored in the database from objects are included in the transaction file analysis. It has the tag value of noun (NN), noun pronoun (NNP), noun pronoun singular (NNPS), or noun singular (NNS). Each object review is represented by a noun in the transaction file. This transaction file is used to calculate the FP Growth contribution.

- Frequent Pattern (FP) Growth: This communication involves understanding the highlights that have received the most reviews. The highlights of this investigation are the properties or characteristics of an item. For example, the highlights of a camera may be the battery, memory card, and so on. This data is available on Amazon, but datasets associated with it are not, and so to get data about an item's component, this technique has used FP-development measurement from affiliation rule mining tool.

2.2. *Graph based* Model **for Fake Review Detection**

This section focuses on graph based fake review detection in context of online product reviews. This model replaces store nodes with product nodes and introduces new scoring criteria to capture the intricate relationships between all kinds of nodes. Moreover, this model exploits more features relating to each kind of nodes to govern the spam city of reviews, reviewers, and products, which can achieve a remarkable precision improvement.

Here are some node features in product review graph. These features are related with reviewers, reviews and products [6].

2.3. Reviewer-related features:

Reviewer's Review Content Similarity (RRCS): Fake reviewers are likely to copy (or slightly modify) their previous reviews across similar products to save time and energy. RRCS of "r" is the average value of the entire cosine similarities between each review text pairs written by reviewer.

$$RRCS(r) = avg_{v_i,v_j} \in V_r, i < j^{cosine(v_i,v_j)} \quad (1)$$

Average Rating Score (ARS): If a reviewer's average rating score is too low, then that reviewer intends to be a spammer.

$$ARS(r) = (5 - avg_{v \in V_r} \Psi_v)/4 \qquad (2)$$

Where Ψv is the rating score of review v in a 5-star rating system.

Reviewer Active Duration (RAD): Opinion spammers are normally newcomers to a website. Let $L(r)$ and $F(r)$ be the last and first review date of reviewer r, T be a user-specified time threshold, then:

$$RAD(r) = \begin{cases} 1 - \frac{L(r) - F(r)}{T}, & L(r) - F(r) \leq T \\ 0, & L(r) - F(r) > T \end{cases} \qquad (3)$$

Ratio of First Reviews (RFR)
Review spammers intend to post reviews early to influence the sentiment of a product. This function is described as the ratio of the number of reviews in the first k% of a product's reviews to the total number of reviews written by the reviewer r:

$$RFR(r) = \frac{|\{v | v \in V_r, v \text{ is in the first } k\% \text{ reviews}\}|}{|V_r|} \qquad (4)$$

Review Date Entropy (RDE): Genuine reviewers usually post reviews spontaneously, while review spammers often submit a significant number of reviews in a limited period of time (e.g., one day) to gain the maximum profits. This model feature by the entropy of review dates of that reviewer. Clearly, spammers have lower review date entropies than those of genuine reviewers.

$$RDE(r) = 1 + \sum_d \frac{|V_r^d|}{|V_r|} \log_{|D_r|}(\frac{|V_r^d|}{|V_r|}) \qquad (5)$$

2.4. Review-related features

1. Review Content Similarity (RCS): For a single review v about product p, let V_p be p's review set, then RCS is defined as the maximum cosine similarity of review v and its earlier review v′ in V_p:

$$RCS(v) = max_{v' \in V_p, date(v') < date(v)} cosine(v, v') \qquad (6)$$

2. Rating Deviation (RD):
Review spammers want to advertise or degrade their products, because their reviews for the same product which differ significantly from those of other reviewers. This model features as:

$$RD(v) = \frac{|\Psi_v - avg_{v' \in V_p} \Psi_{v'}|}{4} \qquad (7)$$

2.5. Product-related features

1. Average Rating (AR): The average rating score of product p reflects the quality of that product. Average rating is defined as:

$$AR(p) = \frac{5 - avg_{v \in V_p} \Psi_v}{4} \qquad (8)$$

Total Number of Reviews (TNR)- Good quality products deserve more reviews. Let P be the product set, then

$$TNR(p) = 1 - \frac{|V_p|}{max_{p' \in \rho} |V_p|} \qquad (9)$$

2.6. *Iterative Computation **with Elimination (ICE)***

The computation framework can also be applied to product review graph model. However, because the number of products are considerably larger than that of stores, the computation will take more time than with stores. Therefore, this framework proposes a refined algorithm ICE to compute the node scores efficiently, as shown below. An advantage of this framework is that it is a novel algorithm to efficiently compute the reputation scores of each kind of nodes in the graph. However, Individual spam features (ISF) had a weak output. Linguistic features (LF) did not do well.
Algorithm: ICE (Iterative Computation with Elimination)

Input: reviewer set R, review set V, product set P, time window Δt, number of reviewers to be returned N, elimination rate ρ

Output: Fake reviewer set F, |F| = N, in ascending order of trustiness of reviewers
Step 1: Set all trustiness of reviewers to 1, set all reliability of products to 1, mark all reviewers in R as not eliminated;

Step 2: Repeat until the top N fake reviewer set F does not change or the maximum number of iterations reached

Step 2.1: Compute the honesty of reviews by non-eliminated reviewers according to

Step 2.2: Compute the trustiness of non- eliminated reviewers according to

Step 2.3: Compute the reliability of products in P according to

Step 2.4: Eliminate top ρ percent of non- eliminated reviewers based on their trustiness scores in descending order.

Step 3: Return F [6]

3. Fake Review Detection Features [7]

The features used for fake review detection are as follows: -

3.1. **Duplicate reviews and** rating: Repeated reviews of a product, comments on a product and rating consistently for a similar product or service by same ID.

3.2. **Username** consisting **numbers**: A true consumer or customer must show his or her actual name, which cannot be made up entirely of numbers. Since the genuine buyer would have a name and an address in alphanumeric notation, the only number means spammer.

3.3. **Only star rating**: Leaving a star rating would not guarantee a positive score and customers will want to suggest more before leaving a star rating. As a result, only star ratings are deemed fraudulent.

3.4. **No** category **Details**: Real purchaser will write the highlights and specifications of product in his reviews and then rate his experience of using the item. As a result, it is marked fake if a person posts a review without specifying the product classification or spotlight.

3.5. **Review** Sentiment **and rating**: These reviews are deemed fraudulent whether they have the highest rating but poor sentiment, or opposite, — for example if a user posts a positive rating but his or her review responses do not fit the rating.

3.6. **Size of** review: An authentic purchaser will leave a comment stating what he liked and did not like about the product or service when addressing the functionality of the product or service. As a result, the review must adhere to a strict size of review.

4. Discussion

Just single factor may not give exact recognizable proof of fake reviews. Consequently, value of all variables is considered in the review measure. This paper presents a review of some most prominent fake review detection method for product rating. Figure 1 below summarizes the capabilities of existing methods along with the limitations that indicate further improvement and future work.

5. Results
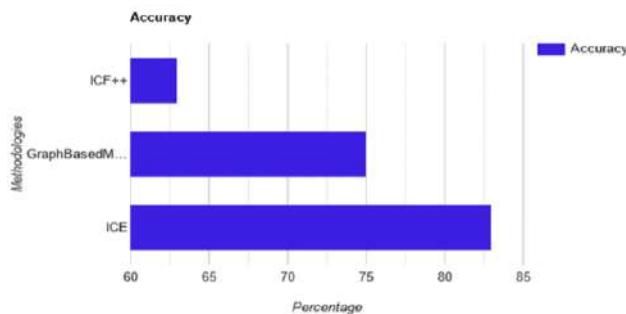
Figure 1. Existing methods along with the limitations that indicate further improvement and future work.

6. Conclusion

It is difficult to go through and review manually to purchase a new product in today's world, where each product has thousands of reviews available. Many of these reviews are spam or fraudulent, or they are focused on consumers' sentimental loyalty to a particular company or competitor. Therefore, there is need for detection of fake reviews. In this survey paper, several proposed SA methods are discussed like ICF++, ICE, Graph based model. Not all papers summarized in this survey have claimed accuracy rate. For ICF++ method, claimed accuracy rate is 63% and the limitation is that the process should be optimized for it to work in short amount of time and for ICE algorithm, the complexity of the system increases very much and accuracy is low. The graph-based model is having 75% accuracy but it significantly slows as the size of that data in use grows. However, accuracy can be increased by using machine-learning techniques like SVM, Random Forest etc. to 85%. For future work, we will apply machine learning techniques and sentimental analysis, which will increase accuracy furthermore.

References

[1] Risk management in software. (2020, October 02). Retrieved from https://www.castsoftware.com/research-labs/risk-management-in-software development-and-software-engineering- projects.

[2] E-commerce manipulation the fake review. (2020, January 2) Retrieved from https://kjk.com/2020/01/25/e-commerce-manipulation-the-fake-review- problem/

[3] Fake-reviews detection. (2021, February 17). Retrieved from https://www.cs.uic.edu/~liub/FBS/fake- reviews.html

[4] Arif Djunaidy and Eka Dyar Wahyuni. Fake Review Detection from a Product   Review Using Modified Method ofIterative Computation Framework MATEC Web of Conferences 58, 03003 (2016)

[5] POS tagger (2021, March 09). Retrieved from https://nlp.stanford.edu/software/tagger.shtml

[6] Dawei SONG, Tingting HOU, Zhun LI, Zhuo WANG. Spotting Fake Reviewers using Product Review Graph Journal of Computational Information Systems 11: 16 (2015) 5759–5767 Available at http://www.Jofcis.com

[7] Ms. Rajshri P. Kashti1, Dr. Prakash S. Prasad Enhancing NLP Techniques for Fake Review Detection International Journal of Pure and Applied Mathematics Volume 119 No. 12 2018, 13203-13209

Preprocessing and Skull Stripping of Brain Tumor Extraction from Magnetic Resonance Imaging Images Using Image Processing

Shweta Suryawanshi [a,1] and Dr. Sanjay B. Patil [b]

[a] *E & TC Department, BSCOER, Narhe, Pune, India*
[b] *Principal, Rajgad Dnyanpeeth's Shree Chhatrapati Shivajiraje College*
of Engineering Pune, India

Abstract. Many neuroimaging processing functions believe the preprocessing and skull strip (SS) to be an important step in brain tumor diagnosis. For complex physical reasons intensity changes in brain structure and magnetic resonance imaging of the brain, a proper preprocessing and SS is an important part. The method of removing the skull is relayed to the taking away of the skull area in the brain for medical investigation. It is more correct and necessary techniques for distinguishing between brain regions and cranial regions and this is believed a demanding task. This paper gives detailed review on the preprocessing and traditional transition to machine learning and deep learning-based automatic SS techniques of magnetic resonance imaging.

Keywords: Diagnosis, brain tumor, magnetic resonance image, machine learning, deep learning.

1. Introduction

In some parts of the human body, there is an unlimited growth of tumor cancer cells. There are different types of tumors that have specific specificities and different treatments. *Brain Tumors (BTs)* are classified as primary BT as well as metastatic BTs. The BT tends to stay in the brain even in the early stages, and then cancer tends to move to other parts of the body as well as spread to the brain. Actually, the majority utilized rating method has been declared by the WHO. It categorizes BTs into grade I to IV using the microscope. Premature detection of BT can be life-threatening, so diagnostic procedures have changed as needed. Medically, treatment for BT consists of surgical treatment, radiation treatment, or chemotherapy [1].

With the advancement of medical imaging, the efficiency of images plays an important role in the assessment of patients with BTs, besides; there is a significant power of patient anxiety. In modern times, increasingly new imaging methods such as Magnetic Resonance Imaging show the whole characteristics of BT and clinical doctors are developing to examine the BT method as desired. Treatment for BT involves surgery, radiation treatment, or chemotherapy. Medical practitioners play an important role in BT assessment, including treatment.

---
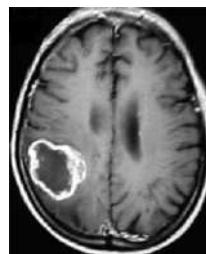
[1] Shweta Suryawanshi, E & TC Department, BSCOER, Narhe, Pune, India
Email: suryawanshi.shweta02@gmail.com.

Once BT is clinically believed, radiological assessments are made to determine the relationship between BT status, BT area, and the structure associated with it. This knowledge is essential as well as important for making decisions in a variety of treatments such as surgery, radiation treatment or chemotherapy. As a result, evaluating BT with imaging methods is one of the important aspects of radiology departments at this time [2].
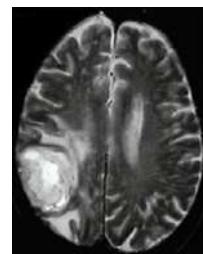
Magnetic Resonance Imaging is not involving the introduction of instruments into the body with superior soft tissue contrast imaging modality, which gives very useful knowledge regarding nature, dimension, and localization of BTs not including exposing the patient to a high ionization radiation. *Magnetic Resonance Imaging* is attracting extra with additional interests for the BT analysis in the medical field. An axial section of four standard sequences for glioblastoma is shown in Figure 1 [3]. In current medical practice, various magnetic resonance images with explanations of tumor segments are used for success analysis. These succession images include T1-Weight Magnetic Resonance Imaging (T1w), T1-Weighted Magnetic Resonance Imaging through Contrast Enhancement (T1wc), T2-Weighted Magnetic Resonance Imaging, Proton Density-weighted *Magnetic Resonance Imaging*, Fluid- Attenuated Inversion Recovery (FLAIR), etc.

Currently, a large number of BT images are being created in health centers, as physicians can't interpret these images promptly [4]. Therefore, automatic segmentation estimates have been developed. BT segmentation is the edema of regular brain tissue with regular cells, necrotic core and white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) [5]. Segmentation of irregular tissues in the segment including machine learning, medical imaging has created significant developments in the field of BT segmentation [6]. In general, most irregular BT tissues can be directly identified by the BT segmentation technique. However, precisely not all systems with reproductive partitioning effects, including instances of deformation, are resolved. BT segmentation has a huge impact on identification, monitoring, action schedules for patients as well as medical tests.

This paper shows detailed review on the preprocessing and traditional transition to machine learning and deep learning-based automatic SS techniques of magnetic resonance imaging.

(a) T1-weighted MRI   (b) T2-weighted MRI

(c) FLAIR   (d) FLAIR with contrast
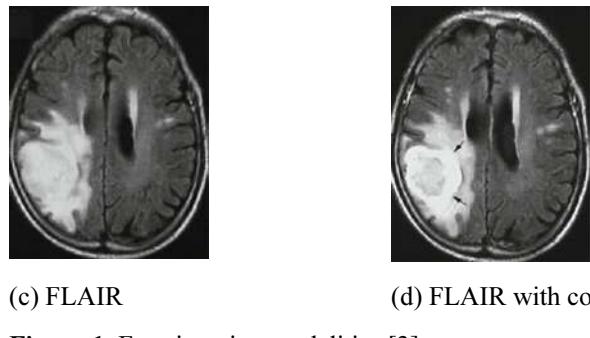
Figure 1. Four imaging modalities [3]

2.   Preprocessing

Before the *Brain Tumor Segmentation* (*BTS)* methods, the *Magnetic Resonance Imaging* preprocessing actions are initiated as it is straightforwardly relayed to the superiority of the segmentation outcomes [7]. Generally, the untreated Magnetic Resonance Imaging images require preprocessing to recognize BTS purposes. Such pre-processing functions consist of filtering, skull-stripping, etc, as well as have straight effect on the effects of BTS. Image filtering is a typical preprocessing part for Magnetic Resonance Imaging [8].

The term noise is used to denote any random attenuation of a signal and is related to images. Noise is sent by image so noise can come through image acquirement or could result from a noisy communication channel over which the image is sent [7]. Either way, the denoising job is to utilize the knowledge that have on the statistical structure of images to eliminate the result of the noise as well as probable. One of the main difficulties facing in the Image Processing (IP) is image denoising. There are several filtering techniques existing. The wide classification of different image denoising techniques is provided in the Figure 2.

In [9], provides a behavioral analysis of the *Anisotropic Diffusion Model (ADM)* of the authors Perona and Malik. The main proposition was to tell the equation from the difficulty of some optimization to find the reality of the unique global minimum reality to maintain the existence of some dense global minima. Also, the smooth, as well as the edge enhancing method of anisotropic diffusion, is shown during the eigenvalue disintegration of the diffusion equation.

Two wavelet-domain filtering (WDF) techniques were given to reduce noise in magnetic resonance imaging [10]. The wavelet coefficient with the corresponding thresholding technique [11] is not different from the previously suggested technique for magnetic resonance imaging evaluation based on soft thresholding [12]. To handle this extra difficult situation, a WDF was obtained that works on an image of square magnitude.

The new *Noise-Adaptive Method (NAM)* verified the reversal of certain filters when incredible unreliable noise was occurring in magnetic resonance imaging [13]. Most filtering magnetic resonance imaging relies on the same distribution over noise distribution. Since this prediction is not true, the resulting filtering suit is suboptimal. This is the case with magnetic resonance imaging with local unrealistic noise levels, for

example, obtained by parallel imaging. An imaginary technique where knowledge of the noise level of a local image is used to reduce the power of the filter. This knowledge is obtained involuntarily from images through new local sound assessment technology. This technique was standardized as well as evaluated with sample nonlocal i.e. filters with authentic magnetic resonance imaging data that saw enhanced rewards in all conditions.

The *Independent Component Analysis* (ICA) has proven immense promise for extracting feature information, basically deciding the building blocks of each provided data [14]. ICA is a statistical method that seeks to find examples of empirical data as the components are as independent as possible. The ICA is applied to the image data to identify the morphology in the image and the ICA gives a representation. That representation can be considered the primary structure in this data. ICA is a highly novel data-analysis tool that has a huge promise to judge the relevant specific features in high-dimensional data.
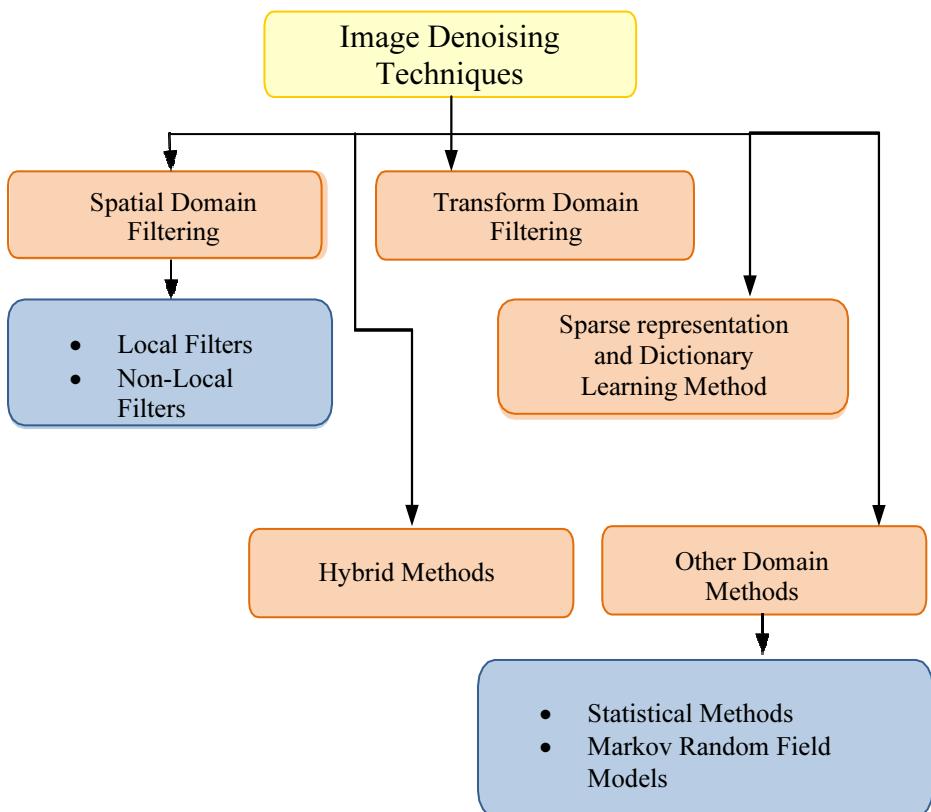
Figure 2. Image Denoising Techniques

It is written in the literature that noise can be reduced in magnetic resonance images by using numerous techniques. The purpose of the paper [15] is to critique modern de-

noising algorithms based on statistical assumptions and their ability to improve the outcome of brain tumor segmentation. ADM is currently the most widely accepted technique for De-noised of BTs magnetic resonance imaging. Although the noise of the images has decreased, it has been suggested that it has persisted permanently as well as turned into a negative result of BT segmentation.

3. Skull Stripping

SS is believed one of the dangerous preprocessing methods that should help in the accurate identification of BT. SS also reduces the likelihood of incorrect classification of brain tissues for periods of the division with incompatible tissues. Accurate brain extraction improves the likelihood of self-diagnosis of numerous neurological disorders such as dementia, as well as schizophrenia. The main idea of SS is that it removes non-brain tissue-dura substance just like the external blood vessels, as well as is lost only in the area of the brain. These non-brain tissues enhance the computational effectiveness of various neuroimaging algorithms. *SS* is categorized into various groups as given below.

- Manual,

- Semi-Automated,

- Automated Methods.

The advances increased for SS or brain *Magnetic Resonance Imaging* extraction has been classified into three main Classes.

1) **Conventional SS (CSS)**: Mathematical operations.

2) **Machine-learning-based SS (MLSS)**: Fuzzy Logic (FL), Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), Bayesian Classifiers (BCs)

3) **Deep-learning-based SS (DLSS)**: use of *Convolution Neural Network* (CNN), *Input Parameters* (IPs) and *Hidden Layers* (HLs).

CSS advances use traditional or regular methods to reach SS. This section covers the content of the usual IP techniques with most SS techniques. The classification of CCS techniques is shown in Table. 1.

Table 1. Conventional skull stripping techniques

Types	Methodology	Discussion	Results
Deformable Surface–Based Skull Stripping Methods (DSSS) [16].	The (DSSS) method primarily says the surface model with frequently distorts the surface in anticipation of it create the best possible result.	This technique does not involve any pre-processing method.	It is extremely robust as well as precise.

Mathematical Morphological Operation–based Skull Stripping Methods (MMOSS)[17].	Effort through Thresholding with morphological erosion and dilation actions in sequence.	One of the drawbacks is that they are parameter needy for example edge constant, diffusion iteration, diffusion, as well as erosion size.	Average performance of method 94.3%.
Intensity-Based Skull Stripping (IbSS) Methods [18].	IbSS advances on modeling intensity sharing of brain Magnetic Resonance Imaging utilized for threshold categorization.	IBSS techniques are responsible for accelerating their smoothness; However, their accuracy can vary in different datasets with incredible levels of image resolution, sound and artwork.	Significantly progress the clinical efficiency.
Template-Based Skull Stripping (TSS) Methods. [19].	TSS or Atlas-based approaches of SS rely on pattern or atlas on Magnetic Resonance Imaging of brain to divide them from non-brain tissues. Making an early estimate for the brain mask (BM), the BM border is segmented once more by a classifier, which improves the last result precision.	Fiducially marker registration based on the maxillary template is accurate enough for image-directed surgery at the base of the ancestral skull, but not for the lateral skull base.	TSS is a precise however noninvasive registration technique for anterior skull base surgical treatment.
Hybrid-Based Skull Stripping (HSS) Methods [20].	The mixture of further than one technique from earlier existing SS techniques.	HSS merges various techniques of SS algorithms that compose them to work jointly to attain an improved solution to a difficulty.	Shows high accuracy.

To get practical results, CSS methods need to be modified in different numerical parameters based on the dataset, where MLSS methods are increasing to get more successful results. The classification of MLSS techniques is shown in Table. 2. Deep learning (DL) with IP and HL is done by CNN. In an in-depth study of DL, as with normal NN, each input of magnetic resonance imaging exceeds the sequence of convolution layers. The different skull stripping algorithms based on CNN are proposed in [21] – [30]. Therefore, it is a challenging task to create SS every time the brain images have different contrasts, scans quality, and intensity. Different SS algorithms are proposed namely manual, semi- automatic, and automated algorithms. Automatic SS greatly improves the accuracy with effectiveness of neuroimaging

algorithms. The results of novel DL based skull stripping algorithms are more accurate and precise than usual presented methods.

Table 2. Machine Learning Based skull stripping techniques

Types	Methodology	Discussion	Results
Model-Based Level Set [31].	Based on the intensity dissimilarity of various brain divisions with curvatures of the brain planes.	Entailed postoperative information to assess the compassion of the system to abnormal structure.	Robust SS outcomes, creation it a hopeful device for utilize in large, multi- institutional, population-based neuroimaging learning.
Fuzzy Logic[32].	Approximate the intensity allocations by the utilize of a priori information depending leading the Bayesian organization and Gaussian mixture form.The previous information is explained by fuzzy membership functions.	Require prior knowledge.	The mean sensitivity is 98.84 %, with false- positive rate was 1.21%.
Support Vector Machine with local and global features [33].	An arrangement of local with global knowledge is utilized. The concluding is required to differentiate between bones moreover air.	Arrange of a little minutes which is capable.	In an association with two additional techniques, one based on MMOSS and the other on DRSS, the SVMSS was found constantly improved segmentations.
Principal Component Analysis-based model [34] .	Common tissue form is incarcerated by PCA pathologies are incarcerated by an entire variation term, the skull with neighboring tissue is incarcerated by a lightly term.	Require of formulations for various image series otherwise modalities.	Best median (97.11) and mean (96.88) Dice scores over all datasets.
FLAIR [35] .	Effectively able to achieve using a random forest of trees classifier to decide with high precision with rapidity whether a pixel is element of the brain tissue or else not.	Superior ventricle and skull size unpredictability.	An extremely superior mean dice score of 97.02%.

4.   Conclusion

Currently, a large number of BT images are being created in health centers, as physicians can't interpret these images promptly. Therefore, automatic brain tumor diagnosis has been developed. This paper gives detailed review on the preprocessing and traditional transition to machine learning and deep learning- based automatic skull Stripping techniques of magnetic resonance imaging. The preprocessing process with the up-to-date methods of magnetic resonance imaging is reviewed. The Magnetic Resonance Imaging preprocessing actions are initiated as it is straightforwardly relayed to the superiority of the segmentation outcomes. Different SS algorithms are proposed namely manual, semi-automatic, and automated algorithms. Automatic SS greatly improves the accuracy with effectiveness of neuroimaging algorithms. The results of novel DL based skull stripping algorithms are more accurate and precise than usual presented methods.

5.   References

[1] H. Chen, Z. Qin, Y. Ding, L. Tian, and Z. Qin, "Brain tumor segmentation with deep convolutional symmetric neural network," *Neurocomputing*, vol. 392, no. xxxx, pp. 305–313, 2020.

[2] J. Y. Zhang *et al.*, "Quantitative Improvement in Brain Tumor MRI Through Structured Reporting (BT-RADS)," *Acad. Radiol.*, pp. 1–5, 2019.

A. Drevelegas, "Imaging of brain tumors with histological correlations," *Imaging Brain Tumors with Histol. Correl.*, pp. 1–432, 2011.

[3] N. F. Interviews *et al.*, "Significant Moments the Chair," no. 25, 2020.

[4] M. J. Pallod, "Survey on Brain Tumor Detection," pp. 1–6, 2020.

[5] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, "Medical image segmentation based on U-Net: A review," *J. Imaging Sci. Technol.*, vol. 64, no. 2, pp. 1–12.

[6] Z. Ahmed, K. Mohamed, S. Zeeshan, and X. Q. Dong, "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine," *Database (Oxford).*, vol. 2020, pp. 1–35, 2020.

[7] J. Li *et al.*, "Automatic fetal brain extraction from 2D in utero fetal MRI slices using deep neural network," *Neurocomputing*, vol. 378, no. xxxx, pp. 335–349, 2020.

[8] Y. L. You, W. Xu, A. Tannenbaum, and M. Kaveh, "Behavioral analysis of anisotropic diffusion in image processing," *IEEE Trans. Image Process.*, vol. 5, no. 11, pp. 1539–1553, 1996.

[9] R. D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Trans. Image Process.*, vol. 8, no. 10, pp. 1408–1419, 1999.

[10] M. Hilton, T. Ogden, D. Hattery, G. Eden, and B. Jawerth, "Wavelet Denoising of Functional MRI Data," *WAVELETS Med. Biol.*, no. December 2014, pp. 93–114, 2019.

[11] C. S. Anand and J. S. Sahambi, "MRI denoising using bilateral filter in redundant wavelet domain," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, 2008, doi: 10.1109/TENCON.2008.4766742.

[12] J. V. Manjón, P. Coupé, L. Martí-Bonmatí, D. L. Collins, and M. Robles, "Adaptive non-local means denoising of MR images with spatially varying noise levels," *J. Magn. Reson. Imaging*, vol. 31, no. 1, pp. 192–203, 2010.

[13] P. Hoyer, "Independent component analysis in image denoising," *Master's Thesis, Helsinki Univ. Technol.*, no. April 1999, 1999.

[14] Diaz, P. Boulanger, R. Greiner, and A. Murtha, "A critical review of the effects of de-noising algorithms on MRI brain tumor segmentation," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 3934–3937, 2011.

[15] S. M. Smith, "Fast robust automated brain extraction," *Hum. Brain Mapp.*, vol. 17, no. 3, pp. 143–155, 2002.

A. E. K. Isselmou, S. Zhang, and G. Xu, "A Novel Approach for Brain Tumor Detection Using MRI Images," *J. Biomed. Sci. Eng.*, vol. 09, no. 10, pp. 44–52, 2016.

[16] Z. Yang *et al.*, "A web-based brain metastases segmentation and labeling platform for stereotactic radiosurgery," *Med. Phys.*, pp. 1–14, 2020, doi: 10.1002/mp.14201.

[17] G. Eggers and J. Mühling, "Template-based registration for image-guided skull base surgery," *Otolaryngol. - Head Neck Surg.*, vol. 136, no. 6, pp. 907–913, 2007, doi: 10.1016/j.otohns.2006.12.021.

[18] S. S. and C. S. Pawar, "An Automatic Brain Tumor Detection and Segmentation using Hybrid Method," *Int. J. Appl. Inf. Syst.*, vol. 11, no. 9, pp. 6–11, 2017, doi: 10.5120/ijais2017451641.

[19] Zhuang, J. Yang, L. Gu, and N. Dvornek, "Shelfnet for fast semantic segmentation," *Proc. - 2019 Int. Conf. Comput. Vis. Work. ICCVW 2019*, pp. 847–856, 2019, doi: 10.1109/ICCVW.2019.00113.

[20] H. Chen, Q. Dou, L. Yu, J. Qin, and P. A. Heng, "VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images," *Neuroimage*, vol. 170, no. April 2017, pp. 446–455, 2018, doi: 10.1016/j.neuroimage.2017.04.041.

[21] Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, 2017, doi: 10.1016/j.media.2016.10.004.

[22] O. Lucena, R. Souza, L. Rittner, R. Frayne, and R. Lotufo, "Silver standard masks for data augmentation applied to deep-learning-based skull- stripping," *Proc. - Int. Symp. Biomed. Imaging*, vol. 2018-April, no. Isbi, pp. 1114–1117, 2018, doi: 10.1109/ISBI.2018.8363766.

[23] O. Lucena, R. Souza, L. Rittner, R. Frayne, and R. Lotufo, "Convolutional neural networks for skull-stripping in brain MR imaging using silver standard masks," *Artif. Intell. Med.*, vol. 98, no. August 2018, pp. 48–58, 2019, doi: 10.1016/j.artmed.2019.06.008.

[24] O. Lucena, R. Souza, L. Rittner, R. Frayne, and R. Lotufo, "Convolutional neural networks for skull-stripping in brain MR imaging using silver standard masks," *Artif. Intell. Med.*, vol. 98, pp. 48–58, 2019, doi: 10.1016/j.artmed.2019.06.008.

[25] N. Imaging and L. Angeles, "Auto-context and Its Application to High- level Vision Tasks and 3D Brain Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1744–1757, 2010, doi: 10.1109/CVPR.2008.4587436.

[26] N. H. M. Duy, N. M. Duy, M. T. N. Truong, P. T. Bao, and N. T. Binh, "Accurate brain extraction using Active Shape Model and Convolutional Neural Networks," 2018.

[27] B. Yilmaz, A. Durdu, and G. D. Emlik, "A new method for skull stripping in brain MRI using multistable cellular neural networks," *Neural Comput. Appl.*, vol. 29, no. 8, pp. 79–95, 2018.

[28] P. Ganaye and I. C. Umr, *Semi-supervised Learning for Segmentation Under Semantic Constraint Pierre-Antoine*, vol. 1, no. August 2019. Springer International Publishing, 2018.

[29] H. Zhuang, D. J. Valentino, and A. W. Toga, "Skull-stripping magnetic resonance brain images using a model-based level set," *Neuroimage*, vol. 32, no. 1, pp. 79–92, 2006, doi: 10.1016/j.neuroimage.2006.03.019.

[30] S. Kobashi *et al.*, "Fuzzy-ASM Based Automated Skull Stripping Method from Infantile Brain MR Images," pp. 632–632, 2008.

[31] J. Sjolund, A. E. Jarlideni, M. Andersson, H. Knutsson, and H. Nordstrom, "Skull Segmentation in MRI by a Support Vector Machine Combining Local and Global Features," *Proc. - Int. Conf. Pattern Recognit.*, pp. 3274– 3279, 2014.

[32] X. Han et al., "Brain extraction from normal and pathological images: A joint PCA/Image-Reconstruction approach," Neuroimage, vol. 176, pp. 431–445, 2018.

The Study of Usage of Hyperledger Fabric in Agricultural Ecommerce

Vivek Ray [a,1], Anmol Singh [a], Mayank Singh [a], Rishi Singh [a], and Dr. Shushila Palwe [b]

[a] Student, Dr. Vishwanath Karad MIT World Peace University, Pune 411047, India
[b] Assistant Professor, School of CET, Dr. Vishwanath Karad MIT World Peace University, Pune 411047, India

Abstract. With the advent of COVID-19, the ecommerce industry in India has seen an inflection point with rise in the demand across all the segments of the industry. Much ecommerce has been started to cater to the supply and demand mismatch in agricultural goods front. In these times, Blockchain is seen to create a trust bridge for all the stakeholder to transact through goods and supplies with minimal risks involved. But, to use this technology the technical, economical and a scalable approach to this technology is still a very prominent requirement for mass adoption. Permissionless Blockchain build on Proof of Work consensus protocol cannot be used due to their slow speed, low scalability, and high energy consumption for network functioning. This paper concludes with the possibility of using a Permissioned blockchain such as Hyperledger Fabric to not only solve the several underlying issues to facilitate efficiency in the Ecommerce architecture, with critical view on the mass adoption.

Keywords. Ecommerce, Blockchain, Permissionless Blockchain, Proof of Work, Permissioned Blockchain, Hyperledger Fabric.

1. Introduction

Ecommerce is a term given to illustrate electronic commerce, which refers to goods and supplies being sold over the internet through channels such as websites and mobile applications. Due to low commutability in COVID-19, the inter-net landscape through mobile applications, experienced a surge in the ecommerce industry because of high demand of goods [1]. But the growth of ecommerce business in agriculture sector is still not matching with all the stakeholder, as the profits are highly concentrated towards the middle parties involved in the whole distribution as well as marketing of the goods. This challenge can be fixed by creating an infrastructure of ecommerce with trust reinstated to all the stakeholders with the help of use of digital distributed ledger in blockchain. A blockchain is a platform of decentralized computation and information sharing platform that enables multiple authoritative domains, who do not trust each other, to cooperate, coordinate and collaborate in a rational decision-making process. Individual transaction data files, called as blocks, are managed through specific software

---

[1] Vivek Ray, Dr. Vishwanath Karad MIT World Peace University, Pune
Email: vivekray903@gmail.com.

platforms that allow the data to be transmitted, processed, stored, and represented in human readable form. This digital ledger forms the foundation for trust in a trustless digital environment. The usage of blockchain network can be seen through its practical application in several cryptocurrencies [2] which are its most dominant usage. In agricultural ecommerce, the blockchain offers to help in different operations such as in traceability, supervision, and management of goods.

The initial blockchain platform such as Bitcoin and Ethereum were permissionless blockchain [3] as they were open to participation by nature and were usually used in application involving multidisciplinary use cases that required trustworthy and decentralized execution of business logic. But with the benefits, permissionless block-chain also suffered from several drawbacks such as (i) low performance because of high throughput of transaction requests, and (ii)confidentiality issues. These drawbacks cause decrease in efficiency and hence, create problem for the utilization of public blockchains by large businesses since these problem prevent the integration of enterprise grade systems that usually require secure, high performance applications that guarantee reliable and consistent transactions to a volatile public network. Therefore, permissioned blockchains were developed for avoiding these shortcomings.

Permissioned blockchain are blockchain networking systems in which the participation in the network and in the transactions of the network required explicit authorization from the administrators of the network. Permissioned blockchains are suitable for competing enterprises which in out cases revolves around the various Ecommerce platform that are, nonetheless, willing to engage in collaborative processes without employing third parties, such as notaries, or centralized settlement networks. There are several factors to prove that permissioned blockchains provide various enhancements over their permissionless counterparts, such as: (i) the participation in the blockchain is limited to a specific group of users that requires explicit system reconfiguration to be modified, permissioned blockchains use Byzantine Fault Tolerant (BFT) protocols, which are a better alternative in terms of transaction latency and throughput. (ii) Furthermore, permissioned blockchains are generally better in terms of data privacy and data sharing as sensitive transaction can be secured from access,(iii) Finally, most permissioned blockchain systems are directed toward transaction finality as well as other transactional properties which are not focused by public blockchain.

2. Literature Review

In the article of Kamilaris and others [4], the solution suggested through blockchain were: (i)traceability in value chains, (ii) participation of small farmers, (iii)fair pricing through whole value chain, (iv)more informed consumer purchasing decision, (v)reduced transaction fees and less dependence on intermediaries, and (vi) more transparent transaction and less frauds.

As suggested by V.S. Yadav and A. R. Singh [5], blockchain based mobile app have the ability to solve the above listed problems of agriculture, however the proposed solution does also have the problems as suggested by Kamilaris[4]: (i) High uncertainties and market volatility of transaction fees, (ii) No regulation in place , (iii)Scalability issues due to latency of transactions,(iv)Privacy issues, and (v)decline of cryptocurrencies prices may lead to decline in growth of block-chain network or even making the whole network out of order.

As Hyperledger Fabric fixes these issues with a predominant permissioned architecture which not only suits the commercial needs but also the business privacy involved, thus it plays a better role for the supported blockchain requirements involved for Ecommerce needs.

2.1. Hyperledger Fabric Overview

Hyperledger Fabric, an open-source project initiated by Linux Foundation and IBM, is the modular blockchain framework which provides de facto standard for enterprise blockchain platforms. It has provided a foundation for developing enterprise-grade applications and industry solutions, the open, modular architecture that uses plug-and-play components to be applied for several use cases [6].

Fabric executes distributed applications written in general-purpose programming languages like Go, Java or Node.js. There are three stages in Fabric Blockchain operation: Execution, Ordering and Validation [7]. Fabric Network securely monitors the execution operation in such a way that the data structure can only be updated without any inherent transaction fees as it has no cryptocurrency built in for the network. Fabric consists of two main parts:

In Hyperledger Fabric, smart contract is referred as chain code, which is program code that implements the application logic and runs during the execution phase. It is one of the integral parts of the distributed application hosted on Fabric Blockchain.

An endorsement policy is a static library for transaction validation during validation phase in Fabric Blockchain, which is instructed by the chain code. As the Fabric Blockchain is permissioned network, so only the designated administrators have the permission to modify system policies that are used in during node designation by the Membership Service Provider (MSP).

.A Fabric blockchain consists of a set of nodes that form a network (Fig. 1). All nodes that participate in the network have an identity, as provided by a modular MSP. MSP has three roles which handle the overall architecture of the network:

- Clients: These are the applications which helps the person to initiate the transactions on network. They have SDK packages which is connection towards the Chaincodes hosted on Fabric Blockchain.
- Peers(P): Peers maintain the state of the network and the copy of ledger. All peers commit blocks to the distributed ledger. In Fabric Blockchain there are two special types of peers:
  - Endorsing Peer: Peers with chaincode installed which stores transaction of chaincodes in an isolated chaincode containers and prepares proposal based on results of chaincode. They produce endorsement signature and simulation of transaction. This signature is sent back to client.
  - Committing Peers: Peers with only purpose of maintaining and updating their full ledger of records. They do not have any association with chaincodes or their functions.
- Ordering Service Nodes (OSN): The ordering service nodes accept verified transaction, orders them into a block, and delivers the blocks to the committing peers.
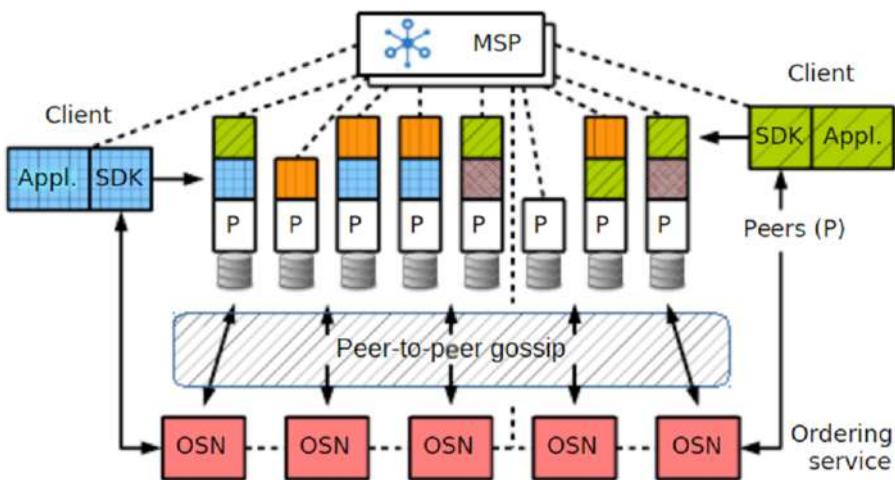
Figure 1. A Hyperledger Fabric Blockchain Network

All these components form the application structure required to operate the overall business processes which will write the data onto the Fabric Blockchain. The process starts with the client sending a transaction to Endorsing peer. Endorsing peers sends back a signature on the valid transaction. Figure 1 this signature transaction is sent to OSN where it is verified. After verification, this transaction is forwarded from OSN to all the committing peers via Peer-to-Peer gossip. This gossip enables the transaction between two parties to be propagated to all the committing peers for maintaining the record of transaction in distributed ledgers of Fabric Blockchain Network.

3. Proposed Model for Ecommerce

Using the Hyperledger Fabric architecture, the proposed model uses chaincode which is a container for deploying Smart contract on Fabric network integrated with mobile application running on smartphones. One or several smart contracts for different functionalities can be defined within a chaincode involved in Ecommerce. Every smart contract has a name that is uniquely identified in a chaincode. Chaincode is executed within a secured Docker container isolated from the endorsing peer process. It initializes and manages ledger state through transactions submitted by application. The chaincode can address the smart contract for:

- **Authentication:** This smart contract can address to the issue of authenticating the seller through pre-defined channel as the regulatory governmental bodies in-scribe. This insures the services reachability to the marginalized/non-marginalized farmers with the required Consumer demand.
- **Traceability:** The factor of traceability is based on the process of the goods transaction as well as the risk involved in case of damaged/spoiled goods. The logistical approach of Hyperledger Fabric in the first scenarios is based on the parties being exchanging the goods on a step and creating changes in the network based on the future step till the goods reaches to the final stakeholder.

In the second scenario, since the data obtained from authentication makes the process of origin of the goods involved traceable, in case of any food hazard involved can further reported to the regulatory.

- **Transaction Records:** The transaction record of the mutual exchange of goods with the financial data can be further pushed to the smart contract which would deal with the data being added to the chain of records in the ledger history of both the parties. This would make both the parties being adhered to the goods being exchanges between them as per the acknowledged terms in all aspect.

4. Limitations

The Agricultural Ecommerce platform cater to the demand of the buyer as well as supply of the farmer. However, the Fabric platform's sustainability also depends on the number as well as the volume of transactions involved. This factor makes the usability of this technology being more concentrated toward bigger corporation or ecommerce platforms where transaction is of higher volume or higher value. The correlation of using more transparency in the platform for having higher trust is debatable based on notions of the entire end stakeholder (Farmers and Buyer) in the whole value chain. Thus, for transparency on the transaction level, where both the parties are assured of successful transaction through application is the prime priority. Blockchain can only add more value and trust if already existing transaction chain is successfully implemented which can make Fabric network optimal for use.

5. Result

On the aspects of the operational as well as financial efficiency, the permissioned blockchain could cater all the functionalities of permissionless blockchain without any issue in terms of transactional latency as well as the cost per transaction. Privacy and transparency are equally important balancing outlook for an ecommerce which depends on the platform developers as to what parameter of a transaction needed to be declared without having a breach in the involved trade information. As per the historical volumes in Agricultural Ecommerce of India, permissioned blockchain could scale to the needs, with providing all the essential required functionalities of a required immutable ledger for the stakeholders in the supply and demand chain. The proposed model distributes workloads of the functionalities involving the authentication, traceability, and transaction record management. These systematic processes can change based on the geographical as well as applicational functionalities to be achieved for stakeholders.

The systematic processes can be explained by the Figure 2 is an example of a vegetable cargo, which is transported from buyer to seller through four steps: placement of order with successful monetary transaction, its acceptance from seller, its delivery and successfully reception from buyer's end. These changes are made to the block of data in Fabric which can only get updated but in its entirety, transactions cannot be deleted. Thus, the integrity of whole transaction is maintained without any

role of middle party in the entire process except the optimal working of the systems involved.
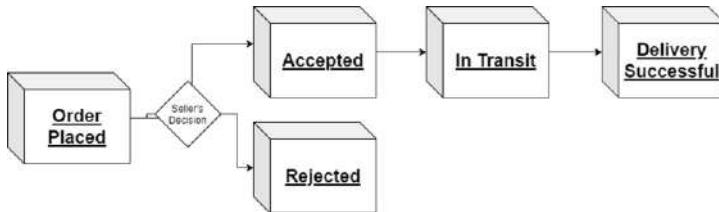

Figure 2. Block changes in Fabric on "Status" variable after every step in Agricultural Ecommerce

6. Conclusion

The proposed model could create a systemic channel for the Ecommerce platform to exchange goods between buyer and farmer. However, these outcomes are also based on the foundational foothold of a platform before its usability of Hyperledger Fabric Blockchain network. The simplified Graphical User Interface as well as dominancy of the Ecommerce application in the region will help this feature towards being accepted and Fabric being mass adopted in other sectors of Ecommerce.

References

[1] Shrey Nougarahiya, Gaurav Shetty, Dheeraj Mandloi, F.: A Review of E – Commerce in India: The Past, Present, and the Future, In: Research Review International Journal of Multidisciplinary. Volume 06 Issue 03, 12-22 (2021).

[2] David Vidal-Tomás,T.: Transitions in the cryptocurrency market during the COVID-19 pandemic: A network analysis, In.: Finance Research Letters (2021)

[3] G. Falazi, M. Hahn, U. Breitenbücher, F. Leymann and V. Yussupov, T.: Process-Based Composition of Permissioned and Permissionless Blockchain Smart Contracts, In.: IEEE 23rd International Enterprise Distributed Object Computing Conference (EDOC), Paris, France, pp. 77-87(2019).

[4] Andreas Kamilaris, Agusti Fonts, Francesc X. Prenafeta-Boldú, F.: The rise of blockchain technology in agriculture and food supply chains, T.:Trends in Food Science & Technology,Volume 91, Pages 640-652(2019)

[5] Vinay Surendra Yadav, and A. R. Singh, , F.: Use of blockchain to solve select issues of Indian farmers, In: AIP Conference Proceedings 2148, 030050 (2019)

[6] IBM (What is Hyperledger Fabric?), https://www.ibm.com/in-en/topics/hyperledger,last accessed 2021/11/04.

[7] Elli Androulaki, Artem Barger, Vita Bortnikov, Christian Cachin, Konstantinos Christidis, Angelo De Caro, David Enyeart, Christopher Ferris, Gennady Laventman, Yacov Manevich, Srinivasan Muralidharan, Chet Murthy, Binh Nguyen, Manish Sethi, Gari Singh, Keith Smith, Alessandro Sorniotti, Chrysoula Stathakopoulou, Marko Vukolić, Sharon Weed Cocco, and Jason Yellick. T.: Hyperledger fabric: a distributed operating system for permissioned blockchains. In.: Proceedings of the Thirteenth EuroSys Conference. Association for Computing Machinery, New York, NY, USA, Article 30, 1–15(2018)

Hybrid Customer-Centric Sales Forecasting Model Using AI ML Approaches

Sanket Londhe [a,1], Sushila Palwe [a]

[a]*Department of Computer Science and Engineering, MIT World Peace University, Pune, MH*

Abstract. Business Intelligence is a process of preparing, analyzing, presenting, and maintaining the data to gain insights for the decision-makers to make informed decisions. While there are many approaches to predict the growth based on the sales figures a very few consider the influence of customer data on the forecasting and the relevance of the same while making the predictions. So, in this study, we will look at some of the existing techniques used so far to make predictions and studies used to understand the customer data. With the analysis, we shall try to devise a hybrid approach to the traditional sales prediction, which would include a customer-centric data analysis. We shall look at some of the techniques which are traditionally used in Market Basket analysis and at the same time look at the techniques like classification, segmentation, regression, etc. to get a perception of the impact of customer data on sales forecasting. We shall highlight all the pros and cons of the algorithms and try to come up with an intelligent approach that would give accurate results

Keywords. Business Intelligence, Sales Forecasting, Customer Analysis, Classifiers, Regression, Segmentation, ARIMA, LSTM, RFM

1. Introduction

Business intelligence (BI) is an analytics-driven process that combines analytics and data processing techniques along with data visualization tools and best practices to help organizations to make data-driven decisions. The refined processes are supported by data to make a case. In practice, data-driven operations eliminate inefficiencies, and quickly adapt to market or supply changes. [1,6] These processes include: Data mining techniques and preparation; Knowledge discovery and transfer by reports; Benchmarking to Gain actionable insights; Querying and Statistical analysis prior to Visual analysis

Analytics experts use BI concepts and techniques to make informed decisions using historical data and its impact on the present modes of operations. Using various data visualization techniques analysts gain insights and observe patterns to ensure the growth of the organizations by taking relevant actions quickly. Today BI is used not only to increase sales but also to ensure the security of the data, understand consumer behavior, devise marketing strategies, and much more. [1]

AI and ML approaches contribute to BI strategy by integrating business insights into the strategy.

---

[1]Sanket Londhe, Department of Computer Science and Engineering, MIT World Peace University

Email: sanketlondhe99@gmail.com.

As companies use more and more data for their business operations, efforts would be made for a shared data strategy and collaborative data projects. A Business Intelligence Framework becomes a critical part of the BI Strategy across the organization.[2]

In this paper, we will look at some of the work done in the field of Customer Analytics and Sales forecasting and see various techniques used in their respective fields. Based on the studies we will come up with an approach that will be a combination of the said concept and compliant with the Business Intelligence concept.

2. Related Work

Customer analysis is a process having huge importance at different stages of a business plan. It is very important to know your product and the target audience who may want or benefit from your product [3,5]. An article from Forbes states that 81% of businesses are dependent on customer analytics insights to improve their understanding of the customers [4].

The approach used in [9] uses various models like Logistic Regression, Linear SVM, Decision Trees, Random Forest, Quadratic Discriminant Analysis, and Neural Networks to estimate sales using customer classification. For sales forecasting the authors in [9] used various regression trees and Ridge Regression. These classifiers can be implemented using the python Scikit-Learn library [13].

Another study [10] uses the RFM (Recency, Frequency & Monetary Value) model for segmentation using the k-means algorithm to understand customer purchase patterns and behavior.

Time-series forecasting is one of the fundamental concepts when it comes to making predictions and analyzing trends based on historic values [7]. Since the intent of this survey is to forecast sales, the data we are considering is traditionally non-stationary in nature. ARIMA (Autoregressive Integrated Moving Average) is one such commonly used technique used to analyze non-stationary data. The study in [11] emphasizes the benefits of ANNs in various fields like finance, engineering, and others where time series forecasting holds critical importance.

We intend to define our model to work on data of a smaller size hence the work done in [11] makes sense and they have proposed their model based on RNN, ARIMA, and part sentiment analysis. They claim their model works well in the short term without compromising the results. The focus of the study in [11] is to assess the consumer perception about the quality of the product and whether they would be interested to purchase the same or not.

Authors in [11] used MAPE (Mean Absolute Percentage Error) to measure the accuracy of the forecast based on the ARIMA-RNN model. MAPE is used as it compares different models and signifies if one is better than the other based on the deviation of predicted values from the observed values. The smaller is the MAPE, the more accurate is the model.

$$MAPE = \left(\frac{| (Actual\ Value\ -\ Predicted\ Value) |}{Actual\ Value} \right) * 100$$

Another study [12] uses a combination of ARIMA (Autoregressive Integrated Moving Average), LSTM (Long short-term memory) models, and concepts of wavelet de-noising on Hydrological time series data. Hydrological time series data is data

related to water flow and mainly used in water management studies. This data is not stationary in nature and also contains lots of noise making it very difficult to forecast. The proposed model in [12] first uses the wavelet de-noising to eliminate the noise from the data. Then this de-noised data is fit to the model using ARIMA and subsequently forecasted. The result if ARIMA produces some residuals which are used to train the LSTM network. Results of the LSTM network are then used to fine-tune the forecasts by the ARIMA model.

3. Mathematical Model for K-Means Clustering

K-means clustering is a method of forming a specific number of clusters (k) represented by their centroids. Based on the k (no. of clusters required) centroids, each point in the data is segregated into a specified cluster. Clustering is done based on the distance between the data point and the centroid. We do this until all the data points are part of a cluster.

Input: k (number of clusters), D (set of lift ratios)
Output: a set of k subgroups (clusters)
Method: Randomly choose k as initial centroids;
Repeat: Assign each object to the cluster centroid based on the similarity features i.e., the mean value of objects in the cluster and update the cluster means
Until no change;

3.1. ***Data in Euclidean Space:*** *The quality of a cluster is measured using Sum of Squared Error (SSE). It is nothing but the Summation of the Euclidean distances between the centroid of the cluster and the data points. It is also known as scatter.*

3.2. ***Cohesion****: The main aim is to increase the similarity between the centroid and the data points in the cluster. For the objective function the centroid is the mean. This ensures that the algorithm is not limited to the Euclidean space. This is called Cohesion.*

3.3. ***Time and Space Complexity*** *- The space complexity of the K-means algorithm is BigO((d+K)n), where d are the data points and n are the attributes. Time taken to run the algorithm is linear in terms of data points in the cluster. The Time Complexity is BigO(i*K*d*n), where i are the no. of iterations.*

3.4. ***Evaluation Criteria*** *- For clusters K=3 and K=5 in [10] we compare the Recency values with Frequency and Monetary values of the sales. Comparing the clusters, we find which set of customers have the highest RFM value.*

4. Mathematical Model for LSTM Model

A typical LSTM network contains memory blocks which are called cells. There are two states which are transferred to the cell next to it. These two states are the hidden state and the cell state. The cell state holds the main flow of data and it allows an unchanged flow of data forward. However, some minor linear transformations may

take place. A sigmoid gate is used to add or remove data from the state in the chain. A gate contains different individual weights. LSTM models avoid the long-term dependency issues. As discussed, before it uses sigmoid gates to deal with the memorizing process. [15]

An LSTM network is initialized by getting rid of the unwanted data in the first step. This Sigmoid Function decides which data to remove. It takes the result of the previous unit (h_{t-1}) at time t − 1 and the input (X_t) at time t. The forget gate (f_t) of the sigmoid function determines which of the previous outputs needs to be omitted relative to the number in C_{t-1} state. [15]

$$f_t = \sigma \ (\ W_f [\ h_{t-1}, \ X_t] + b_f)$$

where σ is the sigmoid function, W_f is the weight matrices and b_f are the bias.

The next step is to select and store data in the next input (X_t) and then update the cell state. One of the two layers decides whether or not to update (0 or 1) the new information. The other one is the tanh function which assigns weight (-1 to 1) to the value which passed by. The product of two is updated in the new cell. This result is added to the previous memory (C_{t-1}) which updates to C_t.

$$i_t = \ \sigma \ (\ W_i [\ h_{t-1}, \ X_t] + b_i),$$
$$N_t - tanh \ (\ W_n [\ h_{t-1}, \ X_t] + b_n),$$
$$C_t = C_{t-1} f_t + N_t i_t$$

where, C_{t-1}, C_t are the cell states, t − 1, t is time at cell states C_{t-1} and C_t respectively, W is the weight matrices and b is the bias of the cell state.

The last step shows the filtered version on the output (h_t) based on the state (O_t). Here initially the sigmoid function finalized the part of cell state which subsequently becomes output and then the sigmoid gate (O_t) and the result if tanh function from cell state (C_t) is multiplied. The resulting value is between -1 and 1.

$$O_t = \sigma \ (\ W_o [\ h_{t-1}, \ X_t] + b_o),$$
$$h_t = O_t \ tanh(C_t)$$

where, W_o is the weight matrices and b_o is bias of the output gate.

4.1. **Evaluation Criteria** *– The performance of the model is evaluated by statistical methods. NSE and RMSE are popularly used to evaluate predicted and observed values.* **The Nash-Sutcliffe Efficiency (NSE)** *measures the magnitude of predicted data variance compared to measured values. NSE values range from* −∞ *to 1.* **The Root Mean Square Error (RMSE)** *is the error-index that shows how closely the predicted and observed values match, based on the range of data. Lower is the RMSE better if the fit of the model and zero implies a perfect fit. An LSTM model gives highly reliable results with NSE values close to 1 and a minimal value of RMSE.*

$$NSE = \left(1 - \frac{\sum_{i=1}^{n} \ (O_i - P_i)^2}{\sum_{i=1}^{n} \ \left(O_i - \underline{O_i}\right)^2} \right) * 100 \qquad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \ (O_i - P_i)^2}$$

where, O_i is the observed discharges, P_i is the simulated discharges at time t, n is the total number of observations.

5. Results and Discussion

The study in [9] shows that algorithms like Linear SVM, Logistic Regression, Neural Network, AdaBoost, and Naïve Bayes showed classifier accuracy of more than 80% as shown in Table 1.

TABLE 1. CLASSIFIERS ACCURACY [9]

Classifier	Success Ratio	Classifier	Success Ratio
Linear SVM	84%	K-Nearest Neighbour	78%
Logistic Regression	84%	RBF SVM	75%
Neural Network	84%	Decision Tree	74%
AdaBoost	84%	Random Forest	74%
Naive Bayes	84%	Quad. Discriminant Analysis	43%

Similarly, work in [10] shows the RFM log on two different values of K (using k-means clustering) and presents an optimal way of getting the result as shown in Table 2. The authors in [10] took reference of the Silhouette Score (It is a metric used to calculate the goodness of a cluster. Its value ranges from -1 to 1.) to present the most optimal results.

TABLE 2. SILHOUETTE SCORE [10]

K = 3				K = 5			
#Clusters	Recency log	Frequency log	Monetary log	#Clusters	Recency log	Frequency log	Monetary log
0	161.1914	16.7619	291.8525	0	210.331	10.2652	170.5903
1	11.3732	209.3714	5316.8004	1	20.3099	95.2432	1913.8276
2	20.323	48.8775	894.3214	2	10.7796	280.0861	7456.3908
				3	81.2908	32.5154	619.0144
				4	3.8476	41.6503	708.6397
Silhouette Score for K= 3 is 0.3621				**Silhouette Score for K= 5 is 0.3491**			

The experiment carried out by authors in [11] used the Mac Revenue dataset consisting of 6757 entries. They divided the data into various batches to train and test the models. The experiments showed that on certain dates in the data ARIMA model performed better and on others RNN models. However, the proposed hybrid model outperformed them all with a minimum Average MAPE value of 3.7354 compared to ARIMA (6.3762) and RNN (9.2642) models as shown in Table 3.

TABLE 3. MODEL ACCURACY TESTING [11]

Model	Mean MAPE Value
Proposed Hybrid RNN Model	3.7354
RNN Model	9.2642
ARIMA Model	6.3762

As the study in [11], here also authors have used MAPE along with MSE (Mean Square Error) to measure the accuracy of the forecast. The experiments in [12] show that the predicted values by the proposed model based on de-noising-ARIMA-LSTM are significantly close to the observed values. Table 4. shows the results of MAPE and MSE and we can see that the values of the proposed model are fairly smaller compared to standalone ARIMA and LSTM models.

TABLE 4. RESULT COMPARISON [12]

Model	MSE	MAPE
de-noising-ARIMA-LSTM	0.0044	0.51%
ARIMA	0.0178	0.72%
LSTM	0.0049	0.56%

There is a need to understand customer behavior and apply the same not only in the marketing segment but also to accurately predict the sales and gain real insight. The research related to sales prediction is very old and is going on for years now. The traditional approaches do not fit right in the cross functional analysis. Considering the growing trend of social media, the customers are more informed of the offerings provided and they have a lot of options to choose from for any given product. Hence there is a need for the companies to incorporate customer analysis within their sales forecast which is where the traditional approaches fail.

6. Proposed Approach

We propose a hybrid approach to traditional sales forecasting, the whole process of prediction is going to be customer-centric. Figure 1 following are the steps to implement the proposed plan:

Step 1: Customer Segmentation
Pre-process and aggregate the data at the customer level. Build RFM Features for each customer. Apply K-means clustering to identify different groups (clusters) and evaluate the nature of each group. Calculate RM score and sort customers. Visualize the results and explore some key numbers.

Step 2: Segment Selection
The main aim is to identify the segments which impact the sales and not include values for the sales of average churn customers. Based on the RMF score and we would define the segments which have relevance to the sales prediction. Since Low RMF customers are likely to churn their sales value has no impact on the prediction hence we must discard them from subsequent predictions.

Step 3: Sales Prediction
Once we have removed the unwanted customer segments, we shall do the time series analysis on the remaining customers using appropriate ARIMA or LSTM model. In the end, all the independent values are consolidated to get the final predicted values.
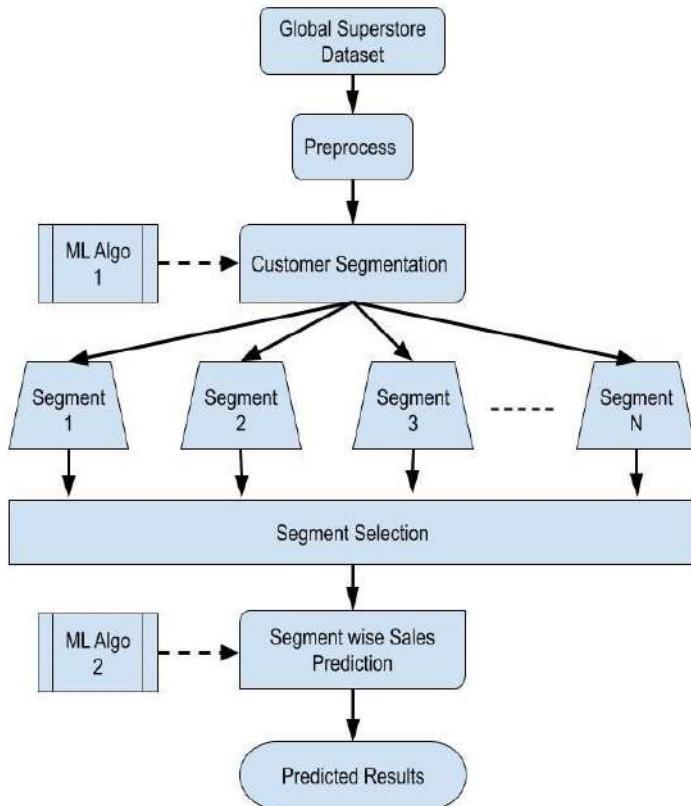
Figure 1. Proposed Model

7. Conclusion

We saw there are many fundamental studies that apply the concepts of Business Intelligence using various AI and machine learning techniques. We saw many algorithms to get concrete results. Many studies are available which presents different ways of segmenting customer behavior, but none to blend the idea for predicting sales. There is a need to understand customer behavior and apply the same not only in the marketing segment but also to accurately predict the sales and gain real insight. The survey work intends to present the analysis in a simplistic way to fuel the growth of any organization by applying the predictive approach. It will be a combination of programming, data analysis, and machine learning concepts; from the customer standpoint.

References

[1]   J. Su and Y. Tang, "Business Intelligence Revisited," 2017 Fifth International Conference on Advanced Cloud and Big Data (CBD), 2017, pp. 1-6, doi: 10.1109/CBD.2017.9.

[2] S. Atul Khedkar and S. K. Shinde, "Customer Review Analytics for Business Intelligence," 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2018, pp. 1-5, doi: 10.1109/ICCIC.2018.8782305.

[3] Customer Analysis | Alameda SBDC (2020). Available at: https://www.acsbdc.org/resources/small-business-topics/marketing/customer-analysis (Accessed: 12 December 2020).

[4] Customer Analysis | CMG Consulting (2020). Available at: https://www.cmgconsulting.com/post/what-is-customer-analysis (Accessed: 12 December 2020).

[5] McColl-Kennedy JR, Zaki M, Lemon KN, Urmetzer F, Neely A. Gaining Customer Experience Insights That Matter. Journal of Service Research. 2019;22(1):8-26. doi:10.1177/1094670518812182

[6] S. Khan, M. R. Qader, T. K and S. Abimannan, "Analysis of Business Intelligence Impact on Organizational Performance," 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), 2020, pp. 1-4, doi: 10.1109/ICDABI51230.2020.9325610.

[7] Mentzer, J.T. and Cox, J.E., JR. (1984), Familiarity, application, and performance of sales forecasting techniques. J. Forecast., 3: 27-36. https://doi.org/10.1002/for.3980030104

[8] Insights, S. (2020) Predictive Analytics: What it is and why it matters, Sas.com. Available at: https://www.sas.com/en_in/insights/analytics/predictive-analytics.html (Accessed: 14 December 2020).

[9] R. Valero-Fernandez, D. J. Collins, K. P. Lam, C. Rigby and J. Bailey, "Towards Accurate Predictions of Customer Purchasing Patterns," 2017 IEEE International Conference on Computer and Information Technology (CIT), Helsinki, 2017, pp. 157-161, doi: 10.1109/CIT.2017.58.

[10] P. Anitha, Malini M. Patil, "RFM model for customer purchase behavior using K-Means algorithm", Journal of King Saud University - Computer and Information Sciences, 2019, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2019.12.011.

[11] R. G. Hiranya Pemathilake, S. P. Karunathilake, J. L. Achira Jeewaka Shamal and G. U. Ganegoda, "Sales Forecasting Based on AutoRegressive Integrated Moving Average and Recurrent Neural Network Hybrid Model," 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Huangshan, China, 2018, pp. 27-33, doi: 10.1109/FSKD.2018.8686936.

[12] Z. Wang and Y. Lou, "Hydrological time series forecast model based on wavelet de-noising and ARIMA-LSTM," 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 2019, pp. 1697-1701, doi: 10.1109/ITNEC.2019.8729441

[13] Verhoef, P. C., & Donkers, B. (2001). Predicting customer potential value an application in the insurance industry. Decision support systems, 32(2), 189-199.

[14] K. T. Ratnayaka, K. N. Seneviratne, W. Jianguo and H. I. Arumawadu, "A Hybrid Statistical Approach for Stock Market Forecasting Based on Artificial Neural Network and ARIMA Time Series Models".

[15] Le, Xuan-Hien; Ho, Hung V.; Lee, Giha; Jung, Sungho. 2019. "Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting." Water 11, no. 7: 1387.

[16] Statistical tree-based feature vector for content-based image retrieval ,Sushila Aghav-Palwe; Dhirendra Mishra, IJCSE 2020,21(4)

A Comprehensive Assessment on IOT Devices with Data Mining Techniques

G. Vijay Kumar [a, 1], M. Sreedevi [a], Arvind Yadav [a], B. Aruna [a]

[a] *Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur-522502, India*

Abstract. Now at present development the entire world using vast variety of smart devices associated among sensors & handful of actuators. There is an enormous progress within the field of electronic communication; processing the data through devices and the bandwidth in internet technologies makes very easy to access and to interact with the variety of devices all over the whole world. There is a wide range research in the area of Internet of Things (IoT) along Cloud Technologies making to build incredible data which are creating from this type of heterogeneous environments and can be able to transform into a valuable knowledge with the help of data mining techniques. The knowledge that is generated will takes a crucial role in making intellectual decisions and also be a best possible resource management and services. In this paper we organized a comprehensive assessment on various data mining techniques engaged with small and large scale IoT applications to make the environment smart.

Keywords. Internet of Things, Sensors, Actuators, Cloud, Data Mining.

1. Introduction

At the present time all kinds of Industries are utilizing and contributing their work and doing Examination on the gadgets which we use daily and throughout the world in the IoT technology. The companies like Cisco, Microsoft, Google, and Research societies together with Internet of Things (IoT), Wireless Sensor Networks (WSN), Mobile Computing (MC), Cyber-Physical System (CPS) etc., the entire world functioning and investigate to contribute their work and curiosity to make the urban areas as Smart Environment [1-3]. In IoT, the things which we can likewise call as the Objects are its availability with Internet. We can say that the web is additionally utilized as mechanism of network and furthermore be utilized as force to run gadgets. For any type of devices, the correspondence plays a essential task in the network. The devices retrieve the information and need to share this information to different gadgets and the server farm needs a medium to do this where web will assist the information with being shared between different gadgets and the server farm. For this the gadgets ought to have a one-of-a-kind personality to recognize it in organization, it should be distinguished naturally, and ought to impart and impart the information to one another and the people. They should settle on choice without help from anyone else or follow the human orders [4].

---

[1] G. Vijay Kumar,Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Email: gvijay_73@kluniversity.in.

Presently the inquiry emerges that why would that be a great deal of examination going on IoT innovation and Industries are moving towards IoT? The research which is going on shows that urban areas will be overlaid among detecting and activation gadgets in the following 10 Years. The sensors and actuators will increment soon after quickly detecting and activating gadgets will expand threefold as much as populace of individuals. Ericsson and Cisco anticipated that "50 billion of little inserted actuators and sensors associated with the web before 2020 year with Internet of Things will make 14.4 trillion dollars of significant worth in question for businesses in the following decade" [5]. So, we can say that, a huge number of connected heterogeneous devices will form IoT environment [6-9]. The detecting gadgets which are in IoT climate produces huge amount of data. This data which is sensed, initially called as Raw data. This information is at first not introduced in Human justifiable language and should be handled. The information may have a place with a framework or any kind of utilization in the Smart IoT climate. The foundation may comprise of gadgets information just as the organization information or any sort of information system in the IoT climate. The information tangles situate with a little sort of utilization or huge kind of use. The information should be changed over into human reasonable language where Data Mining procedures will assist us with recovering the significant data from the information [10]. The installed gadgets in IoT climate may create colossal information might be to serve various applications. Here a unified middleware to improvement cycle simpler, give protection, security inside assorted applications and administrations [11].

2. Related Work

In paper [4] they have reviewed study on data mining technologies with Internet of Things and this paper mainly focused on Application of data mining techniques over Internet of Things and this reference studied on IOT data problems be able to solve through "data mining Technologies like Clustering, Classification and frequent pattern mining for IOT" [4]. In paper [12] "they have assessed IoT and big data analytics. Big data analytics be a emerging technology is a key to IoT for decision making". In paper [13] "they made a heart-beat monitoring system using data mining algorithms like classification" [14], clustering [15], "and frequent/sequential pattern mining" [16,17] "from IoT applications". They also represented "an overview of cloud assisted system architecture and Data mining process".

In reference [18] a mining framework that is able to professionally and confidentially recognize the measures into IoT application. In paper [19] they proposed a model in requisites of accuracy or optimization. Some researchers focused on this model on real time applications. In paper [20] they have a detailed study on data mining applications and 3 different views they are knowledge view, technique view, and application view and data mining technologies are integrated with IoT Technologies for better system optimization and accuracy.

3. Key Data Mining Methodologies

A tremendous measure of information is information is available in the climate around us. It appears to be unusable to make this climate strongly having without suitably using the information mining advancements. Information mining can be a main learning with computerization today. Figure 1 the PC helped learning develops all the more accurately when acted in numerous layers in a progressive way.
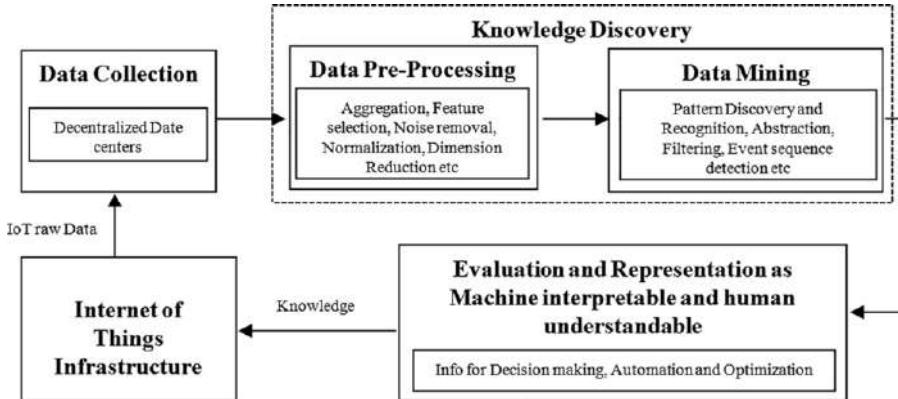

Figure 1. Process of Data Mining in the environment of Internet of Things

3.1 Classification

Characterization is an information mining capacity that allots things in an assortment to target classifications or classes. The objective of characterization is to precisely foresee the objective group in the information for each case. For instance, a characterization model can be consumed to recognize advance candidates as low, medium, or high credit hazards. This assessment is used to recuperate huge and significant information about data, and metadata. This data mining procedure helps with gathering data in different classes.

An order task begins with an informational index in which the class tasks are known. On behalf of instance, an order representation that predicts credit danger might be produced dependent on noticed information for some candidates throughout some extend of time. Notwithstanding the chronicled FICO score of the information may track employment history, number and sort of ventures, home proprietorship or rental, extended periods of living arrangement, etc. The FICO assessment will be the objective for different characteristics and the indicators, and the information for every client would comprise a case.

3.2 Clustering

In clustering or assembling of various objects are delegate comparable items. Informational indexes are isolated into various clusters for examination which depends on the likeness of the information. Later the collection of information into a variety of clusters, a mark is doled out to the cluster causes in regulating to the development by grouping. Bunch investigation in Data Mining Figure 2 involve in discovering the gathering of articles which be like each other in the gathering however are not the same as the item in different gatherings.
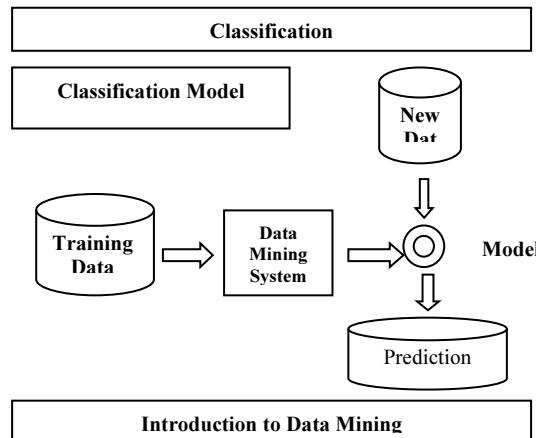
Figure 2. Classification Model

Data Mining cluster analysis applications:

- It helps in distributing records on the web for information revelation.
- It be able to utilized to decide plant and creature scientific classifications, arrangement of qualities with similar functionalities and understanding the structure innate to populaces.
- It helps the ID territories of comparative land to utilize on ground perception and ID of house bunches as per house type, esteem, and topographical area.

3.3. Use of Clustering

Scalability: Adaptability in bunching suggests that as we support the measure of information protests, figure 3 an opportunity to perform grouping ought to around scale to the multifaceted nature request of the calculation.
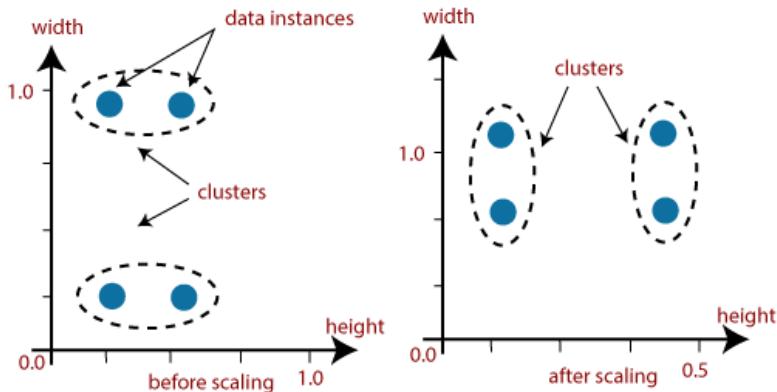
Figure 3. Showing example where Scalability may leads to wrong result

Interpretability: The outcomes of clustering should be interpretable, comprehensible, and usable.

In View of Smart Agriculture: In horticulture creation, the amount and quality necessities of future food and sustenance supplements are significant in our everyday life. As the days are passing the ascent in populace builds step by step. Along these lines, utilizing the idea IOT (internet of things) we can build up this farming framework with huge numbers of the forthcoming cutting-edge innovations. The primary attributes like organization availability, low force sensors and actuators arrangement are the fundamental significant item to be utilized in this venture. For making this agribusiness framework here we use IoT with distributed computing for the better process. The calculation utilized here is the enormous information mining to get the arrangement all the simpler for enhancing and improving shrewd agribusiness framework. Presently in these current days there are a considerable lot of the unexperienced of food should me more and acceptable quality. By the new innovations like completely robotized machines are to be utilized in the horticulture to plant the seeds without the assistance of numerous ranchers. Dribble water system, water reusing and the composts can likewise spread to the plants effectively with these mechanized machines. Getting data (moisture, temperature) from sensor hubs like zigbee sensors and after to lessen the clamor in the detected information play the pressure utilizing the expectation with sifting and afterward the exchange of the information utilizing IoT door and a solid shape introduced on raspberry pi. Expectation about the harvest development utilizing the calculation as choice tree. The keen articles like submerged sensors and AUVs (for boor), Burried sensors (for loUGT), Nano-satellites (for lost). The correspondences , systems administration and confinement of the brilliant submerged investigation and observing , expectation farming, selsmic investigation ,and checking of oil and gas fields, worldwide network past earth, far off detecting and situating. They will be the significant uses of X-IoT.

In View of Smart Transportation: Within urban communities will have a substantial traffic because of the expanding of populace step by step. Along these lines, because of this substantial traffic, numerous mishaps are looking with this traffic. With the idea of the IOT (Internet of Things) we can make the improvement in this transportation to

lessen mishaps. Utilizing the sensors in the vehicle to advise us that to wear safety belt, and when a driver crosses the speed of 80kmph these sensors will be frightened so the driver can diminish speed. To make framework for vehicle aware of try not to glitch condition. For keeping up distance between the vehicle to vehicle is likewise very security insurance to the explorers. The information as well as trade a few administrations information Like security, solace and productivity for social great.

4. Conclusion

This paper demonstrated a systematic and specified evaluate of Data mining algorithms like classification and clustering mining from IoT applications perspective. It is reviewed and organized in tabular issues. We descriptively analyzed the functions like Smart Home, Smart Healthcare, Smart Grid, Ambient Assistant Living, Smart Manufacturing, Smart Agriculture, Industrial IoT, and transportation taking place the establishment of records mining applied sciences employed towards the information translation. Here the conversion will increase more complex interested in today's big information to create IoT environment. In addition, we demonstrated the gadget structure in second section and Data mining manner in the above sections for IoT to pre-processing as well as understanding discovery performs in necessary position amongst the layers. Structures that consist of a range of diverse clever devices and with the purpose of creating heterogeneous data for pre-processing top-notch importance. The machine boosts knowledge discovery and overall presentation with the aid of extra splendid and superior carrier suggestions.

References

[1]   Stankovic, John., "Research directions for the internet of things", IEEE Internet Things Journal, 1 (1): pp. 3–9, 2014.
[2]   Kravchenko Yurii, Olena starkova. et al., "Technology Analysis for smart Home Implementation", In proceedings of 4th International Scientific-Practical Conference on Problems of Infocommunications, Science and Technology, Ukraine, pp.579-584, Oct 2017.
[3]   Miorandi Daniele., De Pellegrini, F., et al., "Internet of things: vision, applications and research challenges", Elsevier Journal. Ad Hoc Netw. 10 (7), pp.1497–1516, 2012.
[4]   Tsai Chun.W., et al., "Data Mining for Internet of Things - A Survey", IEEE Communication Surveys and Tutorials. Vol. 16 issue 1, pp. 451, 2014.
[5]   Google 2017,Big data online Available: https://cloud.google.com/whatis-big-data/.
[6]   Uusitola Mikko A., "Global vision for the future wireless world from the WWRF", IEEE Vehicular Technology Magazine, vol. 1 issue 2, pp. 4–8, 2006.
[7]   Ortiz Antonio M., Hussein Dina., et al., The cluster between internet of things and social networks: review and research challenges. IEEE Internet of Things Journal, vol. 1 issue 3, pp. 206–215, 2014.
[8]   Bijarbooneh Farshid Hassani., Du Wei., et.al, "Cloud-Assisted Data Fusion and Sensor Selection for Internet of Things", IEEE Internet of Things Journal, vol. 3 issue 3, pp.257-268, June 2016.
[9]   Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.
[10] Mohammad Abdul Razzaque, Marija Milojevic-Jevric et al., "Middleware for internet of things-a survey", IEEE Internet Things Journal, vol 3 issue 1, pages 70–95, Feb 2016.
[11] Mohsen Marjani, Fariza Nasaruddin et al., "Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges" IEEE Access, vol 5, pp, 5247-5261, 2017.
[12] Vijay Kumar G, Bharadwaja A, et al., "Temperature and heart beat monitoring system using IOT", Proceedings of ICEI 2017- 2018, pp. 692-695, Jan 2018.

[13] Vijay Kumar G, Sreedevi M., et al., "Incremental mining of popular patterns from   transactional databases", IJET, vol 7 issue 2.7, pp. 636-641, 2018

[14] Vijay Kumar G, Sreedevi M et al., "Regular frequent crime pattern mining on crime datasets" IJET, vol 7 issue 2.7, pp. 972-975, 2018.

[15] Vijay Kumar G, Krishna Chaitanya T, et al., "Mining Popular Patterns from Multidimensional Database", Indian Journal of Science and Technology, Vol 9 issue 17, DOI: 10.17485/ijst/2016/v9i17/93106, May 2016.

[16] Vijay Kumar G, and ValliKumari. V, "Incremental mining for regular frequent patterns in vertical format", IJET, Volume 5, Issue 2, pp. 1506-1511, 2013.

[17] Samer Samarah, Mohammed Gh. Al Zamil et al., "An Efficient Activity Recognition Framework: Toward Privacy-Sensitive Health Data Sensing", IEEE Access, vol 5, pp. 3848-3859, 2017.

[18] Zainal Abedin, Prashengit Dhar et al., "Traffic Sign Detection and Recognition Using Fuzzy Segmentation Approach and Artificial Neural Network Classifier Respectively", International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 518-523, 2017.

[19] Vengatesan K., Kumar A., Parthibhan M., Singhal A., Rajesh R. (2020) Analysis of Mirai Botnet Malware Issues and Its Prediction Methods in Internet of Things. In: Pandian A., Senjyu T., Islam S., Wang H. (eds) Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI - 2018). ICCBI 2018. Lecture Notes on Data Engineering and Communications Technologies, vol 31.

[20] Feng Chen, Pan Deng et al., "Data Mining for the Internet of Things: Literature Review and Challenges", International Journal of Distributed Sensor Networks, vol. 11 issue 8, 2015.

A Survey on Mining Cryptocurrencies

Tejaswini Pawar [a,1], Sagar Shirsat [b], Yaminee Patil [c], Dr. Vijay Sonawane [a],
Dhiraj Birari [a]

[a] *Department of Information Technology, MVPS's KBTCOE Nashik, INDIA*
[b] *Senior Executive - Design, Crompton Greaves Ltd. Nashik, INDIA*
[c] *Department of Information Technology, A P Shah Institute of Technology Thane, INDIA*

Abstract. Advanced monetary standards have acquired huge ubiquity nowadays. Bitcoin is the decentralized, disseminated, distributed virtual cash known cryptographic money. Bitcoin mining chips away at standard of the blockchain, which is believed to be one of this present century's sharp advancement. The blockchain is the arrangement of blocks that are associated so that in the current block there is the hash of the past block. Any adjustment of information in any block in a blockchain brings about a blunder in the entire blockchain. A strategy called mining, where excavators settle a complex numerical riddle, produces Bitcoins. The excavators contend as quickly as time permits to mine the Bitcoin and guarantee the award. Mining should be possible by a solitary individual or by a pool, where a lot of excavators join to mine a solitary block in an organization.

Keywords. Blockchain, Cryptocurrencies, Hash, Mining, Proof of Work (PoW)

1. Introduction

Cryptocurrency is digital asset powered by blockchain technology [1]. Cryptocurrency hold monetary value created by electricity and high- performance computer. It is form of digital money policed by millions of computers called miners on same network and created by mathematical computations. Cryptocurrency works through distributed ledger technology in decentralized manner. The entire cryptocurrency system collectively produces a decentralized cryptocurrency at a pace that is established when the system is generated and widely recognized. Hundreds of Cryptocurrencies, each with its own twist on blockchain technology and numerous intended uses, are available to buy or sell. Bitcoin was first distributed Cryptocurrency released as open-source software in 2009. Crypto currencies use a blockchain technology [2] which is basically a ledger containing a record ofall the transactions on it that have taken place.

The blockchain is decentralized, meaning that it is not hosted in one specific location and can thus not be compromised easily. The smallest unit of a blockchain is a block, and it is a holder containing all the data of the exchange. There are four fields to a block, or essential credits:

---

[1] Tejaswini Pawar, Assistant Professor, Department of Information Technology, MVPS's KBTCOE, India

Email: pawar.tejaswini@kbtcoe.org.

Previous hash: The estimation of the hash of the past block is put away by this trait and that is the means by which the blocks are associated with one another figure 1.

1. **Information:** This is the accumulated arrangement of exchanges that were mined and approved and remembered for the block.
2. **Nonce:** The nonce is an arbitrary worth used to change the presentation of the hash esteem in a "proof of work" agreement calculation that Bitcoin employments. Hash esteem is planned to be created by each block, and the nonce is the boundary used to produce the hash esteem. The confirmation of work is the exchange check measure did in the blockchain.
3. **Hash:** This is the worth acquired through the going through the SHA-256 calculation of the past hash worth, information and nonce; it is the block's advanced mark.
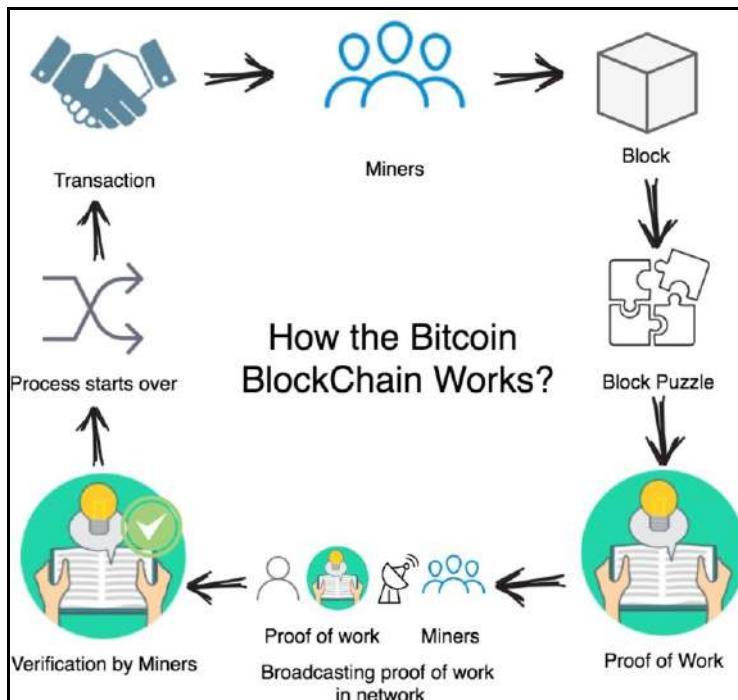

Figure 1. Working of Blockchain in Bitcoin

The means for running the organization incorporate the accompanying [3]:
*   New exchanges are shipped off all nodes.
*   Check if the exchanges are genuine.
*   In a block, every node package new exchanges.
*   Each node is attempting to track down a hard verification of-work.
*   When a node finds a proof-of-work, the block is communicated to all nodes.
*   The block is possibly affirmed by nodes if all exchanges in it are valid and not spent as of now.

- Nodes pass on their endorsement of the block by working utilize the hash to make the following block in the chain, as the earlier hash of the affirmed block.

2. Proof of Work (PoW)

Proof of Work is consensus algorithm consist of complex cryptographic mathematical algorithm. Figure 2 it is introduced by Bitcoin to accumulate the amounts of cryptocurrency. It is verification process which contains complex computations. The value of nonce is hashed with SHA-256 and it will generate hash including number of zeros which is included in particular block in the chain. This high-level mathematical computation calculated by miners using high computing hardware. Bitcoin miners utilize the SHA-256 hashing calculation and determine the hash worth to deliver the hash. On the off chance that it is not exactly the given condition (the objective), at that point it is expected that the puzzle is tackled.

```
┌──────────────────────────────────────────────┐
│   Transactions are bundled together into a block   │
└──────────────────────────────────────────────┘
                        ↓
┌──────────────────────────────────────────────┐
│   Transaction within each block is verified by miners   │
└──────────────────────────────────────────────┘
                        ↓
┌──────────────────────────────────────────────┐
│ To verify transaction miner should solve complex mathematical │
│              puzzle called proof of work              │
└──────────────────────────────────────────────┘
                        ↓
┌──────────────────────────────────────────────┐
│   First miner will get the reward who solved the block   │
└──────────────────────────────────────────────┘
                        ↓
┌──────────────────────────────────────────────┐
│   Verified block by miner is added to blockchain   │
└──────────────────────────────────────────────┘
```
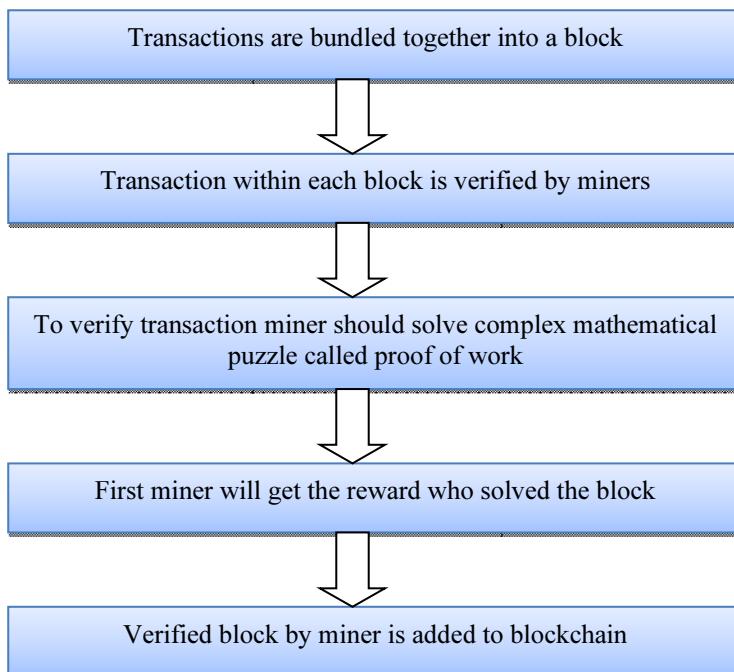
Figure 2. Flow of PoW

3. Bitcoin Mining

Mining is a distributed consensus mechanism which, by including them in the block chain, is used to validate pending transactions. It enforces a sequential order in the block chain, preserves the network's neutrality, and allows the state of the system to be agreed by various computers. Exchanges should be bundled in a block that consents to

extremely severe cryptographic guidelines that will be checked by the organization to be approved. These guidelines forbid the adjustment of past blocks on the grounds that doing so would negate every single ensuing block. Mining additionally creates what might be compared to a serious lottery that keeps any person from progressively adding new blocks to the blockchain without any problem.

Mining guarantees that solitary substantial exchanges in the blockchain of some random digital money are affirmed. Mining is the technique for providing the organization of cryptographic money with a protected repayment instrument. Diggers are gadget proprietors who connect their Figureuring force and assets to the organization of digital money like Bitcoin dependent on "evidence of-work." A part of the cash that is mined as an award is acquired by the main excavator to approve another block for the blockchain. The Figure 3 shows interaction of how Bitcoin blockchain functions [3]. As the Bitcoin organization's hash rate developed, the general measure of 32-cycle nonce was exhausted excessively quickly. The additional nonce arrangement was acquainted with resolve this issue, whereby the coin base exchange is utilized as a wellspring of additional nonce to give a more extensive determination of nonce to be looked by the excavators. By using the following flowchart this method can be visualized:
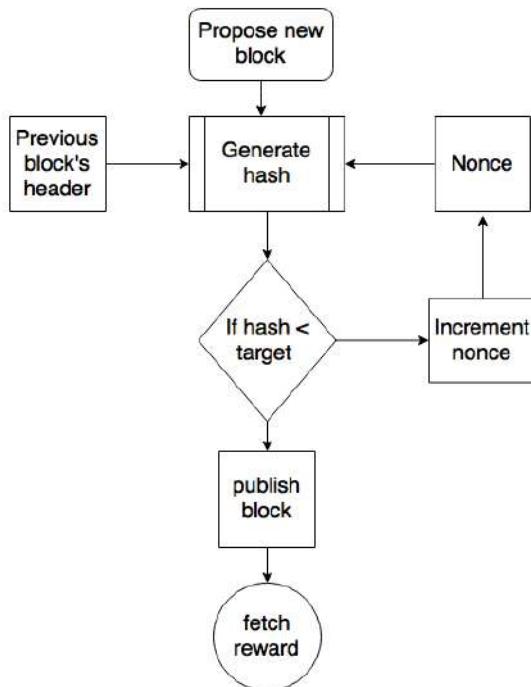

Figure 3. Flowchart of Mining

The calculation for mining comprises of the accompanying advances:
1. A header of the past block is recovered from the Bitcoin network.
2. Assemble a progression of organization exchanges into a block to be proposed.

3. Using the SHA-256 calculation, process the twofold hash of the past block's header joined with a nonce and the recently recommended block.
4. Check that the subsequent hash is lower than the current (target) level of trouble; at that point, PoW is addressed. Because of the fruitful PoW, the block found is communicated to the organization and the award is gotten by excavators.
5. If the subsequent hash isn't not exactly the current degree of trouble (target), rehash the interaction in the wake of expanding the occasions.

4. Hardware for Bitcoin Mining

Bitcoin miners utilize their assets (equipment and power) to approve an exchange, and new Bitcoin are produced in the organization each time a block is mined. Following are the manners in which that portray utilization of various strategies for mining cryptographic money [5].

4.1. CPU Mining

Everything you need to use the CPU method to be able to mine is just a CPU and a couple of programs. Miners used standard processors to overcome the mathematical problems in the early days of Bitcoin, managing processor units (CPUs). It used to require some investment for mining Bitcoin and other cryptographic forms of money, despite the fact that the difficulty levels were less difficult than today. The degree of difficulty proceeds to change and grow, so the excavators have needed to build their preparing power too.

4.2. Cloud Mining

Cloud mining is likely the most famous route to mine digital currencies. Cloud mining has become so mainstream to a great extent since it gives individuals who might not have sufficient cash to purchase their hardware or who may essentially not be keen on claiming an equipment's the capacity to take an interest in the realm of digital currencies. Cloud mining is a technique wherein you pay a specific amount of cash to somebody (most ordinarily a huge organization) and "lease" their mining machine, called a "rig," and the mining cycle itself. This lease goes on for a settled upon span, in which all the income created by the apparatus are moved to your cryptographic money wallet (short the expense of power and upkeep). The people (organizations) that give these cloud mining administrations ordinarily have gigantic mining offices available to them with various ranches (ten or many apparatuses stacked and cooperating) and realize without a doubt how to mine cryptographic money.

There are two alternatives of cloud mining - free and charged. Numerous people searching for approaches to mine digital money would float towards the "free" decisions, however it has its hindrances (exceptionally sluggish mining speeds, additional conditions, and so on) Paying cloud mining ordinarily carries on like this:

A few hosts give you the choice to assemble and design your cloud mining plan. At that point look at the plans that the host gives and go through with the exchange (which

means you pay the host), register digital currency wallet code and that is initial steps to mine

4.3. GPU Mining

The most mainstream and notable strategy for mining cryptographic forms of money is presumably GPU mining. GPU mining is likely the most widely recognized and notable interaction for mining digital forms of money. Designs cards are utilized by GPU apparatuses to mine cryptographic forms of money. A processor, a motherboard, cooling, rig outline and - obviously - a couple (2 - 8) illustrations cards are made of one single apparatus. A normal cost for a well-performing and pleasantly planned GPU mining rig will in general be around the $3000 value range.

4.4. ASIC Mining

Miners use ASIC (application-explicit incorporated circuit) innovation, which was presented explicitly for mining Bitcoin and other digital forms of money. ASICs are very notable and regarded in light of the fact that they make insane amounts of cryptographic money contrasted with the GPU and CPU of their rivals. They burglarize different diggers who use GPU or CPU apparatuses of the capacity to stay aware of both hash paces and income. ASICS is so amazing. ASICs have additionally turned some remarkable cryptographic forms of money's economy - imagine if the bulk of profit will go to one miner with an ASIC ranch

5. Conclusion

Cryptographic forms of money are decentralized and run on the guideline of blockchain. Every one of the exchanges is reasonable and straightforward. The pace of trouble and the opposition between the excavators with the best accessible equipment makes mining more complicated. Digital currency mining needs a great deal of Figureuring and great equipment that can give a great hash rate with low energy. Miners should be cautious about picking equipment prior to beginning to mine digital money due to equipment costs. It is exceptionally high and the other extra expense during mining is the expense of power and fixes.

References

[1] Wikipedia" https:// en.wikipedia.org/ wiki/ History of bitcoin". Accessed, 25 Jul. 2018
[2] https://www.simplilearn.com/bitcoin-mining- explained-article
[3] Suman Ghimire and Henry Selvaraj" A Survey on Bitcoin Cryptocurrency and its Mining". 26th International Conference on Systems Engineering (ICSEng), 18-20 Dec. 2018
[4] https://www.codeproject.com/Articles/125760 1/Introducing-the-Process-of-Mining-in-Blockchain
[5] https://www.bitdegree.org/crypto/tutorials/how-to-mine-cryptocurrency
[6] https://www.bitpanda.com/academy/en/lessons/what-is-bitcoin-mining-and-how-does- mining-work/
[7] https://bitcoin.org/en/how-it-worksA.N. Author, Article title, *Journal Title* **66** (1993), 856–890.

Conformance Checking Techniques of Process Mining: A Survey

Ashok Kumar Saini [a, 1], Ruchi Kamra [b] and Utpal Shrivastava [c]
[a]Dept of CSE, JECRC University, Jaipur, Rajasthan, India
[b]Dept of CSE, Amity University, Manesar, Haryana, India
[c]Dept of CSE, G L Bajaj Group of Inst. Mathura, U.P., India

Abstract. Conformance Checking (CC) techniques enable us to gives the deviation between modelled behavior and actual execution behavior. The majority of organizations have Process-Aware Information Systems for recording the insights of the system. They have the process model to show how the process will be executed. The key intention of Process Mining is to extracting facts from the event log and used them for analysis, ratification, improvement, and redesigning of a process. Researchers have proposed various CC techniques for specific applications and process models. This paper has a detailed study of key concepts and contributions of Process Mining. It also helps in achieving business goals. The current challenges and opportunities in Process Mining are also discussed. The survey is based on CC techniques proposed by researchers with key objectives like quality parameters, perspective, algorithm types, tools, and achievements.

Keywords. Conformance Checking; event log; Petri-net; Process Mining.

1. Introduction

Process Mining (PM) is new research that lies between data science and Business Process Management (BPM)[1]. Generally, BPM processes the model rather than event data. It focuses on the designing, controlling, quantity, and optimization of business processes. Traditional data analytical techniques like machine learning and data mining do not consider the end-to-end process. It focuses mainly on patterns or results. There is a missing link between BPM and data science, namely PM [2][3], to improve the process. Nowadays, most organizations use information systems such as BPM, Enterprise Resource Planning systems, etc. These information systems record each activity and describe a process's underlying behaviour, as shown in Figure 1. Each event is related to a movement that belongs to a particular stage of the process[4][5]. PM uses these events to discover, monitor, and improve the process [4]. The organization of the paper is as follows: Section I is the introduction of PM. Section II is PM techniques and it's applicability in different perspectives. Section III is a discussion of tools used in PM. Section IV is a detailed discussion of various research work carried out in CC techniques. Lastly, section V is the contributions and the scope of future research in this domain.

---

[1] Ashok Kumar Saini, Department of CSE, India
 Email: shksaini@gmail.com.
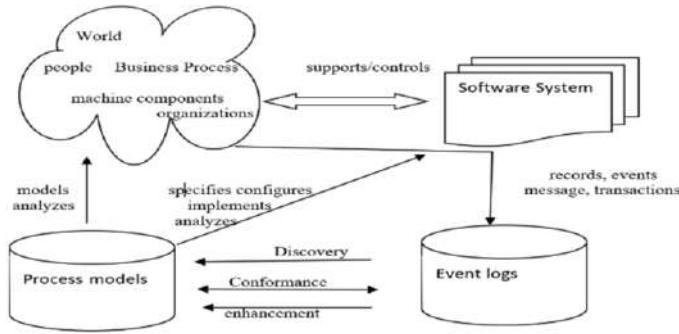
Figure 1. Process mining overview and its techniques

2. Process Mining Techniques Overview

Components of PM in Figure 2 are process discovery that has an event log as input, CC, and enhancement both have log and model as input. Process discovery: First technique of PM is process discovery for discovering the model that replicates log [2]. CC: It checks the conformity of the model with log and assesses whether they described reality. There are four quality parameters [6]: fitness, simplicity, precision, and generalization. A perfect fitness model can replay all traces from beginning to end. For any log (EL) and model (M), then the fitness of the model is:

$$\text{fitness}(\text{EL}, \text{M}) = 1 - \frac{fcost(\text{EL}, M)}{move(\text{EL}) + |\text{EL}| + move(M)}$$

Enhancement: It takes the process model and event log as input and enhances the process model using the observed event log [4][5]. A model is simple if it explains clearly all behaviors [5][6] shown in Figure 3. A precise model does not allow many traces. The flower model is less precise and more generalized. The fitness value varies from 0 to 1. The best-fitted model has one fitness value. The model that is not generalized is also called overfitting.
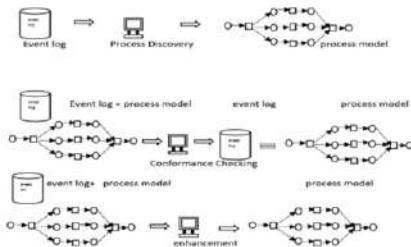

**Figure 2.** PM techniques in terms of input and output   **Figure 3.** Flower model
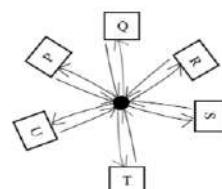
- *PM perspective:* There are mainly four types of PM perspective. First is a control-flow, to find the excellent categorization of numerous promising paths

[5]. The second organizational, shows how movements are associated with each other. The third case, emphasizes cases, and the fourth perspective time related to evaluating cases [4][5].

3. Tools and Algorithms Categories

Comparative usability of ProM and Disco tools of PM are shown in table 1. Some other tools are ProM Lite, RapidProM (both are open source), LANA (Lana Labs), SNP (SNP Schneider-Neureither & Partner AG), EDS (StereoLOGIC Ltd), Icris, ProcessGold, ARIS PPM, Fujitsu, Icaro, Minit, myInvenio, QPR, Rialto, etc. At the time of loading event in ProM framework shown in Figure 4.

Table 1. Comparison of Modulation schemes

Tools	ProM	Disco
Class	Open	Commercial
Purpose	General	General
Discovery	supported	supported
CC Checking	supported	Not supported
Societal Mining	supported	Not supported
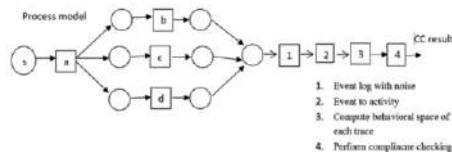
PM algorithm is categorized into three classes: Deterministic, Heuristic, and Genetic algorithms [6]. Deterministic like α-algorithm provide constant output for the specific input of variables. Heuristic algorithms provide a better solution by trial and error. A genetic algorithm is used when the problem starts with an arbitrary point and tries to find a better solution by introducing random variations [7].
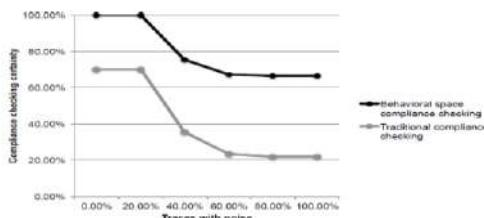
Figure 4. ProM framework when loading event log

4. Proposed Models by Researchers

This section surveys the proposed work on CC. Table 2 has shown the comparative survey. Most CC techniques are created on a control-flow perspective and offline mode, but the conformity of the model also depends on different perspectives like data, time, etc. Online CC framework [9] proposed quantifying the observed behaviour in real-time and controlling the complexity to the constant time of each event. To test the approach, they run a conformance checker for about 70 mins and 256110 events generated by generator PLG about 65 events/sec. Process model [10] discovered from different process discovery algorithms and compared these algorithms on same data streams such as Lossy Counting with Budget, Sliding Window, and Exponential Decay.

Figure 5. Evolution process

In Figure 5, the first and second step is to build in existing plugins and the probabilistic CC approach is implemented as a plugin in third step [11]. For checking conformity of the model, mapping is required. For noise level 0, their compliance checking technique results in 70.2%, and traditional methods provide 29.8%. They are shown in Figure 6.


Figure 6. Comparative analysis between traditional and probabilistic Conformance Checking.

The CC technique is based on alignment [13]; for alignment, it is essential to associate the passage of events with the passage of the process model. One more CC technique [14] is based on replay-token. A new approximation CC technique [15] was proposed to compute conformance value in a faster way. [16] found the value of all four parameters fitness f=0.995, precision p=0.996, generalization g=0.958 and simplicity s=0.387. [17] proposed an approach that detects the anomalies in traces stored in PAIS using the ProM tool. [25] proposed a novel framework for PM analysis that uses advances in-memory data processing and graph algorithms that reduce the cost of taking out and converting the event data present in the information system. [28] Increasing the volume of data becomes a challenge as existing PM techniques cannot handle the high volume of data with many activities. Clustering [30] based approach that overcomes the problem of the complex and imprecise model due to large volume of data. PM technique [29] was used to analyze the process of an emergency room in the hospital. But they [27] neither considered real-time concerns in the behavioural domain nor the resources and relationships between actions. [26] expanded the work done in [28], provided approach of reconstructing process model from audit trail logs.

Table 2. Different approaches to Conformance Checking.

Ref	Key Objective	Achievements	Future work	Tool
[10]	Comparing results visually of two different process discovery algorithms.	Users can analyze the internal data structure for handling the event data stream.	Make CC metrics checking the performance algorithms.	ProM
[11]	Creating a mapping between process model and uncertain events.	Applicable on several real-world procedures where traditional CC methods fail.	Need to extend the approach of mapping and also help in the selection.	ProM 6
[12]	Detect the deviation between modelled and observed behaviour.	The hierarchical approach is compared with decomposition by manual.	Enhance the projected technique to a more significant class.	ProM
[13]	Maintain alignment between events and model.	Alignment makes it possible to replay the event.	Finding an optimal alignment algorithm.	ProM

[15]	Finding the possible behaviour of subset.	Approximation value is close to actual alignment value.	The best subset method.	ProM
[16]	Find the similarities between PM and event log.	Fitness of the model is 0.87, precision=0.9, simplicity=0.3 8 and generalization=0.98.	Finding learning automata for discovering the process model	ProM
[18]	Design rule sets that show the relationship between tasks. Traditional methods are time-consuming.	The rule set considers noise and imbalance in data and the problem with the alpha algorithm recovered.	Planning of performing more real-world case studies on discovering the model	ProM
[21]	Dealing with the log consisting noise.	Experiments with a synthetic and real-time log.	Future work for dealing with duplicate tasks.	ProM
[22]	An alignment-based replay to enhance the state space.	Handled intertwined state space with the help of alignment-based replay.	Enhance the CC matrix by parameters precision and generalization.	ProM
[23]	Process model generates with minimum information.	A block-structured model that is fit and sound replays all the observed behavior.	Length of two-loop removes the restriction of start and end state.	
[24]	Approach provides the instance graph of an individual log instance.	In this approach, the noise is filtered out from the log.	Developing algorithms that integrated multiple instances.	
[27]	Provided an approach of constructing a PM	The modelling technique is compatible.	Need controlling learning conditions	Flow mark

5. Conclusion and Future Work

The study shows that PM techniques are limited for process discovery and check the conformity of the process model. There are many proposed CC techniques. Most CC techniques are based on control flow and do not provide an actual cause of process deviation. The other parameters like data, resources, time need to be considered. Many CC techniques are token-based, but sometimes they give unpredictable and ambiguous results. The CC techniques based on alignment provide more strong conformity. As the volume of data is increased day by day, the alignment-based approach comes with challenges. The available tools also face difficulties. For massive data, these techniques are inefficient, being a complex process model. To overcome such problems, a decompose alignment technique is used. Such methods are based on the decomposition of model and event log into small components and aligned respectively; the decomposition technique shows better computation time. The detailed study shows the need for future work for the automatic decomposition of the process model and event log with minimum error and execution time.

References

[1]   Van der Aalst, W., & Damiani, E. (2015). Processes meet big data: Connecting data science with process science. IEEE Transactions on Services Computing, 8(6), 810-819.

[2]   Van Der Aalst, W. (2016). Data science in action. In Process mining (pp. 3-23). Springer, Berlin, Heidelberg.

[3]   Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). Fundamentals of business process management (Vol. 1, p. 2). Heidelberg: Springer.

[4]   Van, D. (2011). Process mining discovery, conformance and enhancement of business processes. Springer Heidelberg Dordrecht London New York. ISBN.

[5]   Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., & Wynn, M. (2011, August). Process mining manifesto. In International Conference on Business Process Management (pp. 169-194). Springer, Berlin, Heidelberg.

[6]   Lee, W. L. J., Verbeek, H. M. W., Munoz-Gama, J., van der Aalst, W. M., & Sepúlveda, M. (2018). Recomposing conformance: Closing the circle on decomposed alignment-based conformance checking in process mining. Information Sciences, 466, 55-91.

[7]   Mahendrawathi, E. R., Arsad, N., Astuti, H. M., Kusumawardani, R. P., & Utami, R. A. (2018). Analysis of production planning in a global manufacturing company with process mining. Journal of Enterprise Information Management.

[8]   Nagy, Z., & Werner-Stark, Á. (2020, March). A Multi-perspective Online Conformance Checking Technique. In 2020 6th Int. Conference on Information Management (ICIM) (pp. 172-176). IEEE.

[9]   Burattin, A., & Carmona, J. (2017, September). A framework for online conformance checking. In International Conference on Business Process Management (pp. 165-177). Springer, Cham.

[10] Baskar, K., & Hassani, M. (2019, January). Online Comparison of Streaming Process Discovery Algorithms. In BPM (PhD/Demos) (pp. 164-168).

[11] van der Aa, H., Leopold, H., & Reijers, H. A. (2017, June). Checking process compliance on the basis of uncertain event-to-activity mappings. In International Conference on Advanced Information Systems Engineering (pp. 79-93). Springer, Cham.

[12] Munoz-Gama, J., Carmona, J., & van der Aalst, W. M. (2013, June). Hierarchical conformance checking of process models based on event logs. In International Conference on Applications and Theory of Petri Nets and Concurrency (pp. 291-310). Springer, Berlin, Heidelberg.

[13] Van der Aalst, W., Adriansyah, A., & van Dongen, B. (2012). Replaying history on process models for conformance checking and performance analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(2), 182-192.

[14] Rozinat, A., & Van der Aalst, W. M. (2008). Conformance checking of processes based on monitoring real behavior. Information Systems, 33(1), 64-95.

[15] Sani, M. F., van Zelst, S. J., & van der Aalst, W. M. (2020, June). Conformance checking approximation using subset selection and edit distance. In International Conference on Advanced Information Systems Engineering (pp. 234-251). Springer, Cham.

[16] Naderifar, V., Sahran, S., & Shukur, Z. (2019). A review on conformance checking technique for the evaluation of process mining algorithms. TEM Journal, 8(4), 1232.

[17] Bezerra, F., Wainer, J., & van der Aalst, W. M. (2009). Anomaly detection using process mining. In Enterprise, business-process and information systems modeling (pp. 149-161). Springer, Berlin, Heidelberg.

[18] Măruşter, L., Weijters, A. T., Van Der Aalst, W. M., & Van Den Bosch, A. (2006). A rule-based approach for process discovery: Dealing with noise and imbalance in process logs. Data mining and knowledge discovery, 13(1), 67-87.

[19] Accorsi, R., & Stocker, T. (2012, March). On the exploitation of process mining for security audits: the conformance checking case. In Proceedings of the 27th Annual ACM Symposium on Applied Computing (pp. 1709-1716).

[20] Cook, J. E., & Wolf, A. L. (1996). Process discovery and validation through event-data analysis (Doctoral dissertation, University of Colorado).

[21] de Medeiros, A. K. A., Weijters, A. J., & van der Aalst, W. M. (2007). Genetic process mining: an experimental evaluation. Data Mining and Knowledge Discovery, 14(2), 245-304.

[22] Camargo, M., Dumas, M., & Gonzalez-Rojas, O. (2020). Discovering Generative Models from Event Logs: Data-driven Simulation vs Deep Learning. arXiv preprint arXiv:2009.03567.

[23] Leemans, S. J., Fahland, D., & van der Aalst, W. M. (2013, June). Discovering block-structured process models from event logs-a constructive approach. In International conference on applications and theory of Petri nets and concurrency (pp. 311-329). Springer, Berlin, Heidelberg.

[24] Van Dongen, B. F., & Van der Aalst, W. M. (2004, November). Multi-phase process mining: Building instance graphs. In International Conference on Conceptual Modeling (pp. 362-376). Springer, Berlin, Heidelberg.

[25] Berti, A. (2019). Process Mining on Event Graphs: a Framework to Extensively Support Projects. In BPM (PhD/Demos) (pp. 60-65).

[26] Pinter, S. S., & Golani, M. (2004). Discovering workflow models from activities' lifespans. Computers in Industry, 53(3), 283-296.

[27] Agrawal, R., Gunopulos, D., & Leymann, F. (1998, March). Mining process models from workflow logs. In Int. Conference on Extending Database Tech. (pp. 467-483). Springer, Berlin, Heidelberg.

[28] Van der Aalst, W. M. (2013). Decomposing Petri nets for process mining: A generic approach. Distributed and Parallel Databases, 31(4), 471-507.

[29] Badakhshan, P., & Alibabaei, A. (2020). Using process mining for process analysis improvement in pre-hospital emergency. In ICT for an Inclusive World (pp. 567-580). Springer, Cham.

[30] Sani, M. F., Boltenhagen, M., & van der Aalst, W. (2020, September). Prototype Selection using Clustering and Conformance Metrics for Process Discovery. In Int. Conference on Business Process Management (pp. 281-294). Springer, Cham.

Machine Learning: A Quantum Perspective

Aishwarya Jhanwar [a, 1] and Manisha J. Nene [a]

[a] Department of Computer Science and Engineering, Defence Institute of Advanced Technology, Pune-411025, Maharashtra

Abstract. Recently, increased availability of the data has led to advances in the field of machine learning. Despite of the growth in the domain of machine learning, the proximity to the physical limits of chip fabrication in classical computing is motivating researchers to explore the properties of quantum computing. Since quantum computers leverages the properties of quantum mechanics, it carries the ability to surpass classical computers in machine learning tasks. The study in this paper contributes in enabling researchers to understand how quantum computers can bring a paradigm shift in the field of machine learning. This paper addresses the concepts of quantum computing which influences machine learning in a quantum world. It also states the speedup observed in different machine learning algorithms when executed on quantum computers. The paper towards the end advocates the use of quantum application software and throw light on the existing challenges faced by quantum computers in the current scenario.

Keywords. Machine Learning, Quantum Computing, Quantum Machine Learning, Quantum Supremacy, Speedup

1. Introduction

In recent years, owing to increase number of datasets, machine learning has gained utmost success in transforming science and technology [1]. The aim of machine learning is to enable computers to act in a manner that barely involves any kind of human intervention and does not require the need of being programmed explicitly [2]. The vast ap-plications of machine learning ranging from predicting protein structure, drug discovery in biology [3] to black hole detection [4], wave analysis in physics [5] to speech recognition, self-driving cars, etc. in computers exhibits a remarkable contribution in today's scenario. However, with this ever-growing size of dataset, it has posed certain challenges in this field. With the Moore's law moving towards its extinction, we might probably attain a peak where current computational methods would not suffice handling of such humongous datasets [6].

This motivated to use the concept of quantum computing which is based on quantum mechanics. The application of quantum mechanics in the discipline of information processing is termed as Quantum Information Processing [7]. By exploiting the properties of quantum computing such as entanglement and superposition, quantum computers could efficiently solve some problems that are perceived as hard problems for classical computers [8,9,10].The applications developed in the field of machine learning using quantum computing algorithms exceeds the capabilities of any classical computer achieving quantum supremacy. A successful demonstration of quantum supremacy would prove that engineered quantum systems can outperform the most advanced classical computers [11].

---

[1] Aishwarya Jhanwar,Defence Institute of Advanced Technology, Pune, Inda
Email: aishwaryajhanwar2001@gmail.com.

This paper aims to provide technical insights of integrating quantum computing with machine learning that will prove beneficial to readers of both the communities i.e., of machine learning as well as of quantum computing.

2. Classical Machine Learning

Machine learning, is a sub-discipline of artificial intelligence where learning from data is involved. In 1959, Arthur Samuel described machine learning as "a field of study that gives computers the ability to learn without being explicitly programmed" [12]. Machine learning is used to find hidden structures and interesting patterns from given dataset. The intrinsic value of data can be augmented with machine learning. This value can be fetched in different manner depending upon the nature of the dataset available. Broadly, the term learning in machine learning can be divided into three types: Supervised, Unsupervised and Reinforcement Learning. In Supervised Learning a computer is fed with input-output data pairs that helps to infer an algorithm and predict the output. It is generally used for classification of data based on the labels. Examples of supervised learning classifier includes K Nearest Neighbors, Decision Tree, Naive Bayes and Random Forest models. In Unsupervised Learning input-output data pairs are not given to the machine, rather the system is fed with unlabeled data and hence it builds an algorithm to find the patterns. It is used for clustering of data. Some of the examples are Apriori, K-means and Principal Component Analysis. Reinforcement Learning uses software agents to earn cumulative rewards and thus proceed to train the algorithm in a certain manner. In all these learning methods, the indispensable role is of training phase. Training phase is often considered as most costly phase of any machine learning process. When dealing with extremely large datasets efficient training methods plays a vital role.

Irrespective of what learning method is adopted, an optimal machine learning algorithm is the one which produces a minimum error rate by consuming minimum resources. Due to ever increasing rate of data, current machine learning systems are nudging the limit of classical computational resources. Challenges lies in the problem of finding an optimal solution to any problem, that helps to minimize the complexity class of the problem [13]. This is where there arises a need to shift to quantum computing.

3. Quantum Computing

Quantum computing deals with the problem of processing, storing and transferring information which is encoded in quantum mechanical systems. This mode of information is termed as quantum information. Quantum information is processed by acting upon quantum systems. Basic quantum computing concepts that influence any algorithm to perform in an efficient manner [14].

3.1. Qubit

Quantum bits or qubit is the fundamental unit of processing information in quantum systems analogous to a bit in classical systems. A qubit can exist in one or more than one state simultaneously. However, the information content stored in a qubit that is

equivalent to a single bit in classical medium. It is defined as linear superposition of states.

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \qquad\qquad \alpha, \beta \in C$$

where, $|\alpha|^2 + |\beta|^2 = 1$. The complex numbers α and β are probabilities of the basic states of $|0\rangle$ and $|1\rangle$, and $|0\rangle$ and $|1\rangle$ represent the orthogonal states.

3.2. Superposition

A qubit can be in a state of $|0\rangle$ and $|1\rangle$ at the same time. This implies that to store two bits worth of value only one qubit is required. On similar lines for 4 bits, 2 qubits will be required and hence, 2^N *Classical Bits* can be stored on *N Qubits.* This means that a particle has an ability to exist in multiple quantum states and when the measurement is performed, it undergoes certain changes resulting in a probabilistic value, thereby losing its individuality [15,16].

3.3. Entanglement

In a multi qubit system, qubits exist in a manner that they lose their individuality. It means the property of one qubit is connected to the property of another qubit. Thus, this property helps the qubits to be correlated with each other despite of being separated by large physical distances [17,18].

3.4. No Cloning Theorem

This theorem states that it is not possible to create an identical copy of an arbitrary unknown quantum state. It helps to infer that once a measurement is performed, it is not sure to get the same information after another measurement being performed on an already measured state.

Quantum measurement depends on the wave function collapse. It occurs when a wave function, initially in a superposition of several eigen states reduces to a single eigen state due to interaction with the external world. This leads to measurement of a Quantum state which is probabilistic in nature.

These paramount features offered by quantum computers are required for high-speed computing which paves way to delve deeper and integrate machine learning with quan-tum computing.

4. Quantum Machine Learning

Quantum Machine Learning (QML) is the intersection of machine learning process with the concepts of quantum computing. In QML, quantum algorithms are developed to solve intricate problems of machine learning utilizing the potency of quantum computing. This is achieved by enabling expensive subroutines of classical algorithms to be executed on a quantum computer. The properties of quantum computing such as superposition induces parallelism in quantum computers which allows to evaluate function on many inputs in machine learning algorithms simultaneously. Entanglement provides a mechanism for improving the storage capacity as well as retrieving

corrupted or incomplete information [19,20]. These properties thus provide significant speedup of any computation evaluated on the basis of complexity [21]. By speedup it means the advantages obtained in run time by any quantum algorithm as compared to the classical methods used for the same task [22]. Table 1 gives a general overview of the different machine learning algorithms and their speedups when executed on quantum computers. The subroutines used by these algorithms uses the properties of quantum computing as discussed in section 3 to enhance their efficiency. It will thus help to understand how quantum computers can revolutionize machine learning.

Table 1. Speed up of machine learning algorithms in quantum environment

Algorithm	Speedup
Associative Memory [23]	Exponential improvement in capacity
Boosting [24]	Quadratic
Divisive Clustering [25]	Quadratic
Gradient Descent [26]	Exponential
K-Means [21,27]	Exponential
K-Medians [25]	Quadratic
K Nearest Neighbors [28]	Quadratic
Pattern Recognition [29]	Exponential improvement in capacity
Principal Components Analysis [30]	Exponential
Q-means [31]	Exponential
Recommendation Systems [32]	Exponential
Support Vector Machines [33]	Exponential

5. Quantum Application Software and Challenges

The gap between the computational models and its actual implementation has been narrowed down to a great extent with the development of open-source quantum software platforms around the globe. Some leading-edge quantum platforms like QISKIT [34], DWave [35], ProjectQ [36], Forest [37], Strawberry Fields [38], Quantum Development Kit [39] and Cirq [40]. allow these algorithms to be implemented on real quantum computers which can be accessed through cloud-based services or quantum simulators which runs on classical machines [41]. Thus, these software makes it possible to implement machine learning in quantum environment and experience the enhanced outcomes.

QML poses a great challenge in the field of Quantum Information Processing as quantum computing is still in its infancy stage of development. Qubits are affected by various disturbances such as vibrations, electromagnetic waves, temperature

fluctuations, cosmic rays, etc. that are induced by the effects of quantum mechanics which are extremely sensitive to external conditions. Thus, the fragile nature of qubits makes them pone to high errors. The researchers are working in the field of mitigating errors induced due to noise and fragile nature of quantum computers [42]. Another challenge is the limited availability of quantum computers. As of now rigorous study is required to decide upon how many logical qubits are required by quantum computers to exceed the capabilities of classical computers, that are extremely powerful [43].

6. Conclusion

The study in this paper reveals that with the existence of quantum supremacy a drastic change will be seen in the field of machine learning. The immense challenges require different approaches of quantum computing to be devised and investigated to enhance machine learning methods. The properties leveraged by quantum computers helps to achieve considerable speedups in the field of machine learning. Although QML has gained a lot of awareness amongst quantum scientists and machine learning researchers, yet in order to gain maximum benefits areas which remain unfolded may be explored and delved deeper.

References

[1] Silver D et al. 2016 Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489. (doi:10.1038/nature16961)
[2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.
[3] Mohammed AlQuraishi. Alphafold at casp13. Bioinformatics, 35(22):4862–4865, 2019.
[4] Benjamin P Abbott, Richard Abbott, TD Abbott, MR Abernathy, Fausto Acernese, Kendall Ackley, Carl Adams, Thomas Adams, Paolo Addesso, RX Adhikari, et al. Observation of gravitational waves from a binary black hole merger.Physical review letters, 116(6):061102, 2016.
[5] Louis-Franc¸ois Arsenault, O Anatole von Lilienfeld, and Andrew J Millis. Machine learning for many-body physics: efficient solution of dynamical mean-field theory. arXiv preprint arXiv:1506.08858, 2015.
[6] Markov IL. 2014 Limits on fundamental limits to computation. Nature 512, 147–154.
[7] SM. Barnett, "Introduction to Quantum Information", School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK, Oxford University Press, pp. 11-33
[8] Shor PW. 1997 Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM J. Comput. 26, 1484–1509. (doi:10.1137/S0097539795293172)
[9] Nene M.J., Upadhyay G. (2016) Shor's Algorithm for Quantum Factoring. In: Choudhary R., Mandal J., Auluck N., Nagarajaram H. (eds) Advanced Computing and Communication Technologies. Advances in Intelligent Systems and Computing, vol 452. Springer, Singapore. https://doi.org/10.1007/ 978-981-10-1023-1 33
[10] Van Dam W, Hallgren S, Ip L. 2006 Quantum algorithms for some hidden shift problems. SIAM J. Comput. 36, 763–778. (doi:10.1137/S009753970343141X)
[11] Neill C, Roushan P, Kechedzhi K, Boixo S, Isakov SV, Smelyanskiy V, Megrant A, Chiaro B, Dunsworth A, Arya K, Barends R, Burkett B, Chen Y, Chen Z, Fowler A, Foxen B, Giustina M, Graff R, Jeffrey E, Huang T, Kelly J, Klimov P, Lucero E, Mutus J, Neeley M, Quintana C, Sank D, Vainsencher A, Wenner J, White TC, Neven H, Martinis JM. A blueprint for demonstrating quantum supremacy with su-perconducting qubits. Science. 2018 Apr 13;360(6385):195-199. doi: 10.1126/science.aao4309. PMID: 29650670.
[12] Arthur L Samuel. Some studies in machine learning using the game of checkers. IBM Jour- nal of research and development, 44(1.2):206- 226, 2000.
[13] Schuld, Maria & Sinayskiy, Ilya & Petruccione, Francesco. (2014). An introduction to quantum machine learning. Contemporary Physics. 56. 10.1080/00107514.2014.964942.
[14] Sharma, MK, Nene, MJ. "Dual factor third-party biometric-based authentication scheme using quantum one time passwords". Security and Privacy. 2020; 3:e129.https://doi.org/10. 1002/spy2.129

[15] G. Upadhyay and M. J. Nene, "One time pad generation using quantum superposition states," 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, 2016, pp. 1882-1886, doi: 10.1109/RTEICT.2016.7808162.

[16] M. Gupta and M. J. Nene, "Random Sequence Generation using Superconducting Qubits," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 640-645, doi: 10.1109/ICICV50876.2021.9388495.

[17] M. Gupta and M. J. Nene, "Quantum Computing: An Entanglement Measurement," 2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI), 2020, pp. 1-6, doi: 10.1109/ICATMRI51801.2020.9398441.

[18] P. Nema and M. J. Nene, "Pauli Matrix based Quantum Communication Protocol," 2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI), 2020, pp. 1-6, doi: 10.1109/ICATMRI51801.2020.9398393.

[19] Adcock J et al. 2015 Advances in quantum machine learning. (http://arxiv.org/abs/ 1512.02900).

[20] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S. 2017 Quantum machine learning. Nature 549, 195–202. (doi:10.1038/nature23474)

[21] Lloyd, S., Mohseni, M. & Rebentrost, P. Quantum algorithms for supervised and unsupervised machine learning. arXiv:1307.0411 (2013).

[22] Ciliberto, Carlo & Herbster, Mark & Ialongo, Alessandro & Pontil, Massimiliano & Rocchetto, An-drea & Severini, Simone & Wossnig, Leonard. (2017). Quantum machine learning: A classical per-spective. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science. 474. 10.1098/rspa.2017.0551.

[23] D. Ventura and T. Martinez, "Quantum Associative Memory," Int. J. Mod. Phys. B, vol. 17, no. 12, pp. 2447–2472, Jul. 1998.

[24] H. Neven, V. S. Denchev, G. Rose, and W. G. Macready, "Training a Large Scale Classifier with the Quantum Adiabatic Algorithm," pp. 1–11, 2009.

[25] E. A¨ımeur, G. Brassard, and S. Gambs, "Quantum clustering algorithms," in Proceedings of the 24th international conference on Machine learning - ICML '07, 2007, vol. 227, pp. 1–8.

[26] I. Kerenidis and A. Prakash, "Quantum gradient descent for linear systems and least squares," 2017.

[27] Khan, Sumsam Awan, Ahsan Javed Vall Llosera Gemma 2019 K Means Clustering on Noisy Intermediate Scale Quantum Computers.

[28] N. Wiebe, A. Kapoor, and K. M. Svore, "Quantum algorithms for nearest neighbor methods for supervised and unsupervised learning," Quantum Inf. Comput., vol. 15, no. 3–4, pp. 318–358, 2015.

[29] C. A. Trugenberger, "Quantum Pattern Recognition," in Invited Talk at the 1st Feynman Festival, Univ. of Maryland, College Park, August 2002, 2002.

[30] S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum principal component analysis," Nat. Phys., vol. 10, no. 9, pp. 631–633, 2014.

[31] I. Kerenidis, J. Landman, A. Luongo, and A. Prakash, "q-means: A quantum algorithm for unsupervised machine learning," 2018.

[32] I. Kerenidis and A. Prakash, "Quantum recommendation system," Leibniz Int. Proc. Informatics, LIPIcs, vol. 67, pp. 1–22, 2017.

[33] D. Anguita, S. Ridella, F. Rivieccio, and R. Zunino, "Quantum optimization for training support vector machines," Neural Networks, vol. 16, no. 5–6, pp. 763–770, 2003.

[34] Qiskit https://qiskit.org/ last accessed 14 April 2021

[35] DWave The Quantum Computing Company https://www.dwavesys.com/quantum-computing last accessed 14 April 2021

[36] ProjectQ https://projectq.ch/ last accessed 14 April 2021

[37] Rigetti https://www.rigetti.com/ last accessed 14 April 2021

[38] Strawberry Fields https://strawberryfields.ai/ last accessed 14 April 2021

[39] Q# and The Quantum Development Kit https://azure.microsoft.com/en-us/resources/developmentkit/quantum-computing/ last accessed 14 April 2021

[40] Cirq https://quantumai.google/cirq last accessed 14 April 2021

[41] LaRose, R. (2019). Overview and Comparison of Gate Level Quantum Software Platforms. Quantum, 3, 130.

[42] Fowler, A. G.; Mariantoni, M.; Martinis, J. M. & Cleland, A. N. "Surface Codes: Towards Practical Large-scale Quantum Computation". Physical Review A, 86(3), 32324 (2012)

[43] Muller, R. P. & Blume-Kohout, R. "The Promise of Quantum Simulation". ACS Nano (2015)

Deep Learning Based Object Recognition in Real Time Images Using Thermal Imaging System

Rohini Goel [a, 1] , Avinash Sharma [a], and Rajiv Kapoor [b]

[a] *CSE Department, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India*
[b]*Professor, Department of ECE, Delhi Technological University Delhi, India*

Abstract. An efficient driver assistance system is essential to avoid mishaps. The collision between the vehicles and objects before vehicle is the one of the principle reason of mishaps that outcomes in terms of diminished safety and higher monetary loss. Researchers are interminably attempting to upgrade the safety means for diminishing the mishap rates. This paper proposes an accurate and proficient technique for identifying objects in front of vehicles utilizing thermal imaging framework. For this purpose, image dataset is obtained with the help of a night vision IR camera. This strategy presents deep network based procedure for recognition of objects in thermal images. The deep network gives the model understanding of real world objects and empowers the object recognition. The real time thermal image database is utilized for the training and validation of deep network. In this work, Faster R-CNN is used to adequately identify objects in real time thermal images. This work can be an incredible help for driver assistance framework. The outcomes exhibits that the proposed work assists to boost public safety with good accuracy.

Keywords. Thermal Imaging, HSV color space Segmentation, Faster Region-Convolution Neural Network (Faster-RCNN), Region Proposal Network, Intersection over Union (IoU).

1. Introduction

Driver assistance frameworks are probably the important applications object recognition. Individual in front of vehicles face a high danger of injury or demise during the night or because of restricted perceivability. Having a warning framework [1] to hinder interlopers is pivotal for both the safety of public and to avoid monetary losses. To forestall the mishaps, several intelligent frameworks have been created utilizing close circuit TV cameras combined with machine learning innovations [2]. However, most frameworks [3] work along with database of images captured in proper light, while the detection performance in front of vehicle in the restricted perceivability [4] is extensively more awful. In thermal imaging system, the radiations as infrared light are detected by the device called thermal camera to capture the thermal images. To recognize individuals in the restricted perceivability, several investigations have

---

[1] Rohini Goel, CSE Department, Maharishi Markandeshwar (Deemed to be University), Mullana

Email:27rohinisharma@gmail.com

adopted machine learning based methods [5] utilizing thermal IR images. The machine learning based approaches can be distinguished on the basis of feature descriptors & classifiers such as HOG[6], SIFT[7], INERTIA[8], SVM [9][10] or Adaboost [11].
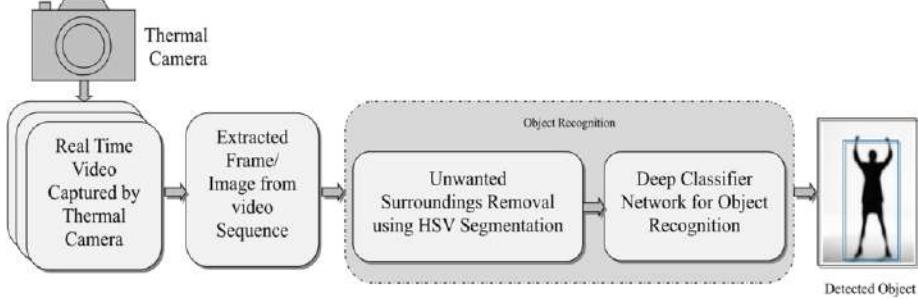

Figure 1. Basic block diagram of the proposed method used to recognize the object.

In some human recognition applications haar feature based classifiers [12][13] is used for thermal IR images. Although the machine learning methods [14] performs satisfactorily but for images having noisy/ ambiguoused features[15] suffer from degraded performance. To handle this problem convolutional neural network [16] [17] based deep learning methods [18] have been used for object recognition in thermal images. The recognition performance for thermal images can be upgraded using deep network based methodologies [19]. It has observed that human detection methods based on CNN [20][21] give better recognition results as compared to methods based on SVM. The selection of proper training data affects the recognition performance of deep network methods[22]. Several investigations have created thermal image datasets, e.g. CVC-14 [24], LSI [25] and KAIST [23] for person recognition in limited visibility. The work presents a precise as well as proficient strategy to recognize objects in real time thermal images. In this paper, a system is intended to segment the moving objects from the image using color HSV segmentation. Finally, the Faster-RCNN is utilized to distinguish the objects in the real time thermal images.This paper is organized in various subsections as: section II presents the proposed procedure in detail including collection steps using thermal imaging, HSV segmentation and deep network [26] (Faster R-CNN) for recognition of object. The outcomes and discussions are explained in section III. Finally, the conclusion of paper is discussed in the section IV.

2. Proposed method

The proposed method is used to recognize the object on it from scene sequence captured by the thermal camera [27]. The basic block diagram of the proposed work is illustrated in figure 1. The "VINCENT HD and Night vision IR camera" as shown in figure 2 acquires the video sequence with the resolution of 720p in night vision IR camera mode. The frames are extracted from the video sequences for the further processing. The various steps of the proposed approach for the recognition of the objects are illustrated in algorithm 1 and discussed in the following subsections.

2.1   Object Recognition

For the real time identification of the objects in the thermal images extracted from the video sequence by the "VINCENT HD and Night vision IR camera: Initially the noise (unwanted surroundings) is removed from the image and then objects are identified using Faster R-CNN [28] network.

Figure 2. The "VINCENT HD and Night vision IR camera" used for the thermal imaging

- Noise (unwanted surrounding) Removal Using HSV Segmentation

The unwanted surroundings of the object are considered as noise. In the thermal video frames the living objects appear brighter as they emit greater infrared radiations than other. Hence HSV segmentation [29] is applied to remove the unwanted surroundings. The image is converted to HSV color space form RGB color space as given below:

$$H = \begin{cases} 1 + (G - B)/\Delta \; (if \; R = max(R,G,B)) \\ 3 + (B - R)/\Delta \; (if \; G = max(R,G,B)) \\ 5 + (R - G)/\Delta \; (if \; B = max(R,G,B)) \end{cases}$$
$$S = [max(R,G,B) - min(R,G,B)]/max(R,G,B)$$
$$V = max(R,G,B) \tag{1}$$

Where $\Delta = [max(R,G,B) - min(R,G,B)]$ and (R, G, B) is the RGB color space at pixel location *(x,y)* in the image. At that point, the lower limit l_T and upper edge h_T range of the white color is characterized in HSV color space. The thresholding of the HSV color space is done to separate the white color in the image that addresses living objects.

$$I_{Seg}(x,y) = \begin{cases} 1 & (if \; l_T \leq I_{HSV}(x,y) \leq h_T) \\ 0 & else \end{cases} \tag{2}$$

Where, '1' and '0' addresses white and black tones respectively, $I_{HSV}(x,y)$ input image in HSV color space and $I_{Seg}(x,y)$ is output image after noise removal. Finally, the objects are segmented and the surrounding objects is taken out in the output image.

- Deep Network for Object Recognition

Faster R-CNN [28] is utilized to recognize objects in the real time thermal images. The Region Proposal Network (RPN) utilizes 'attention' system to tell the Fast R-CNN [30] detector network where to look. In this work, VGG-16 network model [31] is utilized in Faster R-CNN model [28]. In default arrangement, there are anchors at position of image having 3 scales and 3 ratios. The anchors with an IOU more than 0.7 are classified as foreground and the anchors don't cover any ground truth object (IOU under 0.3) are classified as background. The loss function F_{Loss} to be limited is given as:

$$F_{Loss}(\{m_v\}, \{n_v\}) = \frac{1}{C}\sum_v L_c(m_v, m_v^*) + \gamma \frac{1}{R}\sum_v m_v^* L_r(n_v, n_v^*) \tag{3}$$

Where, m_v is the predicted anchor probability of anchor v being an object. Vector n_v signifies the parameterized coordinates of predicted bouncing box. L_r and L_c are the

regression and classification loss respectively. For classification, Cross Entropy Loss is utilized and for Bounding Box Regression smooth L1 loss is utilized. The normalization parameters C and R are mini batch size (i.e. C=256) and anchor locations (i.e. R=2400). The term 'γ' is the weight balancing parameter set to 10. The Fast R-CNN network utilizes cross entropy loss in between numerous object classes with end to end back propagation and Stochastic Gradient Descent with momentum having value of 0.9.

3.   Result and discussion

For this work, VINCENT HD and Night vision IR camera is used to record the real time thermal video that has an approximate range of 1.5 km and resolution up to 720p but the video is recorded at a resolution of 360*450. The camera is set at channel number 101 assigned to night vision IR camera mode and IP address 192.168.1.65 is assigned to access camera remotely. The 749 frames are used evaluate the performance of proposed work as shown in Fig. 3. Faster R-CNN implementation is performed with configuration having Intel(R) Core(TM) i5-7200U CPU @ 2.50 GHz, 2701 MHz.

Figure 3. Extracted frames from the video sequence captured by thermal Night vision IR camera.

(a)   (b)

Figure 4 (a). Input frame **(b)** Surrounding removal performed at threshold values l_T =0.85 and h_T =1.

3.1 Object Recognition

HSV segmentation is used for the removal of the unwanted surroundings. As already discussed, living objects appear more white than other objects in thermal images because the living objects radiates more infrared rays than other objects. For the masking of white color space two thresholds (l_T and h_T) are used. As the obstacles appear whiter than others, so the upper threshold h_T is set to '1' but the lower threshold l_T is varied between 0.0-1.0 to get the optimum result. These results are recorded for HSV segmentation at threshold l_T=0.85, h_T =1 as shown in figure 4(b). Finally the Faster R-CNN network is applied to the output of the HSV segmentation to recognize objects in the thermal images. The performance of the network is revamped to enhance the accuracy by optimizing the parameter IoU threshold. The initial results of the proposed algorithm are obtained using threshold value 0.7 as shown in figure 5(c) but at the shorter distance from the camera. The lower IoU values provides better results in terms of timely identification but faced lower accuracy. The IoU threshold value '0.5' is used which timely recognize the objects with good accuracy as shown in figure 5(g). The final results for the IoU=0.7 and 0.5 are shown in Fig. 5(d) and 5(h) respectively.

| (a) | (b) | (c) | (d) |
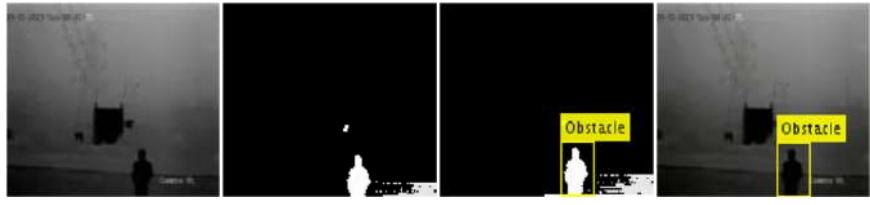
Figure 5 (a). Input frame **(b)** HSV segmented image after Surrounding removal **(c)** Faster R-CNN result with at IoU=0.7 **(d)**The final results of object recognition with Faster R-CNN at IoU=0.7


| (e) | (f) | (g) | (h) |
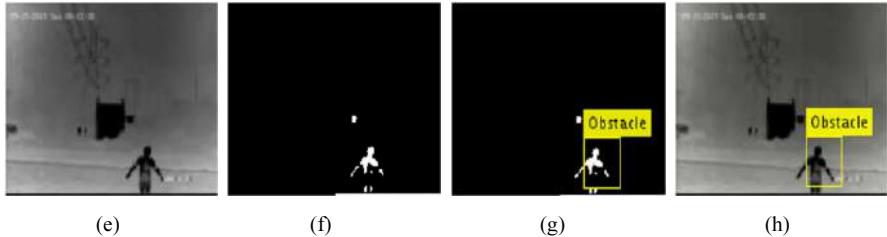
Figure 5 (e). Input frame **(f)** HSV segmented image after Surrounding removal **(g)** Faster R-CNN result with at IoU=0.5 **(h)**The final results of object recognition with Faster R-CNN at IoU=0.5

The proposed work performance is estimated with the help of parameters like accuracy, precision and recall. These parameters are evaluated for different values of the Intersection over Union (IoU) as shown in Table 1. The visual results and Table 1 concludes that the proposed system has better performance at IoU=0.5. Figure 6 illustrates the graphical comparison which also shows that the performance of this with accuracy of approximately 84.7%.

Table 1. The proposed method performance for different IoU

IoU	Accuracy	Precision	Recall
0.5	0.847	0.851	0.873
0.7	0.818	0.799	0.832
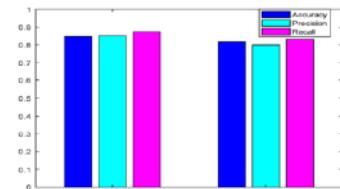
Figure 6. Performance at different values of IoU

4.   Conclusion and future scope

The proposed strategy turns out sufficiently for the object recognition real time thermal images. This paper introduced a novel methodology dependent on deep learning network to distinguish objects in images caught utilizing thermal imaging framework. This work can be robust technique to build up an early warning framework to forestall mishaps for the public safety upgrade. This strategy will be likewise cost effective as it doesn't need any huge change in the architecture as well as will decrease the monetaryin terms of accidental compensation. The framework will be extended for recognition of other object classes near the vehicle to improve the framework.

References

[1] Y. Wu., X. He, and T.Q. Nguyen, Moving Object Detection with a Freely Moving Camera via Background Motion Subtraction, *IEEE Trans. on Circuits and Systems for Video Technology* **27**(2), (2017), 236–248.

[2] R. Goel, A. Sharma, and R. Kapoor, State-of-the-Art Object Recognition Techniques: A Comparative Study. *Advances in Intelligent Systems and Computing (Springer)* (2020).

[3] J. Lee, J. Park, I. Kim, D.Y. Kang, Application of vision-based safety warning system to haeundae beach, *Korea. J. Coast. Res.* **91**(2019), 216–220.

[4] N. Hautiere, J. Tarel, J. Lavenant, and D. Aubert, Automatic fog detection and estimation of visibility distance through use of an onboard camera. *Machine Vis. and Applications (Springer)*, **17**(2006), 8–20.

[5] F. Suard, A. Rakotomamonjy, A. Bensrhair, A. Broggi, Pedestrian detection using infrared images and histograms of oriented gradients. *In Proceedings of the 2006 IEEE Intelligent Vehicle Symposium, Tokyo, Japan*, (2006), 206–212.

[6] N. Dalal, and B. Triggs, Histograms of oriented gradients for human detection. *In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA*, (2005), 886–893.

[7] K. Mikolajczyk, C. Schmid, A. Zisserman, Human detection based on a probabilistic assembly of robust part detectors. *In Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, (2004), 69–82.

[8] Y. Fang, K. Yamada, Y. Ninomiya, B.K.P. Horn, I. Masaki, A shape-independent method for pedestrian detection with far-infrared images. *IEEE Trans. Veh. Technol.* **53**(2004), 1679–1697.

[9] H.K. Ragb, V.K. Asari, T.H. Aspiras, Human detection in infrared imagery using intensity distribution, gradient and texture features. *In Proceedings of the Mobile Multimedia/Image Processing, Security, and Applications 2018, Orlando, FL, USA*, (2018).

[10] F. Xu, X. Liu, K. FujiMura, Pedestrian detection and tracking with night vision. *IEEE Trans. Intell. Transp. Syst.* **6**(2005), 63–71.

[11] L. Zhang, B. Wu, R. Nevatia, Pedestrian detection in infrared images based on local shape features. *In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA*, (2007), 1–8.

[12] C.H. Setjo, B. Achmad, F. Djoko, Thermal image human detection using Haar-cascade classifier. *In Proceedings of the 7th International Annual Engineering Seminar (InAES), Indonesia*, (2017), 1–6.

[13] A. Neves, and R. Ribeiro, Algorithms for face detection on infrared thermal images. *Int. J. Adv. Softw.* **10**(2018), 499–512.

[14] S. Guennouni, A. Ahaitouf, A. Mansouri, A comparative study of multiple object detection using haar-like feature selection and local binary patterns in several platforms. *Model. Simul. Eng.*, **8**(2015).

[15] M. Hasan, S. Ullah, M.J. Khan, K. Khurshid, Comparative analysis of SVM, ANN and CNN for classifying vegetation species using hyperspectral thermal infrared data. *In Proceedings of the International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, Enschede, The Netherlands*, (2019), 1861–1868.

[16] O. Yardimci, B.C. Ayyıldız, Comparison of SVM and CNN classification methods for infrared target recognition. *In Proceedings of the Automatic Target Recognition XXVIII, Orlando, FL, USA*, (2018).

[17] J.H. Kim, H.G. Hong, K.R. Park, CNN-based human detection in nighttime images using visible light camera sensors. *Sensors*, **17**(2017).

[18] H. Qu, L. Zhang, X. Wu, X. He, X. Hu, X. Wen, Multiscale object detection in infrared streetscape images based on deep learning and instance level data augmentation. *Appl. Sci.*, **9**(2019).

[19] B. Khalid, A.M. Khan, M.U. Akram, S. Batool, Person detection by fusion of visible and thermal images using convolutional neural network. *In Proceedings of the 2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE), Islamabad, Pakistan*, (2019).

[20] C. Herrmann, M. Ruf, J. Beyerer, CNN-based thermal infrared person detection by domain adaptation. *In Proceedings of the Autonomous Systems: Sensors, Vehicles, Security and the Internet of Everything, Orlando, FL, USA*, (2018).

[21] J. Liu, S. Zhang, S. Wang, D.N. Metaxas, Multispectral deep neural networks for pedestrian detection. *arXiv* (2016).

[22] A.D'Acremont, R. Fablet, A.Baussard, G. Quin, CNN-based target recognition and identification for IR imaging in defense systems. *Sensors*, **19**(2019).

[23] S. Hwang, J. Park, N. Kim, Y. Choi, I.S. Kweon, Multispectral pedestrian detection: Benchmark dataset and baseline. *In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, USA*, (2015), 1037–1045.

[24] A. Gonzalez, Z. Fang, Y. Socarras, J. Serrat, D. Vazquez, J. Xu, A.M. Lopez, Pedestrian detection at day/night time with visible and FIR cameras: A comparison. *Sensors*, **16**(2016).

[25] A. Khellal, H. Ma, Q. Fei, Pedestrian classification and detection in far infrared images. *In Proceedings of the International Conference on Intelligent Robotics and Applications, UK*, (2015), 511–522.

[26] R. Goel, A. Sharma, R. Kapoor, Object Recognition Using Deep Learning. *Journal of Computational and Theoretical Nanoscience*, **16**(2019).

[27] S. Singh, A. Khosla, R. Kapoor, Object Tracking with a Novel Visual- Thermal Sensor Fusion Method in Template Matching. *International Journal of Image, Graphics & Signal Processing*, **7** (2019), 39-47.

[28] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans.on Pattern Analysis and Machine Intelligence*, **39**(2017), 1137–1149.

[29] D. D. Burdescu, M. Brezovan, E. Ganea, L. Stanescu, A New Method for Segmentation of Images Represented in a HSV Color Space. *Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems (Springer)*, (2009), 606-617.

[30] R. Girshick, Fast R-CNN. *Proceedings of IEEE International Conference on Computer Vision*, (2015).

[31] S. Tammina, Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. *International Journal of Scientific and Research Publications*, **9**(2019).

Computer Vision Approach for Detecting Adulteration of Ghee with Foreign Fats – A Survey

Aditya UPADHYAY [a, 1], Dr. Neha CHAUDHARY [a]

[a] *Manipal University Jaipu, India*

Abstract

Ghee is pure clarified fat derived from milk, yogurt and fresh cream. It is most commonly used milk fat product in India. The consumption and production of ghee is consistently increasing by 10% in our country in every year. In comparison to other milk fat product, ghee is expensive and short in demand because of its pleasant taste or high nutrition value. Due to its high cost and demand in market, there are high possibilities to adulterate it with cheap fats like vegetable oil/animal body fats. The adulteration detection of ghee is becoming a serious issue to chemists. Several analytical and instrumental methods are available for the detecting adulteration in ghee based on chemical principles. On the basis of study, it was observed that analytical methods are not suitable to detect the adulteration level of <15%. In recent time, digital image analysis is introduced in the field of adulteration detection in food products. A very few studies found in the area of milk fat adulteration detection with foreign fats using image analysis. Various studies found related to detection of adulteration in Oils (like Extra virgin olive oil, sesame oil etc.) with cheap oil using the various color models (like CIELAB, RGB, HSV, CMYK) and machine learning algorithms.

keywords: - Ghee adulteration, Image analysis, Oil adulteration, Computer Vision, image analysis, physico-chemical, milk fat adulteration

1. Introduction

Clarified butter, or Ghee (Originated from Sanskrit word Ghrta), is the most widely used lipophilic dairy product (98.9% lipids) in South Asian countries [2]. Mostly cow milk, buffalo milk is used to prepare ghee. Ghee is mostly used for frying and in cooking food because of its pleasant taste, delicious aroma and high nutrition value. It is also used in the preparation of traditional medicines and performing religion rituals. According to reports, 35-40% of produced milk is converted into Ghee and 170 thousand metric tons ghee produced in fiscal year 2020. But in summer season in India [3], a very complex situation arises due to fall in Milk production. In comparison to other dairy products, Ghee is expensive and high in demand. Therefore, it is highly susceptible to adulterate with Cheap vegetable fats, animal body fat and inedible Oils.

---

[1] Aditya Upadhyay, Department of Computer Science, Manipal University Jaipur, Rajasthan, India;
E-mail: adityaupadhyay144@gmail.com.

Other motives for adulterating Ghee are gaining more profit and to fill the gap between demand and supply. The detection of adulteration in ghee is very important because it is directly related to our health. Adulterated ghee contains hydrogenated fats, essence colors, carcinogenic chemicals that affects immune system, increase bad cholesterol, increases the risk of cancer.

2.   Methods for Preparing Ghee

There is various method used commercially for preparing a ghee [7].

Indigenous Milk-Butter Process: - This method is mostly used at domestic level. In this method, Makkhan is used made by direct churning of yogurt at room temperature. Makkhan is placed in a vessel or ghee boiler and heated to about 110-120 ºC with constant stirring in order to evaporate all the moisture. After that, the ghee is filtered using muslin cloth and packaged in containers.

Direct Cream heating Method: - In this method, fresh cream is heated in a stainless-steel jacketed ghee vessel and heating end point can be determined manually based on the color of ghee residue. This method needs long heating time to remove moisture. After this, ghee is separated from ghee residue by the pressure filters and centrifugal separators. Then ghee is allowed to cool to 40-42 ºC and then packaged in individual containers or tins.

Creamery Butter method: - In this method, butter is melted in a butter-Melter at 60ºC, and then pumped into a ghee boiler. In the boiler, the moisture in butter is evaporated with constant stirring to reach a temperature at 110-120 ºC. Further steps are same as above mentioned.

For all the above-mentioned methods, 110-120 ºC temperature is suggested as clarification temperature for producing good quality of ghee. The crystalline phenomenon of high melting triglycerides and granularity of ghee is responsible for its purity and quality. The micro structure and textural properties of ghee are based on factors like crystallization, the existence of phospholipids and additives, milk fatty acid profile, clarification temperature followed by rate of its cooling.

Chemically, ghee is a complex lipid of glycerides, phospholipids, sterols, sterol esters, fat soluble vitamins, carbonyls, free fatty acids, hydrocarbons, carotenoids (available in cow ghee only). Glycerides constitute 98% of the total material. Granularity is also a measure of good quality ghee. A pure ghee should have a soft and granular texture and granularity of ghee appeared due to high melting triglycerides (HMT). Pure ghee texture is better because of bigger grain size and higher proportions (9.0–12.0%) of high melting triglycerides (HMT). Commercially available ghee has poor granularity. There are various other factors on which the quality of ghee depends, such as

- Animal feed type
- Raw material used for producing ghee– makkhan, cream, white butter
- Method of preparation
- Clarification temperature
- Storage conditions of ghee

3.   Various fats used for Adulteration

Due to unique taste, high demand and high nutritional value, Ghee is expensive in comparison to other edible fats available in the market. It increases the chances of mixing cheap fats for economic advantages by traders. Various fats are used as an adulterant like: -

- vanaspati fats/oil, Refined oil like palm oil, Soyabean oil, coconut oil.
- animal body fats like lard or tallow
- mixture of animal body fats and plant oil
- mineral oils- solid paraffin and liquid paraffin
- Colors

4.   Adulteration Detection Method for Ghee

For detecting adulteration in ghee by foreign fats, several analytical and instrumental methods are available. These methods are based on the determining of physical and chemical constants, specific fatty acids, fatty acid composition, glyceride structure and color reactions etc. Analytical methods [1][15], based on chemical principles, are Melting Point, Apparent Solidification time (AST) test, Baudouin test, Butyro Refractometer (BR) Reading, Critical temperature of dissolution (CTD), Reichert Meissl (RM) value, Iodine value, Polenske value, Saponification Value etc. Instrumental methods [19] like High Pressure Liquid Chromatography (HPLC), Spectroscopy, RP-TLC, Gas Chromatography (GC) are costly and need technical skills. Gandhi et. al. (2014) [20] coupled solvent fraction technique with RM value for the detecting presence of palm oil and animal body fat in ghee.

For the state of Rajasthan except Jodhpur district [17], FSSAI has prescribed the following standards for the quality of produced ghee: BR reading at 40C should be in the range of 40-43, minimum RM value should be 26, Baudouin test Negative, Maximum Moisture 0.5%, free fatty acid (oleic acid) 3%. Analytical methods are not capable in identifying all adulterants with different level of adulterations. Physico chemical constant values also varying geographically and seasonally. In Table 1, Physico chemical constants of pure cow and buffalo ghee are mentioned seasonally [7]. In Table 2, the values Physico-chemical parameters of pure cow, buffalo ghee and their mixture mentioned [7].

Table 1. Physico-Chemical constants of Cow and Buffalo pure ghee in different months of Year

Physico Chemical constants	Pure Cow ghee						Pure Buffalo Ghee					
	Jan	March	May	July	Sept	Nov	Jan	March	May	July	Sept	Nov
RM	29.63	29.28	28.52	29.48	29.82	30.93	33.39	31.70	30.87	32.40	33.42	33.73
BR	41.5	42.1	42.5	41.0	41.2	40.7	40.7	41.5	41.6	40.4	40.6	40.2
PV	1.7	1.5	1.1	1.4	1.3	1.4	1.6	1.2	1.0	1.2	1.1	1.4
IV	36.5	39.6	40.2	34.4	35.9	32.8	34.6	36.2	37.3	33.2	32.3	30.1
SV	229.12	228.14	226.07	231.50	228.00	230.00	232.57	230.75	230.76	232.82	231.66	233.63

Table 2. Physico-Chemical constants of Pure Cow ghee, pure Buffalo ghee & (Cow+Buffalo) Ghee

Physico Chemical Constant	Pure Cow Ghee	Pure Buffalo Ghee	Cow + Buffalo Ghee
RM Value	29.61±0.32	32.59±0.46	31.10±0.74
BR Reading	41.50±0.29	40.83±0.25	41.16±0.20
PV Value	1.40±0.08	1.25±0.09	1.33±0.06
Iodine Value	36.61±1.18	33.98±1.09	35.29±0.86
Saponification Value	228.81±0.76	232.03±0.48	230.42±0.64
CTD	51.6-54.6	52.4-56.2	NA
AST (min-sec) at 18ºC	2-56 to 3-26	2-30 to 2-48	NA

5.   Digital Image Analysis for food adulteration detection

An image is made up of pixels arranged in matrix and these pixels can be seen by zooming an image [18]. A numerical value is assigned to each pixel in the range from 0(black) to 255(White) representing pixel intensity value. All intermediate value between the range of 0-255 are different grey shades from black to white. It is expected that the mixing of any extraneous material, may changes pixel intensity which may changes the surface texture quantifiers of digital image of the food sample. This shows a scope in the domain of analysis of digital image as a technique for checking the quality of food products. In literature, many studies have used digital image analysis techniques for detection of adulteration in daily used food products like Butter, Cheese, Ghee, Ketchup, Oils, species, meat, milk, grains etc.

Lukinac et. al. (2018) [16] reviewed the application of digital image analysis in cheese quality evaluation. Reile et. al. (2020) [14] used the concept of color histograms with SPA-LDA for classification to detect the adulteration in ketchup with Sudan I dye. Sano et. al. (2003) detected adulteration in roast coffee powder using digital image processing. Justin et. al. (2017) [21] proposed an approach for the detection of adulteration in food items like oil, chilli powder, black pepper, milk and processed the images using MATLAB to extract feature like Mean, Standard deviation, Entropy, Smoothness, Texture feature using local tetra pattern.

In image processing [18][4], meaningful information of our use can be extracted from processed image by adding some values and performing mathematical operations on raw image. In this process, only rearranging of data is done. The area of image analysis lies between image processing and computer vision, without a clear-cut boundary.

6.   Literature Review

With the advancement of technology, food adulteration detection methods are also changed with time. In now days, digital image analysis has become a relevant tool in food science. A very few studies have been found in the area of ghee adulteration detection using image analysis. Based on the studies, it is observed that image analysis

methods provide better results than physico chemical methods in the detection of adulteration of ghee especially at levels ≥15%.

Wasnik et. al. (2019) [3] proposed a model for detection of adulteration level of vegetable fat in cow ghee using image analysis. A concept of multiple regression was used to detect adulteration level with four selected features (pixel intensity, branches, luminance and equivalent diameter). For the study, samples of cow ghee with different level of adulteration (0%, 5%, 10%, 15%, 20%) with vegetable fat were prepared. Images were captured using flatbed scanner (Canon Canoscan 9000F Mark II) in transmission mode and saved in TIFF format. Total 105(5 samples x 3 replications x 7 days) captured images were analyzed using ImageJ software and Adobe photoshop in L-a-b mode.

Wasnik et. al. (2017 a) [1] derived skeleton parameters (branches, junctions, end point voxels), texture parameters (fractal dimension) and morphological (pixel intensity, particle count, equivalent diameter) of 105 images using ImageJ software. Using the defined image analysis protocol, results shows that pixel intensity, fractal dimension and skeleton parameters were found vary significantly with coefficient of determination >0.95 to detect level of adulteration.

Wasnik et. al. (2017 b) [2] extracted particle count, equivalent diameter and color parameters of 105 sample images using ImageJ software and Adobe photoshop in lab mode respectively. CIELAB parameters (L*, a*, b*) were calculated using the value of L, a, b parameters. From the extracted values of CIELAB (L*, a*, b*), color descriptors like hue, chroma, yellowness index and whiteness index were calculated. The result shows that particle count, luminance, equivalent particle diameter and whiteness index can be considered for detecting adulteration level of vegetable fat in cow ghee.

Ayari et al. (2018) [12] used electronic nose (e-nose) with PCA and ANN for detecting adulteration in cow ghee with sunflower oil and cow body fat. E-nose devices mimic the sense of smell and uses a sensor's array to detect and differentiate odors precisely at low cost. For study, Test samples were created by adding sunflower oil and cow body fat in cow ghee in proportions of 10%, 20%, 30%, 40%, 50% with 15 replications for each adulteration level. The result showed that ANN correct classification rate was 82.5% and 91.3% for cow body fat and sunflower oil, respectively.

Karla et. al. (2015) [13] proposed a methodology based on images analysis and multivariate classification based on discriminant analysis for the detection and classifying extra virgin olive oil (EVOO) samples with respect to brand and adulteration with soybean oil. For classification purpose, RGB color space is used with linear discriminant analysis (LDA) in association with successive projections algorithm (SPA). With full histogram, partial least squares-discriminant analysis (PLS-DA) was employed. For the study, total 53 EVOO samples was collected with three different brand and 99 adulterated samples with soybean oil having different level of adulteration.

Mahdi Rashvand et. al. (2019) [11] used ANN with different hidden layers for the detection of adulteration in Sesame Oil with sunflower oil and Canola oil. MATLAB was used for image processing and RGB, CMY, L*A*B* color components were used for the study.

Marchal et al. (2013) [9] proposed an approach for detecting adulteration in olive oil samples based on computer vision and ANN. For the study, the authors used three different input parameters which was the histogram of every channel of the image in three color spaces: CIELAB, RGB, and HSV, and four feature extraction methods- Principal Component Analysis, Kernel PCA, Linear Discriminant Analysis and Kernel LDA.

Karagiorgos et. al. (2017) [8] developed image processing algorithm for detecting adulteration of soybean oil in olive oil from captured image. Later this algorithm was planned to be implemented into mobile application of smart phones, where anyone can measure the olive oil quality by taking the picture of samples by smart phones. For determining the adulteration percentage in sample, the captured image was divided into two parts- first part contains the oil sample area and another part contains the rest of the image. For the study, 11 different adulterated samples were prepared having 0% to 100% with the step of 10% and three images captured for each sample. CIELAB color space was used to design a model for the prediction of adulteration percentage with 3% to 5% absolute error.

Table 3 Review Studies

Author	Base product	Adulterant	Methods	Accuracy
Wasnik et al. (2019)	Ghee	Vegetable fat	Multiple regression	92.2%
Ayari et al. (2018)	Ghee	Sun flower oil and cow body fat	electronic nose method with PCA and ANN	91.3% (for sun flower oil), 82.5% (for cow body fat)
Karla et al. (2015)	Extra virgin olive oil	Soybean oil	SPA-LDA and PLS-DA	--100% according to brand (A, B, C) --For brand A, 88%(SPA-LDA), 94%(PLS-DA) -- For brand B and C, 100% in both models
Mahdi Rashvand et al. (2019)	Sesame Oil	Sunflower oil and Canola Oil	ANN with different hidden layer	12-6-7 for Sunflower oil, 12-10-7 for canola oil
Marchal et al. (2013)	Virgin Olive Oil	Impurities	ANN, SVM	87.66% (SVM with KPCA), 82.38% (ANN with PCA)

7.   Conclusion

Milk fat adulteration is a serious issue for chemist in now days because it is directly related to our health. Several analytical and instrumental methods are available for detecting adulteration in milk fat or ghee. But these methods are not capable to detect all kind of adulterants. The values of physico- Chemical parameters like RM value, BR reading, CTD, AST, PV, IV, SV is influenced by season, storage conditions and many more factors. Image analysis is new tool in the domain of food adulteration detection and provide better results in comparison to existing methods.

References

[1] Wasnik PG, Menon RR, Surendra Nath B, Balasubramanyam BV, Manjunatha M, Sivaram M (2017a) Application of particle analysis and colour parameters for detection of adulteration of cow ghee with vanaspati derived from image analysis. Indian J Dairy Sci 70(2):200–208

[2] Wasnik PG, Menon RR, Surendra Nath B, Balasubramanyam BV,Manjunatha M, Sivaram M (2017b) Application of pixel intensity, fractal dimension and skeleton parameters for detection of adulteration of cow ghee with vanaspati derived from image analysis. Indian J Dairy Sci 70(3):331–337

[3] Wasnik PG, Menon RR, Surendra Nath B, Balasubramanyam BV, Manjunatha M and Sivaram M (2019) Development of mathematical model for prediction of adulteration levels of cow ghee with vegetable fat using image analysis. J Food Sci Technol 55: 2320–2325

[4] Yam KL, Papadakis S (2004) A simple digital imaging method for measuring and analyzing colour of food surfaces. J Food Eng 61(1):137–142

[5] Ropadi AI, Pavlidi DE, Mohareb F, Panagou EZ, Nychas GJE (2015) Multispectral image analysis approach to detect adulteration of beef and pork in raw meats. Food Res Int 67:12–18

[6] Sano EE, Assad ED, Cunha SAR, Correa TBS, Rodrigues HR (2003) Quantifying adulteration in roast coffee powders by digital image processing. J Food Qual 26:123–134

[7] Kumar, A. (2008) Detection of adulteration in ghee. Ph.D. Thesis submitted to National Dairy Research Institute (Deemed University), Karnal, India.

[8] Karagiorgos N, Nenadis N, Trypidis D, et al. (2017) An approach for estimating adulteration of virgin olive oil with soybean oil using image analysis. Published in: 2017 6th International Conference on Modern Circuits and Systems Technologies (MOCAST), Thessaloniki, Greece, 1–4.

[9] Marchal PC, Gila DM, García JG, et al. (2013) Expert system based on computer vision to estimate the content of impurities in olive oil samples. *J food Eng* 119: 220–228.

[10] Silva SF, Anjos CAR, Cavalcanti RN, et al. (2015) Evaluation of extra virgin olive oil stability by artificial neural network. *Food Chem* 179: 35–43.

[11] Mahdi Rashvand, Abbas Akbarnia. The feasibility of using image processing and artificial neural network for detecting the adulteration of sesame oil[J]. AIMS Agriculture and Food, 2019, 4(2): 237-243. doi: 10.3934/agrfood.2019.2.237

[12] Ayari F, Mirzaee-Ghaleh E, Rabbani H, Heidarbeigi K. Detection of the adulteration in pure cow ghee by electronic nose method (case study: sunflower oil and cow body fat). Int J Food Prop 2018; 21(1):1670–9; https://doi.org/10.1080/10942912.2018.15057 55

[13] K.D.T. De Melo Milanez, M.J.C. Pontes, Classification of extra virgin olive oil and verification of adulteration using digital images and discriminant analysis, Anal. Methods. 7 (2015) 8839–8846. doi:10.1039/c5ay01765c

[14] Reile CG, Rodríguez MS, Fernandes DDS, Gomes AA, Diniz PHGD, Di Anibal CV. Qualitative and quantitative analysis based on digital images to determine the adulteration of ketchup samples with Sudan I dye. Food Chem. 2020 Oct 30;328:127101. doi: 10.1016/j.foodchem.2020.127101. Epub 2020 May 20. PMID: 32480258.

[15] Hazra T, Sharma V, Saha P and Parmar M (2017a) Physico-Chemical Properties Analysis Based Approaches to Ascertain the Purity of Ghee- A mini Review. Int J Sci Env 6: 899 – 907

[16] Lukinac J. Jukić M. Mastanjević K. Lučan M. 2018 Application of computer vision and image analysis method in cheese-quality evaluation: a review Ukrainian Food Journal 7(2) pp 192-214 DOI: 10.24263/2304-974x-2018-7-2-4

[17] https://www.fssai.gov.in/upload/uploadfiles/files/Compendium_Food_Additives_Regulations_29_03_2 019.pdf

[18] Wasnik, P.G.; Menon, R.R.; Meshram, B.D. Digital Image Analysis: Tool for Food Quality Evaluation.In Processing Technologies for Milk and Milk Products, 1st ed.; Agrawal, A.K., Goyal, M.R., Eds.; Apple Academic Press: Florida, FL, USA, 2017; pp. 65–104.

[19] Anupama R, Sharma V, Arora S, Lal D, Anil K (2015) A rapid reversed phase thin layer chromatographic protocol for detection of adulteration in ghee (clarified milk fat) with vegetable oils. J Food Sci Technol 52(4): 2434-24

[20] Gandhi K, Upadhyay N, Aghav D, Sharma V, Lal D (2014) Detection of adulteration of ghee (clarified milk fat) with palmolein and sheep body fat using Reichert-Meissl (RM) value coupled with solvent fractionation technique. Indian J Dairy Sci 67 (5): 387-393.

[21] A. Justin Diraviam, G. Angel and K. Rajappan, 2017. A Novel Method in Detection of Food Adulteration in Selected Food Items using Digital Image Processing Technology with Digital Camera. Asian Journal of Information Technology, 16: 374-382.

Aero Engine Performance Monitoring Using Least Squares Regression and Spectral Clustering

Kunaal SAXENA [a,1] and Manisha J NENE [a]

[a] *Department of CSE, Defence Institute of Advanced Technology, Pune, Maharashtra*

Abstract. Threshold-based flight data recorder analysis techniques have been widely used across the aerospace industry for fault detection and accident prevention. These techniques can detect pre-programmed events but fail to capture unknown patterns in the dataset. This research proposes a machine learning (ML) algorithm to analyze and detect unusual aero engine performance of a turboshaft engine mounted on a single engine rotorcraft. The performance is first modelled from an FDR dataset consisting of hundred flights, using least squares regression (LSR). A technique to scale the model by adding flight data from subsequent flights is thereafter discussed. Spectral Clustering is used for testing and validating the hypothesis derived from the regression model, by employing synthetically generated FDR data for twenty-five flights.

Keywords. Machine Learning, aero engine, FDR, Least Squares Regression, Spectral Clustering.

1. Introduction

1.1. FDR Analysis

The flight data recorder (FDR) is an integral part of all modern aircraft. It records various aircraft and aeroengine parameters every second during flying and stores it in the FDR database. The analysis of this data from the FDR, is termed as Flight Data Monitoring (FDM). The purpose of FDM is to monitor the performance of an aircraft accurately and predict its serviceability. Most aircraft employ a threshold-based fault detection system for the same, which reports an event as a fault only when the threshold for an FDR parameter is triggered.

1.2. Problem Statement

Multiple facets to performance monitoring of aeroengines such as unusual fuel consumption, unusual take off power, abnormal gas generator rotation speed etc. go unnoticed by the existing threshold-based analysis.

---

[1] Kunaal SAXENA, Department of CSE, Defence Institute of Advanced Technology, Pune
Email: motto877@gmail.com.

This is because the present FDM system analyzes FDR data only from a single flight at a time and is not able to observe any changes in the trend of the operating parameters, from previous flights. This study aims to overcome this limitation by utilizing ML techniques, as they are able to draw inferences from a large number of flights and thus capable of observing the minutest of changes in the trend of aircraft and aero engine performance.

1.3. Literature Review

Until now, FDR analysis has not utilized ML techniques on a comprehensive basis [1]. However, over the years, efforts have been made towards research in ML for FDM. Chunxiao Zhang et al. proposed the use of support vector machines (SVM) regression to monitor aero-engine health and condition by comparing flight data of 100 flights with an engine baseline model, in [2]. Li-Hua Ren et al. proposed a data-driven method for modelling the aero-engine aerodynamic performance by combining stochastic gradient descent with SVM in [3]. The studies undertaken in both [2] and [3] utilize SVM Regression, which suffers from a major drawback of higher computational burden because of the required constrained optimization programming [4].

Lishuai Li et al, in [5] have carried out great research in the field of cluster-based anomaly detection [6] for detecting abnormal flight operations, using the technique of density based spatial clustering of applications with noise. Julian Oehling et al. used local outlier probability detection method to discover anomalies in flight performance from FDR data and demonstrated scaling up the method for 1.2 million flights in [6]. Although both techniques discussed in [5] and [6] were made scalable to accommodate additional flight data, they utilized unsupervised learning, which only helps in identifying anomalies and local outliers, without modelling the actual performance of an aircraft.

According to a study undertaken by Clinton et al. in [7], 64 percent of fatal aviation accidents due to material failure are caused because of failure of the aeroengine. The study undertaken in this paper therefore focuses on monitoring the performance of an aero engine. Least squares regression supervised learning is used to model the performance of an aero engine installed on a rotorcraft, from an FDR dataset consisting of 100 flights. The learned model is then used to detect unusual aero engine performance from a set of 25 synthetically generated flight datasets, using spectral clustering. The term 'unusual' refers to an aero engine performance that does not adhere to the inferences drawn from the developed ML model. The remainder of this paper is structured as follows. Section 2 discusses the overall workflow adopted for the research. Section 3 describes the implementation of regression modelling to discover relationships between the engine related FDR parameters. Further, a method to scale the regression algorithm is discussed, followed by the anomaly detection technique used to detect unusual engine performance. Section 4 discusses the results obtained from the applied regression-based model and anomaly detection algorithm. Section 5 concludes.

2. ML Workflow

Figure 1 depicts the overall workflow for building a ML model for learning the normal aeroengine performance. The aircraft FDR data is obtained from historical

database. The database consists of a set of dependent and independent features. Only the aeroengine related features are selected and passed through a feature selection algorithm, to identify the degree of correlation between them. The subject data is preprocessed using a feature scaling algorithm, to scale all features in the dataset, within a uniform range. Using ordinary least squares regression technique, the data is mined to produce a regression-based model. The subject model is then used on test flight datasets to detect anomalies.
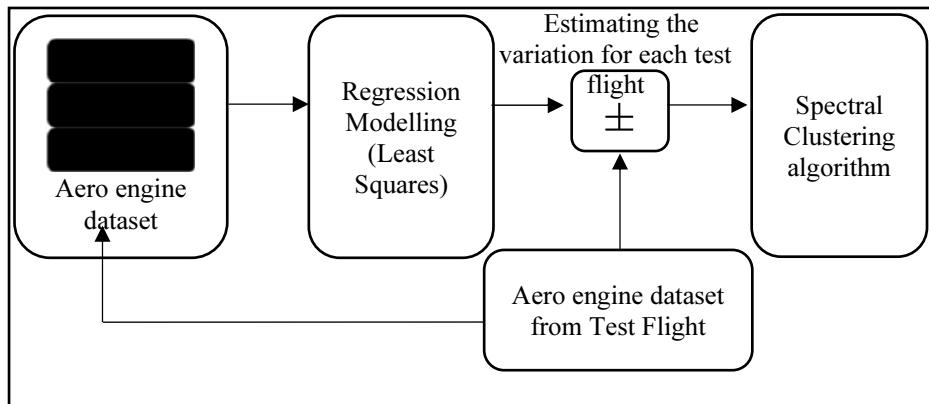
Figure 1. Workflow depicting the proposed ML technique

3. Implementation of Regression Modelling and Anomaly Detection Algorithms

3.1. Regression modelling

Aero-engine performance of turboshaft engines can be ascertained by the relative performance of some of its key parameters such as the gas generator turbine rotation speed (N1) and the generated torque (Q). Subject parameters depend upon the opening of the air flow inlet guide vanes (GV), fuel Flow (FF) into the combustion chamber and the engine exhaust gas temperature (EGT). Thus, we get a set of dependent and independent features of the FDR dataset, which can be defined by the relationship: -

$$y = F(x) + e \qquad\qquad \text{Eq. (1)}$$

Here, 'y' represents the set of dependent aeroengine features of N1 and Q, while 'x' represents the set of independent features GV, FF and EGT in the data set. The first objective of the regression problem is to identify the function F. In terms of a regression problem, we write,

$$y = F. \, \phi(x) + £ \qquad\qquad \text{Eq. (2)}$$

where, 'ɸ(x)' represents a matrix of independent engine features, 'F' represents the matrix that defines the relationship between 'y' and 'ɸ(x)'. '£' represents some noise which may be present in the FDR dataset, and is used as an error reducing coefficient for the regression overfit. Then, the matrices 'y' and 'ɸ(x)' are constructed.

$$
\begin{pmatrix}
N1\ t1 & Q\ t1 \\
N1\ t2 & Q\ t2 \\
\vdots & \vdots \\
\vdots & \vdots \\
\vdots & \vdots \\
N1\ tn & Q\ tn
\end{pmatrix}
\qquad
\begin{pmatrix}
EGT\ t1 & FF\ t1 & GV\ t1 \\
EGT\ t2 & FF\ t2 & GV\ t2 \\
\vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots \\
EGT\ tn & FF\ tn & GV\ tn
\end{pmatrix}
$$

$$\text{Matrix y} \qquad\qquad t\,n \times 2 \qquad\qquad \text{Matrix } \phi(x) \qquad\qquad t\,n \times 3$$

The FDR dataset contains data for approximately 100 hours of flying, thus providing data for more than 3 lakh time samples, represented by the indices 't1' to 'tn' in the matrix. The solution for finding 'F' is obtained using the regularized least squares approximation technique [8].

$$F = inv[(\phi(x)^T\,\phi(x)) + \pounds].\,\phi(x)^T\,y \qquad\qquad\qquad \text{Eq. (3)}$$

The developed model 'F' models the performance of the aero engine i.e., it defines the true relationship between the dependent and independent features of the FDR dataset, with respect to the aero engine.

3.2. Scaling for additional flight data

The proposed technique is capable of handling large FDR data sets as well as continuously updating the model 'F' with new flight data, by updating the matrices 'ɸ(x)' and 'y'. Let 'N' be the number of flights for which the model 'F' was initially calculated and let and let 'D_N' be the number of data points for each of the 'N' flights. Then, for the $(N+1)^{th}$ flight, the model 'F' needs to be updated with data points $D_{(N+1)}$ from the $(N + 1)^{th}$ flight. Updating the model with new flight dataset only requires the computation of the products as shown below, and then adding the obtained products to Eq. (3).

$$\phi(x)_{(N+1)}^T\,\phi(x)_{(N+1)} \qquad\qquad\qquad \text{Eq. (4)}$$

$$\phi(x)_{(N+1)}^T\,y_{(N+1)} \qquad\qquad\qquad \text{Eq. (5)}$$

At each step, dataset of the new flight is added to the existing dataset by calculation of only the inner transpose products. In this way, the data processing technique is made scalable to include FDR datasets from all subsequent flights, without much additional computational overheads. This technique enables the derived

regression model 'F' to keep updating, correcting and tuning itself to the subsequent flights' aeroengine performance.

3.3. Anomaly Detection

Once the model 'F' has been identified, the aim is to use it to discover anomalies in aero engine performance from test flight data. The anomaly detection algorithm also provides a way of evaluating the developed regression model. The subject test data has been generated synthetically and consists of a set of 10 flights with derated aeroengine performance and 15 flights showing normal performance. The first step is to find the variation 'V' between each set of test flight data and trained data.

$$V_{i} = y_i - F \, \varphi(x)_i \qquad\qquad\qquad \text{Eq. (6)}$$

Here, V_i represents the variation observed in the ith test flight. Variations for all sets of test flight data are calculated sequentially. These are applied to a spectral clustering algorithm, enabling separating out the unusual aero engine performances, which are marked as outliers and require further inspection and investigation by aviation experts. Spectral clustering uses the connectivity approach to clustering, wherein communities of nodes (i.e. data points) that are connected are identified in a graph. These data points are then mapped to a low-dimensional space that are segregated to form clusters [9]. A detailed explanation of the spectral clustering algorithm, however, is beyond the scope of this paper.

4. Results and Discussions

The FDR dataset consists of the multiple sets of features, out of which the engine related features include N1, Q, GV, FF, EGT, and some general features like altitude (Alt) and the indicated air speed (IAS). The right set of features to be utilized for regression modelling were selected using the Spearman's rank correlation coefficient.

The features GV, EGT and FF showed a high degree of correlation (>0.6) with the dependent features Q and N1 of the FDR dataset. Therefore, those features have been chosen as independent and dependent set of features, respectively. Regression modelling for the selected features yields the matrix 'F' of the order of 3x2.

Table 1. Regression Model Parameters (Matrix 'F')

	Q	N1
GV	0.96534358	0.7342996
FF	0.34389898	0.61503091
EGT	-0.31262837	-0.3499808
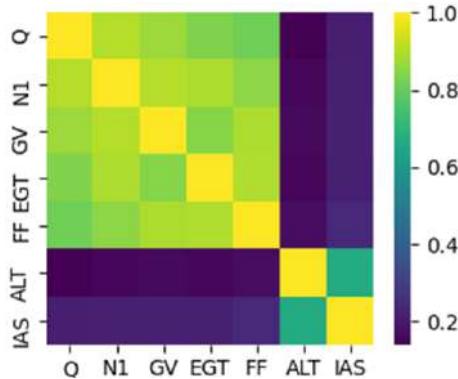
Figure 2. Correlation between FDR dataset features
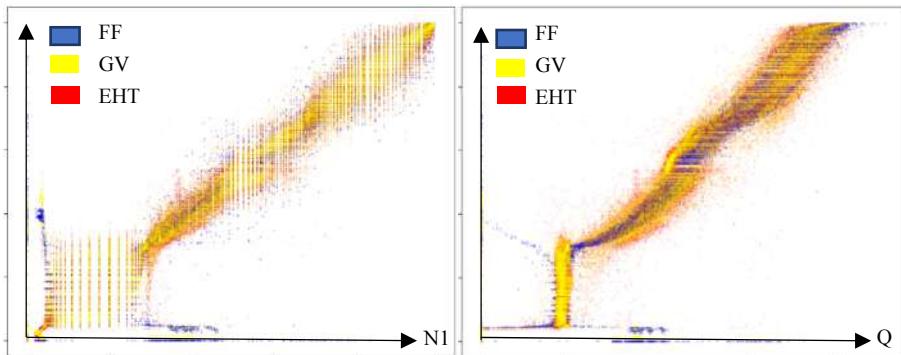
Figure 3. Linear relationship of N1 and Q with engine parameters FF, GV and EHT
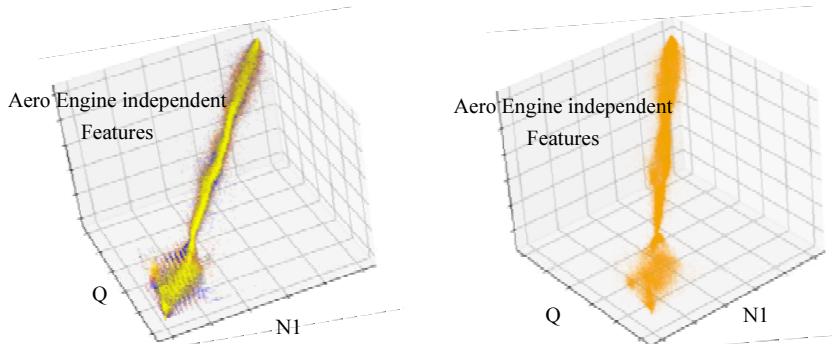
Figure 4. 3D plot showing relationship of Q and N1 with independent aero engine features GV, FF and EGT and the developed regression model on the right.

A set of twenty-five Test Flight datasets were applied to the regression model. Variations from the model were clustered using Spectral Clustering, which yielded two sets of clusters, containing 10 flights and 15 flights, respectively. The two clusters represent unusual and normal aero engine performance in the test flights.
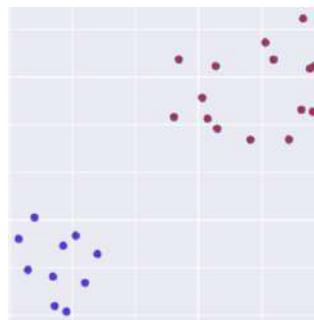
Figure 5. Spectral Clustering for 25 Test Flights

The 10 flights identified as displaying unusual aero engine performance were compared with the traditional threshold-based exceedance detection system. The maximum values reached by Q and N1 during the entire flight duration were not sufficient to trigger the threshold.

Table 2. Comparison of identified unusual performance test flights (TF) with N1 and Q thresholds

	TF1	TF2	TF3	TF4	TF5	TF6	TF7	TF8	TF9	TF 10	Threshold
Q	40.8	44.3	46.9	51.4	53.7	55.5	60.8	67.3	72.2	77.4	80
N1	46.6	50.4	54.6	58.1	61.3	63.4	69.8	76.2	82.3	88.5	92

5. Conclusion

The data processing approach proposed in this paper uses least squares regression to model the performance of a turboshaft engine installed on a rotorcraft from a set of 100 flights. The model is validated by employing spectral clustering to detect unusual engine performance in 25 test flights. Spectral clustering was able to treat every node/ data point in a graph independently, which enabled in effectively separating out dis similar test flight data sets. Using the synthetically generated test flight data, the discussed ML algorithm was able to identify 10 flights showing unusual or derated aeroengine performance. The maximum values attained by Q and N1 in all the identified unusual test flights were less than the threshold values and would have gone unnoticed by traditional threshold-based analysis. This data processing approach, however, does not aim to replace the existing threshold-based analysis. But it provides a different perspective to FDM, by utilizing ML for establishing true relationships between the different FDR features.

References

[1]   Jasra, Sameer & Gauci, Jason & Muscat, Alan & Valentino, Gianluca & Zammit-Mangion, David & Camilleri, Robert. Literature review of machine learning techniques to analyse flight data. AEGATS Conference. 2108 Oct. Toulouse, France.
[2]   Chunxiao Zhang, Nan Wang. Aero-Engine Condition Monitoring Based on Support Vector Machine. Physics Procedia. 2012; Volume 24, Part B, Pages 1546-1552. https://doi.org/10.1016/j.phpro.2012.02.228.

[3]   Li-Hua Ren, Zhi-Feng Ye, Yong-Ping Zhao. A modeling method for aero-engine by combining stochastic gradient descent with support vector regression. Aerospace Science and Technology. 2020; Volume 99. https://doi.org/10.1016/j.ast.2020.105775.

[4]   Wang, Haifeng & Hu, Dejin. Comparison of SVM and LS-SVM for regression. International Conference   on   Neural   Networks   and   Brain   Proceedings.   2005; doi.org/10.1109/ICNNB.2005.1614615.

[5]   Lishuai Li, Santanu Das, R. John Hansman, Rafael Palacios and Ashok N. Srivastava. Analysis of Flight Data Using Clustering Techniques for Detecting Abnormal Operations. Journal of Aerospace Information Systems. 2015; p. 587-598. doi:10.2514/1.I010329.

[6]   Julian Oehling, David J. Barry, Using machine learning methods in airline flight data monitoring to generate new operational safety knowledge from existing data, Safety Science. 2019; Volume 114, Pages 89-104. https://doi.org/10.1016/j.ssci.2018.12.018.

[7]   Clinton V. Oster, John S Strong, Zorn Kurt. Why Airplanes Crash: Causes of Accidents Worldwide. 51st Annual Transportation Research Forum. 2010. Doi: 10.22004/ag.econ.207282.

[8]   Shawe-Taylor J, Cristianini N. Kernel Methods: An Overview. Kernel Methods for Pattern Analysis. Cambridge: Cambridge University Press; 2004. doi:10.1017/CBO9780511809682.

[9]   D. Huang, C. Wang, J. Wu, J. Lai and C. Kwoh. Ultra-Scalable Spectral Clustering and Ensemble Clustering. IEEE Transactions on Knowledge and Data Engineering. 2020 Jun 01; vol. 32, no. 6: pp. 1212-1226. doi: 10.1109/TKDE.2019.2903410.

Distributed Denial of Service Attacks Detection and Mitigation in Software Defined Mesh Networks

Santosh Mani [a, 1], Manisha J Nene [a]

[a] *Department of Computer Science and Engineering, Defence Institute of Advanced Technology (DU)*
Girinagar, Pune - 411025, India

Abstract. Networks configured in Mesh topology provide Network security in the form of redundancy of communication links. But redundancy also contributes to complexity in configuration and subsequent troubleshooting. Critical networks like Backbone Networks (used in Cloud Computing) deploy the Mesh topology which provides additional security in terms of redundancy to ensure availability of services. Distributed Denial of Service attacks are one of the most prominent attacks that cause an immense amount of loss of data as well as monetary losses to service providers. This paper proposes a method by which using SDN capabilities and sFlow-RT application, Distributed Denial of Service (DDoS) attacks is detected and consequently mitigated by using REST API to implement Policy Based Flow Management through the SDN Controller which will help in ensuring uninterrupted services in scenarios of such attacks and also further simply and enhance the management of Mesh architecture-based networks.

Keywords. DDoS attacks, SDN, Attack detection and mitigation, Flow Tables, sFlow-RT, sFlow Agent, sFlow Collector, Policy Based Flow Management (PBFM).

1. Introduction

DDoS attacks on Next Generation Networks (NGNs) leads to serious consequences for critical users like Defence networks and Cloud computing networks with the Core/Backbone infrastructure in a Mesh topology for redundancy against failures to ensure uninterrupted services. Intelligent systems like Internet of Things (IoTs) use Cloud computing environments for data transmission and reception but security measures lacunae and IoT devices easy accessibility cause the Cloud network to be compromised and exploited [33]. Examples of highly impactful DDoS attacks include Estonia attack[6], Sweden Transportation network attack [6], US Department of Health and Human Services (HHS) [7] website attack, *Dyn attack* on high-profile websites such as GitHub, Twitter, Reddit, Netflix, Airbnb and many others [33].With the Covid-19 pandemic crisis, Digital Platform utilisation has skyrocketed to a new high, opening up new avenues for malicious attacks like DDoS. As of 2020, recovery from DDoS attacks takes around four hours [4]. DDoS attacks continue to grow in various dimensions like size (bandwidth), frequency of attack, complexity of attack which threaten and businesses and service providers all over the world.

---

[1] Santosh Mani, Department of Computer Science and Engineering, Defence Institute of Advanced Technology, Email: santorajim@gmail.com.

In this paper a method by utilizing the SDN capability of Policy Based Flow Management (PBFM) with sFlow-RT application has been proposed for detection and mitigation of DDoS attacks for Full Mesh Networks. DDoS attacks, its classification and the related literature review is given in Section 2. A description of SDN Core concept, flow table concept and sFlow-RT tool utilised is given in Section 3. Section 4 gives the description of the proposed methodology to detect and mitigate DDoS attacks. Section 5 presents the experimentation carried out and the observations. Section 6 concludes the paper and explains how SDN helps in better detection and mitigation of DDoS attack while opening up further scope for better troubleshooting of affected links. In this paper the proposed methodology using SDN capabilities has been implemented for Layer 3 and Layer 4 of the OSI layer.

2. Related work – DDoS attacks and classification, literature review

A brief description of types of DDoS attacks [8] and its classification is discussed here along with related literature review for enhancing the network security.

The DDoS attacks are classified as follows [9] [10] [11]: -
Volume based attacks – Overwhelming a resource (server's website) with a huge amount of traffic to prevent access, e.g., UDP flood/reflection attack, ICMP flood.
Protocol attack – Involves exploiting a protocol's weakness, e.g., TCP SYN attack.
Application layer attacks -Used to take down web servers. Also called Layer 7 (OSI Layer 7) or Application layer attack. In this paper, Volume based attack and Protocol attack have been utilised to demonstrate DDoS attacks detection and mitigation in SDN based Mesh Networks.

A review of literature addressing the challenges of DDoS attacks by leveraging SDN capabilities is as follows:
A proposed solution for detecting and mitigating DDoS attack for a tree architecture [11] was the main motivation for developing a DDoS Attack detection and mitigation mechanism in SDN for a Mesh Network which would prove extremely beneficial for Cloud Computing and Cloud based Services. An efficient anomaly detection and mitigation by firewall implementation in all network devices is proposed in [12]. Detection and mitigation of DDoS attacks in Legacy networks is given in [13] and for SDN environments is proposed in [16]. A comprehensive study in existing and new proposed techniques for DDoS attack detection and mitigation are discussed in [14],[15].

3. Preliminaries

3.1. Sdn Core concept and Flow table concept

SDN separates the Control Plane (network environment discovery) from Data Plane (Data/traffic flow) and places the network intelligence at a centralised location called Controller [1][2][18][19][20][21][34]. SDN utilises Flows present in Flow Tables of forwarding network devices in SDN to route packets. A Flow entry/Flow in a Flow Table is a set of packet field values (Figure 1.) which have a match (filter) criterion and a set of instructions / actions which help in formulating policies for services implemented in the Application Layer.

Match Fields	Priority	Counters	Instructions	Timeouts	Cookie	Flags

Figure. 1. Flow entry

3.2. sFlow-RT

sFlow is a sampling technology and can be embedded in switches and routers and continuously monitors traffic flows on all interfaces of networks devices simultaneously [3][21][22][23][24][25][26]. It has two components sFlow Collector and Analyser (Central data collector which analyses the sampled traffic sent by the sFlow Agent) and sFlow Agent (embedded in a switch/router or function as a standalone probe).The sFlow-RT analytics engine receives a continuous telemetry stream from sFlow Agents in network devices, hosts and applications and converts the raw measurements into actionable metrics, accessible through the RESTflow API to configure customized measurements, retrieve metrics, set thresholds, and receive notifications [24]. sFlow-RT detects anomalies in traffic and uses RESTflow API to inform the SDN controller to undertake the mitigation actions. [24]
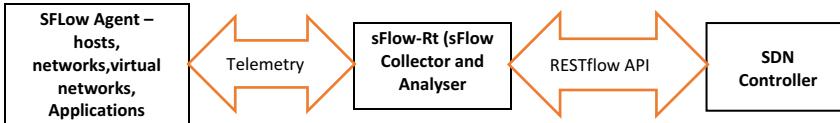
Figure. 2. Logical connectivity between sFlow-RT and SDN Controller

4. Methodology

The methodology proposes the utilization of SDN's PBFM to provide a solution for DDoS attacks detection and mitigation for Full Mesh Networks by combining sFlow-RT (sFlow Agent, sFlow Collector and Analyser) which monitors and using RESTflow API transmits the information to the Controller which then manipulates Flow Table entries in the concerned network devices in a particular flow/path.

4.1. DDoS attack scenario: Detection and mitigation methodology

The logic (Figure 3) of the DDoS Attacks detection and mitigation has been built over the underlaying base Mesh network using routing protocol Shortest Path First (SPF) with Link-failure and Link-flapping detection and mitigation incorporated [34].

First the Controller and mesh network topology are activated with Link-failure and Link-flapping detection and mitigation logic in the background. Next, the threshold is set for DDoS attacks detection on number of traffic packets being sent from a host. Then, the key parameters to classify the DDoS attack type are set. In this paper, three kinds of threat scenarios – UDP reflection, ICMP flood and TCP SYN attack are studied.

After the DDoS attack is initiated, the threshold limit is checked. If crossed, using the defined key parameters, the attack type is identified from the packet samples taken using sFlow-RT. Then based on malicious attacker connected, the router/switch is identified. Flow entries are made in the identified router/switch for that particular
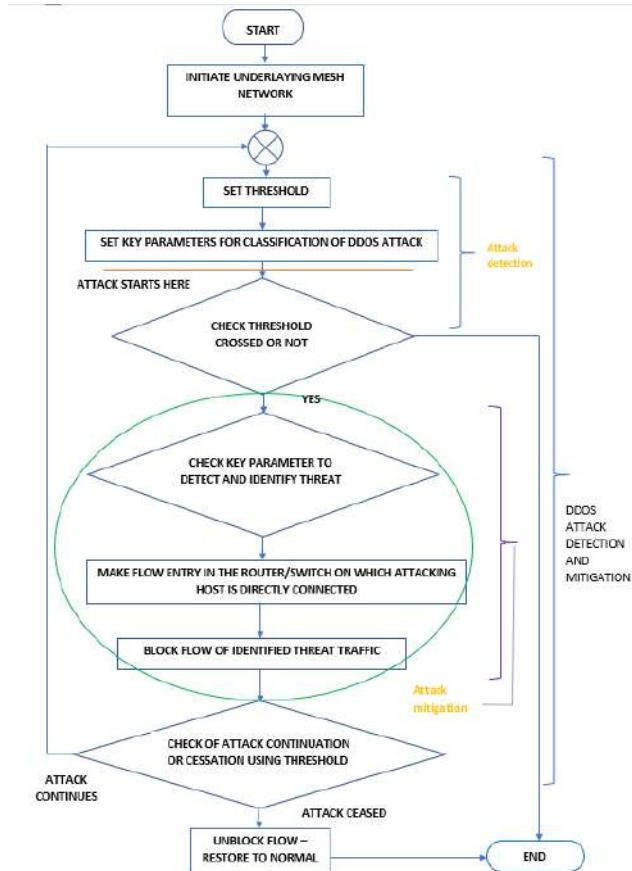
Figure. 3. Methodology for detection and mitigation of DDoS Attacks

attack type to block the flow of threat traffic. The logic applied further to distil the traffic based on each type of attack is given below in this section as test cases.

The logic for intermittent check to see whether the attack traffic is in progress or has ceased is additionally implemented so that once the attack has stopped, the flow is restored to normal, else if not, the detection and mitigation cycle will repeat.

4.2. Test cases for different types of DDoS attacks: -

The logic applied is as follows for the three cases (Figure 4): -

Case 1: UDP reflection attack detection and mitigation logic. First, threshold limit is checked for all the cases. If threshold is crossed and UDP reflection attack traffic is detected and identified based on keys 'ip destination' and 'udp source port' a flow entry with the filters as the key parameters to stop the flow of attack traffic is made to stop the UDP reflection DDoS attack. If the attack is not identified as UDP reflection attack traffic then the logic shifts to the next attack ICMP flood attack.
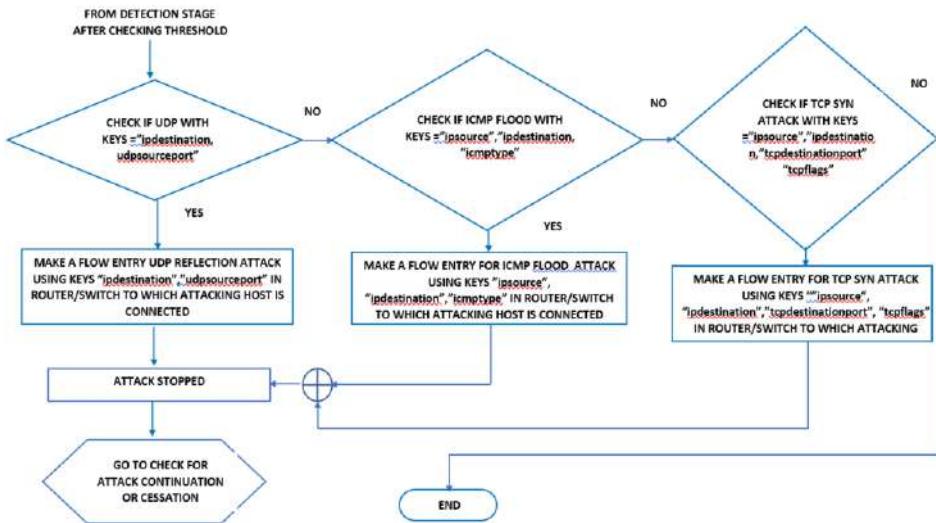
Figure. 4. Test cases for DDoS attacks

Case 2: ICMP flood attack detection and mitigation logic. If threshold is crossed and ICMP flood traffic is detected and identified based on the keys 'ip source', 'ip destination' and 'icmp type' a flow entry with the filters as the key parameters is made to stop the ICMP flood DDoS attack. If the attack is not identified as ICMP flood attack traffic then the logic shifts to the next attack TCP SYN attack.

Case 3: TCP SYN attack detection and mitigation logic. If the threshold is crossed and TCP SYN attack is detected and identified based on the keys 'ip source', 'ip destination','tcp destination port' and 'tcp flags'   a flow entry with the filters as the key parameters is made to stop the TCP SYN DDoS attack. If the attack is not identified as TCP SYN attack traffic, then the logic shifts back to the next step in detection and mitigation methodology (Figure 3)

5. Experiments and Observations

This section describes the experimentation carried out to realise the proposed concepts in the previous section. The mesh topology (Figure 5) has been made in Mininet [27][31][32] and RYU controller utilised for controlling and managing the network. Underlaying routing protocol is SPF with Link-failure and Link-flapping detection and mitigation capability [34]. Attacks are carried out using packet crafter tool *hping3*. sFlow-RT detects anomalies in traffic for DDoS attack detection and uses RESTflowf API to inform the SDN controller to undertake the mitigation actions [28][29][30]. The same topology when viewed from sFlow-RT dashboard is shown in Figure 6.
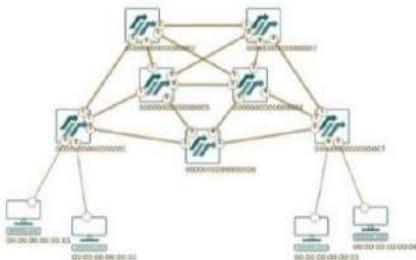
Figure. 5. Topology for experiment
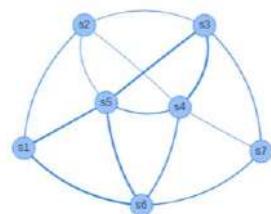
Figure. 6. Topology as seen in sFlow-RT

Following is the illustration of the sequence of events executed during the experimentation: -
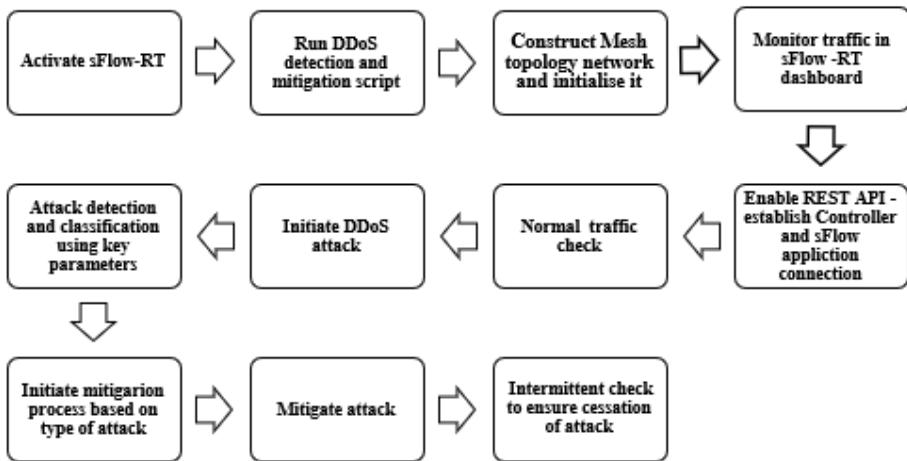
Figure. 7. Experimentation steps

First, the sFlow-RT application for monitoring and analysing the sampled data is activated. The next important step is to initialise and activate the DDoS attack detection and mitigation script. This script is responsible for Setting the threshold which is monitored by the sFlow-RT application, Setting the parameters for classifying the type of DDoS attack, Utilising the sampled data packet taken by the sFlow-RT to identify and detect the attack and Using RESTflow API to carry out DDoS attack mitigation by Policy Based Flow Management.

Next the Mesh topology is initialised and activated with SPF routing protocol and the sFlow-RT traffic monitoring dashboard is started along with enabling REST API connection between Controller and sFlow-RT application. A normal traffic check is carried out to ascertain the working of the sFlow-RT application.

Next, DDoS attack is initiated. Three different types of attacks i.e., UDP reflection attack, ICMP flood attack and TCP SYN attack are utilised for experimentation. The sFlow-RT keeps monitoring the threshold. Once the threshold is crossed, the key parameters in the packets are checked to determine the attack traffic type whether it is UDP reflection attack packets, ICMP flood attack packets or TCP SYN attack packets. After detection, then the mitigation phase is activated.

In the mitigation phase, using PBFM, a flow entry having very high priority is made in the router/switch (Figure 10) to which the attacking host is connected to stop the flow of the attacking traffic only. Following is an example for TCP SYN attack.

Figure. 8. TCP SYN attack initiated

Figure. 9. TCP SYN attack detected and blocked.

Figure. 10.   Flow entry made to block the flow
        of TCP SYN attack traffic

Figure. 11.   Unblock traffic flow

Additional functionality (Figure 11.) of periodically checking whether the attack is in progress or not is implemented. It helps to bring the status of the network devices and flows to normal by unblocking the traffic, if within threshold limits, else the traffic will be kept in blocked state only.

Following were the observations: -

Using PBFM only the malicious traffic is blocked. Normal traffic detected within the threshold is allowed without any interruptions. The methodology only stops the flow of malicious traffic and does not completely isolate the attacking host by shutting down the switch ports. It instead opens up an opportunity for the network administrators and network security experts who access the infected device and try to analyse the attack reasons while normal traffic flows through the network.

If both malicious traffic and normal legitimate traffic within threshold limits are traversing through the same path and network devices only the malicious traffic is selectively stopped and the normal legitimate traffic is not disturbed. It again provides an opportunity for troubleshooting without disturbing the network.

The mechanism of intermittent check of the event whether the attack is in progress helps to gauge the duration of the attack once the attack is ceased. After checking the traffic threshold limits, on ceasing of attack, the traffic is unblocked.

Figure 12. illustrates the graph of how the attacks started in a burst and how it is brought down using Policy Based Flow Management.
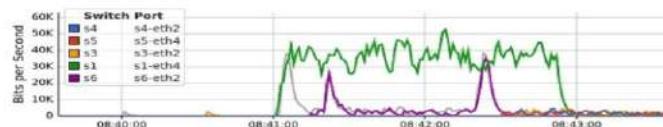
Figure. 12. Graphical representation of TCP SYN attack initiation and mitigation

The highlighted graph (green colour) shows that the attack was started by a host in router/switch s1 and sFlow recognised it as the traffic originating network device. The traffic initiated increased quickly to an average of around 45000 bps. It is detected and mitigated using PBFM and sFlow-RT, by making a flow entry using REST API, through the Controller in the identified network device and the graph dips down to zero.

6. Conclusion

In this paper the study and experimentation using SDN capabilities of PBFM with sFlow-RT application was carried out to detect and mitigate the DDoS attacks and the results found are summarised as follows: -

The network device to which the attacking host is connected is identified and pinpointed accurately and helps in detecting and mitigating the attack at the source of attack itself. Detection is carried out using a script with predefined threshold and the parameters to classify DDoS attack types defined. Mitigation of DDoS attack is realised by making a flow entry in the identified switch/router using PBFM capability of SDN.

Using predefined threshold and classification parameters, PBFM capability of SDN and sFlow-RT with RESTful API, the malicious DDoS Attack traffic is identified and isolated from the normal traffic which passes uninterrupted. With traffic isolation and affected device pinpointed, methodology adopted opens up troubleshooting opportunity without disturbing the network and network traffic for infected devices and all network devices from attacking source host to target destination host.

Intermittent checking functionality for the DDoS attack continuity or cessation helps in gauging attack duration by acting as a gatekeeper to monitor the attack. When the attack ceases, the traffic for that particular kind of flow is unblocked if within threshold limits. With the proposed methodology in the paper, the Controller enables the router/switch, to which the attacking host is connected, to function as a dynamic firewall which gets activated once a trigger in the form of threshold crossing is detected. Better control over the network will help in facing challenges like mapping of network devices and forensic artefacts for attacker tracking due to Cloud Service Utilisation by IoT devices [33].

References

[1]   RFC 7426 – Software Defined Networking: Layers and Architecture Technology
[2]   RFC 7149 – Software Defined Networking: A perspective from within a Service Provider Environment
[3]   RFC 3176 - InMon Corporation's sFlow: A Method for Monitoring Traffic in Switched and Routed Networks
[4]   https://www.comparitech.com/blog/information-security/ddos statistics-facts/
[5]   Rajeev Singh, Department of Computer Engineering College of Technology,G. B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India ,rajeevpec@gmail.com T. P. Sharma Department of Computer Science & Engineering ,National Institute of Technology, Hamirpur, Himachal Pradesh, India ,teekparval@gmail.com. Present Status of Distributed Denial of service (DDoS) Attacks in Internet World. International Journal of Mathematical, Engineering and Management Sciences Vol. 4, No. 4, 1008–1017, 2019, https://dx.doi.org/10.33889/IJMEMS.2019.4.4-080
[6]   DDoS during the COVID-19 pandemic: attacks on educational and municipal websites tripled in Q1 2020 | Kaspersky
[7]   https://www.cloudflare.com/learning/ddos/what-is-a-ddos-attack/
[8]   https://www.indiumsoftware.com/blog/how-ddos-attack-protect-your- business
[9]   Q. Yan, F. R. Yu, Q. Gong and J. Li, "Software-Defined Networking (SDN) and Distributed Denial of Service (DDoS) Attacks in Cloud Computing Environments: A Survey, Some Research Issues, and Challenges," in *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 602-622, Firstquarter 2016, doi: 10.1109/COMST.2015.2487361.
[10]   Christos Douligeris, Aikaterini Mitrokotsa, DDoS attacks and defense mechanisms: classification and state-of-the-art, Computer Networks, Volume 44, Issue 5, 2004, Pages 643-666, ISSN 1389-1286, https://doi.org/10.1016/j.comnet.2003.10.003.

[11] B. H. Lawal and A. T. Nuray, "Real-time detection and mitigation of distributed denial of service (DDoS) attacks in software defined networking (SDN)," *2018 26th Signal Processing and Communications Applications Conference (SIU),* Izmir, Turkey, 2018, pp. 1-4, doi: 10.1109/SIU.2018.8404674.

[12] A. Kumar and N. K. Srinath, "Implementing a firewall functionality for mesh networks using SDN controller," *2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, Bengaluru, India, 2016, pp. 168-173, doi: 10.1109/CSITSS.2016.7779417.

[13] K. Giotis, G. Androulidakis and V. Maglaris, "Leveraging SDN for Efficient Anomaly Detection and Mitigation on Legacy Networks," *2014 Third European Workshop on Software Defined Networks*, Budapest, Hungary, 2014, pp. 85-90, doi: 10.1109/EWSDN.2014.24.

[14] Bawany, N.Z., Shamsi, J.A. & Salah, K. DDoS Attack Detection and Mitigation Using SDN: Methods, Practices, and Solutions. Arab J Sci Eng 42, 425–441(2017). https://doi.org/10.1007/s13369-017-2414-5

[15] N. I. G. Dharma, M. F. Muthohar, J. D. A. Prayuda, K. Priagung and D. Choi, "Time-based DDoS detection and mitigation for SDN controller," *2015 17th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, Busan, Korea (South), 2015, pp. 550-553, doi: 10.1109/APNOMS.2015.7275389.

[16] K. Giotis, C. Argyropoulos, G. Androulidakis, D. Kalogeras, V. Maglaris,Combining OpenFlow and sFlow for an effective and scalable anomaly detection and mitigation mechanism on SDN environments, Computer Networks, Volume 62, 2014, Pages 122-136, ISSN 1389-1286, https://doi.org/10.1016/j.bjp.2013.10.014.

[17] Software-defined network for dummies – Mykola konrad, Dan teichman with Brian Underdahl . Sonus special edition.(A Wiley brand).

[18] D. Kreutz, F. M. V. Ramos, P. E. Veríssimo, C. E. Rothenberg, S. Azodolmolky and S. Uhlig, "Software-Defined Networking: A Comprehensive Survey," in *Proceedings of the IEEE*, vol. 103, no. 1, pp. 14-76, Jan. 2015, doi: 10.1109/JPROC.2014.2371999.

[19] R. G. Rao and M. J. Nene, "SEDoS-7: A proactive mitigation approach against EDoS attacks in cloud computing," *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, India, 2017, pp. 965-970, doi: 10.1109/WiSPNET.2017.8299905.

[20] Brian Underdahl and Gary Kinghorn. (2015) Software Defined Networking For Dummies®, Cisco Special Edition Published by John Wiley & Sons, Inc. 111 River St.Hoboken, NJ 07030-5774 www.wiley.com

[21] https://inmon.com/products/sFlow-RT.php (Last accessed 11-06-2021)

[22] Phaal et al.(2014) Method,System and Computer Program Product for identifying common factors associated with network activity with reduced resource utilization. US Patent 8,838,774, 16 Sept 2014

[23] Phaal P. (2017) Method for asynchronous calculation of network traffic rates based on randomly sampled packets. US Patent 9,509,583, 29 Nov 2016

[24] Peter P. (2017) Distributed traffic quota measurement and enforcement. US Patent 9,712,443, 18 Jul 2017.

[25] Peter P. (2017) Method and system of large flow control in communication networks. US Patent 9,722,926, 1 Aug 2017.

[26] https://sflow-rt.com/ (Last accessed on 11-06-2021)

[27] https://github.com/mininet/mininet/wiki/Documentation (Last accessed 11-06-2021)

[28] https://sflow-rt.com/ (Last accessed on 11-06-2021)

[29] Lantz, Bob; Heller, Brandon; McKeown, Nick (2010). *[ACM Press the Ninth ACM SIGCOMM Workshop - Monterey, California (2010.10.20-2010.10.21)] Proceedings of the Ninth ACM SIGCOMM Workshop on Hot Topics in Networks - Hotnets '10 - A network in a laptop. , (), 1–6.* doi:10.1145/1868447.1868466

[30] sFlow-RT Writing Applications (sflow-rt.com)

[31] sFlow-RT Defining Flows (sflow-rt.com)

[32] sFlow-RT Reference (sflow-rt.com)

[33] Girija Devi M.S., Nene M.J. (2019) Security Breach and Forensics in Intelligent Systems. In: Satapathy S., Joshi A. (eds) Information and Communication Technology for Intelligent Systems. Smart Innovation,Systems and Technologies, vol 107. Springer, Singapore. https://doi.org/10.1007/978-981-13-1747-7_33

[34] S. Mani and M. J. Nene, "Data Loss Prevention due to Link-flapping using Software Defined Networking," *2021 6th International Conference for Convergence in Technology (I2CT)*, 2021, pp. 1-5, doi: 10.1109/I2CT51068.2021.9417990.

Quantum Information Transmission Using CNOT Gate

Ankit Sharma [a,1] and *Manisha J Nene* [a]
[a] *Department of Computer Science & Engineering Defense Institute of Advanced Technology, Pune, Maharashtra, India*

Abstract. We are at the dawn of quantum era; research efforts are been made on quantum information transmission techniques. Properties of quantum mechanics poses unique challenges in terms of wave collapse function, No cloning theorem and reversible operations. Quantum teleportation and quantum entanglement swapping based architecture are utilized to transmit qubit. In this paper we propose an approach to transmit qubits using controlled NOT gate (CNOT) gates and implement it on quantum machine.

Keywords. Qubit, Quantum information transmission, Quantum teleportation, Quantum networks, Entanglement swapping, Quantum gates, CNOT gate, Wave collapse function, No cloning theorem, Reversible operations.

1. Introduction

One of the early quantum applications in field of quantum cryptology was demonstrated in 1992 [1], since then there has been a constant interest and research done by the scientific community in the field of quantum. Recent development of concepts and application has been seen in the areas of quantum authentication [2-4], quantum cryptology [5], quantum random number [6], distributed quantum computation [7], and quantum networks [8-10].

Quantum machines are located over different geographical locations. These machines are connected with each other over classical channels. One of the major disadvantage of such connections over classical channel is that quantum machine performs operations over qubits which has inherent parallelism and other quantum features like superposition. The qubit cannot be transferred over a classical channel and when the value of qubit is measured it collapse to a single finite value which results in loss of its quantum properties [11-15]. The inability of existing communication channels to transmit qubit has put a limit on realization of quantum computation over a network to its full potential.

One important component for any quantum network is qubit transferring/exchanging technique. In quantum mechanics operations on qubits are influenced by wave function collapse [13], No cloning theorem [14] and reversible Operations [15]. These properties pose challenges to measure or copy quantum state in between transmission and also puts restriction to use only reversible operations.

---

[1] *Ankit Sharma*, *Department of Computer Science & Engineering, Defense Institute of Advanced Technology, Pune, Maharashtra, India*; E-mail: vividankit@yahoo.com.

The outline of paper is as follows we first take an insight into various methods of quantum information exchange, the fundamentals of quantum mechanics. Then we cover our proposed approach, mathematical modelling, experiments and then discuss its results.

2. Related work

Quantum communication [16] and Quantum internet [17] has been proposed to connect various quantum devices spread over different geographical location with exchange of qubits or quantum state as the source of information.

To transmit qubit over an extended range there is a requirement of quantum repeater also there is a need of quantum router [18] for routing of information from its source to destination. Various Schemes for implementation of quantum information transmission involving quantum repeater and quantum router have been proposed. For routing function qubits are classified into control and signal qubits in which control qubits controls the path of signal qubits [18]. In some schemes function of control qubit are implemented with classical bits and classical channels which results in semi quantum devices [19]. Techniques have been demonstrated in domain of light-matter interaction [20]. Entanglement based router schemes have been proposed involving changes in signalling information as the routing process progresses [21] or not involving any changes in signalling information as the routing process progresses [22]. Routing schemes have been proposed which does not involve entanglement between control and signal qubits [23]. Other schemes for quantum information exchange have also been proposed which differ in their architectural implementation [24].

3. Quantum preliminaries

Before we begin to explore quantum space for quantum information exchange, we need to be familiar with few preliminaries [25].

3.1. Qubit representation

- In Qubit representation state 0 or |0> and state 1 or |1> is represented by a 2-by-1 matrix as follow:

$$|0> = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad |1> = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

- For Multi Qubits value of |00>, |01>, |10> and |11> can be found by taking tensor product (represented by $\otimes$). For example, value of |00> can be found as:

$$|00> = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

- Logical gates for quantum operations are also modeled with matrices.

3.2. Existing Techniques for quantum information transmission

- Quantum Teleportation. It is one of the methods which can be used to transport quantum state from one place to another [26].
- Quantum entanglement swapping utilizes bell state measurement to extend the entanglement & hence the quantum information [27-28].

4. Proposed approach

In this section we will delve upon our proposed approach of using CNOT gates or controlled NOT gate for quantum information transmission. We will study CNOT gate, truth table, mathematical modelling and the proposed approach.

4.1. CNOT gates

CNOT gate or controlled NOT gate [25] is drawn as shown in Figure 1.
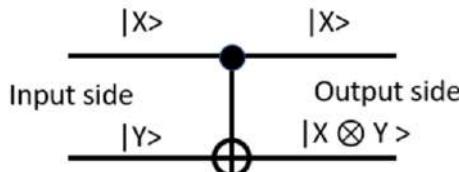
Figure 1. CNOT gate.

- This gate has two inputs and gives two outputs, the top input ($|X>$) can be called as control qubit and bottom input ($|Y>$) can be called as target qubit. If the value of control $|X> = 1$ then the value of target flips and if the value of control is $|X> = 0$ then value of target does not change and remains same as the input. The output value of target can be depicted by the expression $|X>$XOR$|Y>$ where XOR is the exclusive OR. It can be summarized as shown in table 1:

Table 1. Truth table CNOT gate

S.no	Input		Output	
	Control	Target	Control	Target
1	$|0>$	$|0>$	$|0>$	$|0>$
2	$|0>$	$|1>$	$|0>$	$|1>$
3	$|1>$	$|0>$	$|1>$	$|1>$
4	$|1>$	$|1>$	$|1>$	$|0>$

4.2. Mathematical modelling

- CNOT gate can be mathematically modelled as:

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

e.g. if we provide it with input |11> (|X>=1 control and |Y>=1 target) we get output as |10> (|X>=1 remains same and |Y>=0 flips).

- Passing |11> to CNOT gate can be shown as matrix multiplication as shown:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \text{ which gives result equal to } \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

Result obtained can be simplified as $\begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ which is equal to |10>

- Similarly, other values of Table 1 can be verified by this mathematical modelling.

4.3. Proposed methodology

CNOT gate uses two qubits as input and gives two qubits as output if we use control qubit to pass signaling information and target qubit to pass quantum information then we can use CNOT gate for quantum information transmission.

Such an implementation would be helpful in quantum information transmission system which is working in quantum domain connected to quantum devices.

- It can be used for sharing networking information between quantum devices.
- It can be utilised for establishing quantum information transmission system.
- Control and target qubit both can be passed in the quantum network or either one of them can be passed forward depending upon the requirement and architecture of the quantum network employed.
- Such a passing of quantum information in quantum network can help build quantum networks independent of classical channels.
- Operations based on CNOT gate are faster and low cost in terms of quantum memory and processing when compared with other quantum operations like quantum teleportation.

5. Experiment and results

We have used IBM quantum experience an open-source platform to conduct our experiments. Figure 2 we designed a quantum circuit in which the different parameter values generated by Hadamard gate and entanglement circuit in qubits q0 and q1 are transferred/passed through different qubits q2, q3, q4, q5 signifying different values/paths.
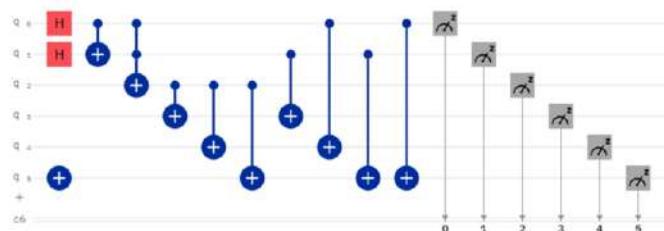
Figure 2. CNOT gate-based quantum circuit.
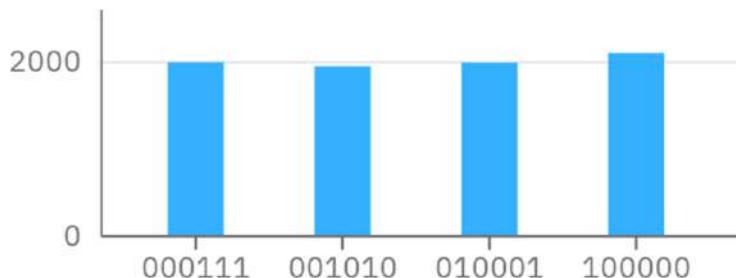
Figure 3. Result obtained after execution of circuit.

Figure 3 shows the results obtained after running the circuit on IBM quantum experience simulator. It can be observed in results that depending on values of q0 & q1 other qubits (q2, q3, q4 & q5) representing different values/paths are getting activated. These can be summarized as given in table 2.

Table 2. Analysis of result obtained from execution of circuit.

S.no	Result value	Value of q0 (Control/ Signaling information)	Value of q1 (Target/ quantum information)	Path activated
1	000111	1	1	q2
2	001010	0	1	q3
3	010001	1	0	q4
4	100000	0	0	q5

In this circuit we have chosen to activate a path on the value of control and target qubit. Similarly, we can draw circuit in which we can transfer the values of either both control and target or any one of them.

6. Conclusion

For transferring of qubits in quantum network, circuit based on CNOT gate can be utilized. Such a circuit would consume less resources, and would use simple quantum operations and hence would be easy to implement. Information from passing of qubits can further be enhanced with predefined quantum functions.

7. Acknowledgement

References

[1] Bennett, C., Bessette, F., Brassard, G., Salvail, L., & Smolin, J. (1992). Experimental quantum cryptography. Journal of Cryptology, 5(1). doi:10.1007/bf00191318.

[2] Sharma, M. K., & Nene, M. J. (2019, October). Quantum One Time Password with Biometrics. In International Conference on Innovative Data Communication Technologies and Application (pp. 312-318). Springer, Cham

[3] Sharma, MK, Nene, MJ (2020). Dual factor third-party biometric-based authentication scheme using quantum one time passwords[†]. Security and Privacy. 2020; 3:e129. https://doi.org/10.1002/spy2.129.

[4] Upadhyay, G., & Nene, M. J. (2016). One time pad generation using quantum superposition states. IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology.

[5] Ekert, A. K. (1991). Quantum cryptography based on Bell's theorem. Physical Review Letters, 67(6), 661–663. doi:10.1103/physrevlett.67.661

[6] M. Gupta and M. J. Nene, "Random Sequence Generation using Superconducting Qubits," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 640-645, doi: 10.1109/ICICV50876.2021.9388495.

[7] Buhrman, H., & Röhrig, H. (2003). Distributed Quantum Computing. Lecture Notes in Computer Science, 1–20. doi:10.1007/978-3-540-45138-9_1.

[8] Chiribella, G., D'Ariano, G. M., & Perinotti, P. (2009). Theoretical framework for quantum networks. Physical Review A, 80(2). doi:10.1103/physreva.80.022339.

[9] Vedral, V., Barenco, A., & Ekert, A. (1996). Quantum networks for elementary arithmetic operations. Physical Review A, 54(1), 147–153. doi:10.1103/physreva.54.147.

[10] Wallquist, M., Hammerer, K., Rabl, P., Lukin, M., & Zoller, P. (2009). Hybrid quantum devices and quantum engineering. Physica Scripta, T137, 014001. doi:10.1088/0031-8949/2009/t137/014001.

[11] Makhlin, Y. (2002). Quantum Information Processing, 1(4), 243–252. doi:10.1023/a:1022144002391.

[12] Horodecki, R., Horodecki, P., Horodecki, M., & Horodecki, K. (2009). Quantum entanglement. Reviews of Modern Physics, 81(2), 865–942. doi:10.1103/revmodphys.81.865.

[13] Bassi, A., Lochan, K., Satin, S., Singh, T. P., & Ulbricht, H. (2013). Models of wave-function collapse, underlying theories, and experimental tests. Reviews of Modern Physics, 85(2), 471–527.

[14] Kuzyk, M. G. (2019). Quantum no-cloning theorem and entanglement. American Journal of Physics, 87(5), 325– 327.

[15] Brandão, F. G. S. L., & Gour, G. (2015). Reversible Framework for Quantum Resource Theories. Physical Review Letters, 115(7). doi:10.1103/physrevlett.115.070503.

[16] P. Nema and M. J. Nene, "Pauli Matrix based Quantum Communication Protocol," 2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI), 2020, pp. 1-6, doi: 10.1109/ICATMRI51801.2020.9398393.

[17] Kimble, H. J. (2008). The quantum internet. Nature, 453(7198), 1023–1030. doi:10.1038/nature07127.

[18] Behera, B. K., Seth, S., Das, A., & Panigrahi, P. K. (2019). Demonstration of entanglement purification and swapping protocol to design quantum repeater in IBM quantum computer. Quantum Information Processing, 18(4). doi:10.1007/s11128-019-2229-2

[19] Hall, M. A., Altepeter, J. B., & Kumar, P. (2011). Ultrafast Switching of Photonic Entanglement. Physical Review Letters, 106(5). doi:10.1103/physrevlett.106.053901

[20] Aoki, T., Parkins, A. S., Alton, D. J., Regal, C. A., Dayan, B., Ostby, E., … Kimble, H. J. (2009). Efficient Routing of Single Photons by One Atom and a Microtoroidal Cavity. Physical Review Letters, 102(8). doi:10.1103/physrevlett.102.083601.

[21] Chang, X.-Y., Wang, Y.-X., Zu, C., Liu, K., Duan, L.-M. (2012): Experimental demonstration of an entanglement-based quantum router. arXiv preprint arXiv:1207.7265v1.

[22] Yan, G.-A., Cai, Q.-Y., & Chen, A.-X. (2016). Information-holding quantum router of single photons using natural atom. The European Physical Journal D, 70(4). doi:10.1140/epjd/e2016-60724-6

[23] Lemr, K., Bartkiewicz, K., Černoch, A., Soubusta, J., & Miranowicz, A. (2012). Experimental linear-optical implementation of a multifunctional optimal qubit cloner. Physical Review A, 85(5).

[24] Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.

[25] N. S. Yanofsky and M. A. Mannucci, Quantum Computing for Computer Scientists. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[26] Bouwmeester, D., Pan, J.-W., Mattle, K., Eibl, M., Weinfurter, H., & Zeilinger, A. (1997). Experimental quantum teleportation. Nature, 390(6660), 575–579. doi:10.1038/37539.

[27] Gupta, M., & Nene, M. J. (2020, December). Quantum Computing: An Entanglement Measurement. In 2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI) (pp. 1-6). IEEE.

[28] Su, X., Tian, C., Deng, X., Li, Q., Xie, C., & Peng, K. (2016). Quantum Entanglement Swapping between Two Multipartite Entangled States. Physical Review Letters, 117(24).

Chennai Water Resource Management Using GIS

Saral Purohit [a,1], Gadhiya Jay Dipakbhai [a] and G. Abirami [a]

ªDept of CSE, SRM Institute of Science and Technology, Tamilnadu, India

Abstract. Water scarcity is one of the biggest problems in the country of India, this problem is often caused by poor water resource management. Notably in India, the city of Chennai has had a water crisis going on for the past few years. On June 19th, 2019, the city officials declared that the city had run out of water. This was termed as 'Day Zero' where there was no water left. Most of the lakes of Chennai dried out. Poor water management and very less rainfall were the major reasons for this scarcity. Water resource management thus is of critical importance to ensure that water is utilized in the right way. So the aim of the proposed system is to evaluate the nature, significance, and rate of change in the water bodies of Chennai over a period of time using GIS/Remote sensing and assess the past and present conditions of water bodies in Chennai and to understand the dynamics and trends of change through various forms of spatial analysis. This is achieved by using different functionalities of ArcGIS and ArcMap and then analysing the obtained data and maps.

Keywords. ArcGIS, Water Resource Management, Spatial Analysis, Water Requirement

1. Introduction

The water resources in Chennai are mainly provided by Madurantakam Lake and Puzhal lake. The other small lakes near Puzhal such as Kaliveli, Pulicat and Sholavaram these all lakes are located in a 50km square area. However, they are now almost dry, according to the Chennai Metropolitan Water Supply and Sewerage Board. Chennai has approximately 40 lakhs (0.4 million) wells according to a research in 2012. People living in Chennai have faced a huge scarcity of water due to this crisis. And much of Chennai's water resources are outside the city, it needs better water resource management.

---

[1] Saral Purohit, Dept of CSE, SRM Institute of Science and Technology, Tamilnadu India
Email: gd7339@srmist.edu.in.

Figure 1. Puzhal lake in Chennai on June 2018 and April 2019

So the aim of the project is to evaluate the nature, significance, and rate of change in the water bodies of Chennai over a period of time using GIS/RS and assess the past and present conditions of water bodies in Chennai and to understand the dynamics and trends of change. Plus create an accurate and up-to-date map of all the water bodies of the city Chennai using GIS Technology with ArcGIS and ArcPy so the tracking of water resources can be much more efficient figure 1.

Geographical Information System is a computer science branch which provides the framework to analyze geographic data. It is also used to display data related positions on the surface of the Earth. The Geographical Information System applies in software and hardware components. GIS applications in water resources include storage of geographic data, hydro-logical management, modeling of groundwater, quality of water, water resource management.

This paper is organized as follows, section 1 is introduction about the research, section 2 is literature review, section 3 is system design, section 4 is implementation and section 5 are conclusion.

2.   Literature Survey

Cheng Wang [1] provides a real time water resource assessment using WRCC. Tian Yu [2] used the EPANET model which improves calculation efficiency and accuracy and does hydraulic calculation. Da Kong [3] provides a database inquiry for water information and improves the condition of heavy workload with low efficiency.

Rong Li [4] used a three-dimensional display using a GIS platform which provides hydrological analysis. Jiping Liu [5] used a multi water analysis model which helps in water resources allocation. Xuwei Ru [6] uses a Grid GIS which divides the map in rows and columns.

Tao Sun [7] also provides a 3-dimensional water resource system. The system breaks 2D display and is independent of the GIS platform. Lejiang Guo [8] provides water monitoring and decision analysis for water resources. It also provides social and environmental benefits. Yingzhan Hu [9] provides flood control decision support, demand analysis and prediction of water resources

Zhao Mai-huan [10] provides a data analysis function for water resource management which helps in supporting decisions. Qiu Agen [11] provides site selection for water resource management using a real time monitoring system. Majid Akram [12] visualizes the impacts of dewatering on the groundwater. The results show drastic but changes in water resources.

Xianwei Wang [13] uses video image classification and Global Precipitation Measurement for groundwater bodies. Chandan Kumar Singh [14] uses multi criteria analysis for groundwater bodies monitoring areas. A. D. Sheena [15] uses GIS and remote sensing to study the impact of the 2015 Chennai flood. It suggests some water management measures to avoid the impact of floods.
Ashwani Kumar Tiwari [16] uses GIS and water quality index to analyse the groundwater bodies of Ganga and Gomti rivers in India. Kuldeep Tiwari [17] uses an effective and cost saving method for rainwater harvesting in water bodies Rajasthan.

3. System Design

3.1 Study Area Analysis

Water resource management is one of the most major problems in Chennai. Chennai depends on four main major reservoirs for water which are Red Hills, Chembarambakkam, Cholavaram and Poondi. These reservoirs contribute up to 60 percent of the city's water requirements. But over the past few years, all water bodies in and around the city of Chennai have shrunk and dried up. The average levels of groundwater in the city have also declined rapidly over the fast years. Chennai receives rains in the months of October to January, apart from that there are showers throughout the year.

The city has also been a victim of floods in the year 2015 as well, this explains Chennai's uneven monsoon season over the years. Good water resource management is required in the city, this can be tracked and measured by implementing GIS (Geographical Information System) based models. GIS is excellent for revealing different patterns and dynamic changes that occur in geographical systems over time.

3.2 Methodology

The system works on the basis of data management, analysis, capturing, and visualizing data regarding position of points on the surface of Earth. By combining data from various points, GIS helps in understanding spatial data. GIS possesses various powerful tools for analysis this includes analysis of water resources as well. Therefore, a large amount of data can be stored and analyzed using GIS.

3.3 Architecture

Considering the need for mapping and tracking water bodies in a city, the water resource management system has a certain set of components needed for its proper functioning, so it's able to map and analyse all data points associated with the water

bodies of the city of Chennai. Therefore, the system would contain various different components such as -

1. Data Capture and Input Unit
2. GIS Data Files Unit
3. Data Analysis Unit
4. Data Visualization Unit
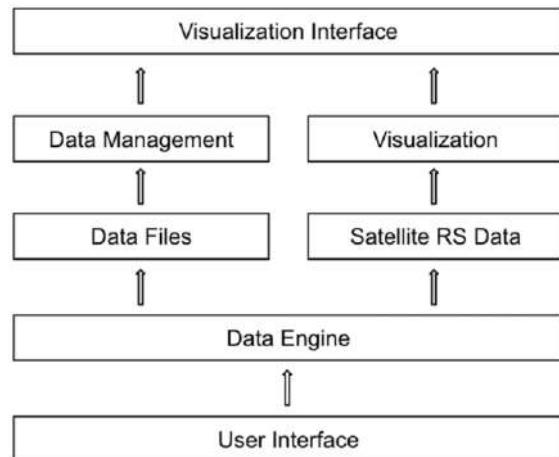
Figure 2. The workflow of the system

In Figure 2. the workflow of the system is shown. Visualization provides analytic tools which helps in 3D modeling and spatial analysis. Data Files include vector data, raster data and images of maps which is used for input analysis. User interface provides users input and output tools to interact with the system.

4. Implementation

4.1 Data Capture and Input

GIS is used to highlight spatial and geographic relationships among the map objects. The difference between GIS and other mapping systems is that GIS regularly analyze the map and keep the real time mapping updated. The weather features and their relationships all are key analysis of GIS. Thus, GIS is used to map new objects and variables. GIS can also predict the amount of water pollution in river basins. Spatial data include information of roads, mountains, rivers, etc. These data can directly be mapped in ArcGIS. Mainly the data is collected by satellites. Arc GIS allows every data in every format and can be laid on another map. Summing this information together forms Data Capture. Data obtained from satellites in the form of images and objects and converting them to digital format can be provided to GIS technology. Remote sensing is another tool which integrates data with GIS. Remote Sensing includes data from balloons, drones, satellites and planes.

4.2 GIS Data Files

GIS Data files are stored in a database obtained from satellite using software EPANET and ERDAS. By applying ArcView to the map, the map coordinates of the basin are noted. This is the basic data storage in GIS. The major type of data stored is in the format of raster and vector. Raster image represents a grid of pixels and images. They can be viewed in a bitmap display. Commonly used raster images are jpg, gif and png. While vector image is defined on a plan and connected by curves and lines. They are formed based on mathematical equations. And thus, advantageous when zooming in and zooming out plane remains smooth. Commonly used vector images are CDR, AI and SVG.

4.3 Data Analysis

GIS makes it easy to perform spatial measurements. Spatial relationship is a distance between two points. These measurements can be done easily by evaluating overlapping areas of two or more maps.

4.4 Data Visualization

Visualization is an essential way to visually communicate to others using tools, representation and methods. Visualization can be done in multiple ways, it could be numbers, tables, graphs and charts. For GIS, data visualization is used in spatial and geographical images. These images help in technical analysis of GIS. Many visual tools have been created to support knowledge discovery and data mining.

4.5 Spatial Analysis

Spatial Analysis is conducted for the generated map. Where various analysis techniques are utilized. Such as
1. Density Analysis
2. Watershed Analysis
3. Trace Analysis
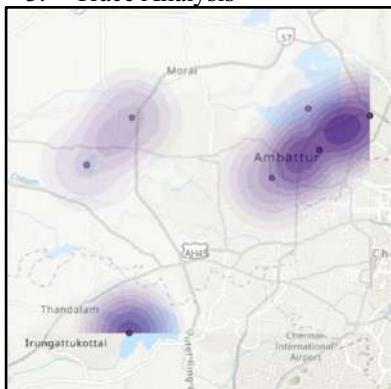
Figure 3. Density Analysis of lakes of Chennai   **Figure 4.**The Watershed analysis of lakes of Chennai
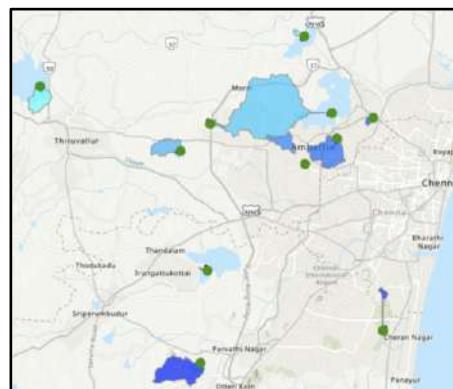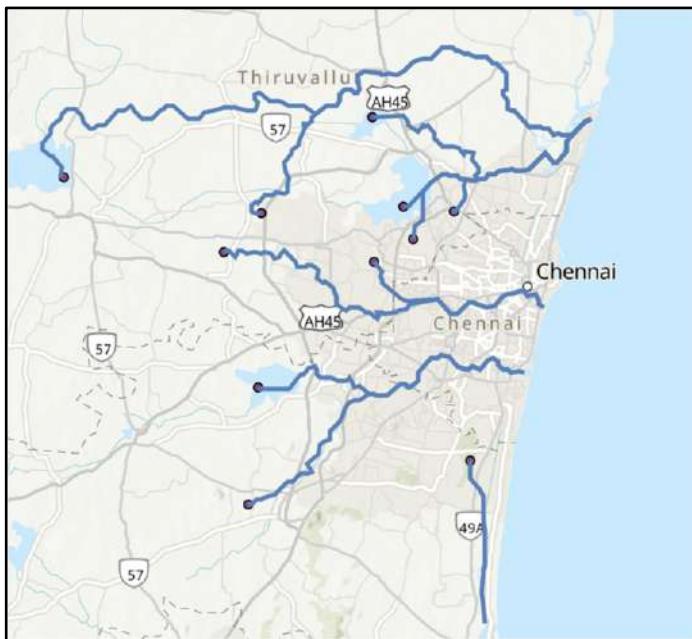
Figure 5. The Downstream analysis of lakes of Chennai

In Figure 3. Density analysis is conducted and it shows concentration and frequency of water bodies' presence across Chennai. In Figure 4. Water analysis is used to create catchment areas based on locations specified for all the water bodies present in Chennai. In Figure 5. Downstream analysis is used to determine the flow paths of all the water bodies in Chennai in a downstream direction.

5.   Conclusion

Based on the conducted research the idea of implementing a water resource management system using GIS in the city of Chennai. This system can be used for various purposes and projects. The maps and analysis produced by GIS can be quite useful to the government and city officials in getting a broader understanding of Chennai's water resources. It is necessary to continue this research as there are multiple avenues of improvement such as - incorporating different satellite data and using the system for disaster analysis such as flood analysis.

6.   Acknowledgement

References

[1]. Cheng Wang, "Development and Application of GIS-based Analysis System for Studying Water Resources Carrying Capacity in the Shule River Basin", IEEE Paper.

[2]. Tian Yu Ma Liya, "Construction of Water Supply Network Based on GIS and EPANET Model in Fangcun District of Guangzhou0", IEEE Paper

[3]. Da Kong, "Design of water resource management information system based on Web GIS", IEEE Paper.

[4]. Rong Li, Tao Sun, Guo Li, "Application of Three-dimensional GIS to Water Resources", IEEE Paper.

[5]. Jiping Liu, "Research on Optimal Allocation of Water Resources in the Qaidam Basin Based on GIS", IEEE Paper.

[6]. Xuwei Ru, Hongzhi Wu, "Research on Integration and Issuance of Water Resources Information based on Grid GIS", IEEE Paper.

[7]. Rong Li, Tao Sun, Guo Li, "Research on Three-dimensional Water Resources Information System Based on ArcGIS", IEEE Paper.

[8]. Lejiang Guo,Lei Xiao, "Application of GIS and Remote Sensing Techniques for Water Resources Management", IEEE Paper.

[9]. HU Yingzhan, "The Design of Monitoring Management Decision Support System of Water Resources Based on GIS", IEEE Paper.

[10]. Xu Chen-guang, "Geographic Information System Application for Water Allocation", IEEE Paper.

[11]. Qiu Agen, "A Site Selection Algorithm for Water Resources Points Based on GIS Technology", IEEE Paper.

[12]. Elgaali Elgaali; Majid Akram, "Mapping of dewatering impacts on ground water and soil quality using GIS", IEEE Paper.

[13]. Xianwei Wang 1, and Hongjie Xie, "A Review on Applications of Remote Sensing and Geographic Information Systems (GIS) in Water Resources and Flood Risk Management", MDPI, 2018

[14]. Chandan Kumar Singh, Yashwant B. Katpatal, "A GIS Based Design of Groundwater Level Monitoring Network Using Multi-Criteria Analysis and Geostatistical Method", 2017

[15]. A. D. Sheena, M. Ramalingam, B. Anuradha, "Remote sensing and GIS in the recent Chennai food study", 2018

[16]. Ashwani Kumar Tiwari, Abhay Kumar Singh, Mukesh Kumar Mahato, "Assessment of groundwater quality of Pratapgarh district in India for suitability of drinking purpose using water quality index (WQI) and GIS technique", 2017

[17]. Kuldeep Tiwari, Rohit Goyal, Archana Sarkar, "GIS-based Methodology for Identification of Suitable Locations for Rainwater Harvesting Structures", 2018

Data Privacy Preservation and Security Approaches for Sensitive Data in Big Data

Rohit Ravindra Nikam [a,1], Rekha Shahapurkar [b,2]

[a] Research Scholar,Computer Science and Engineering Department, Oriental
University, Indore, India

[b] Computer Science and Engineering Department, Oriental University, Indore,
India

Abstract. Data mining is a technique that explores the necessary data is extracted from large data sets. Privacy protection of data mining is about hiding the sensitive information or identity of breach security or without losing data usability. Sensitive data contains confidential information about individuals, businesses, and governments who must not agree upon before sharing or publishing his privacy data. Conserving data mining privacy has become a critical research area. Various evaluation metrics such as performance in terms of time efficiency, data utility, and degree of complexity or resistance to data mining techniques are used to estimate the privacy preservation of data mining techniques. Social media and smart phones produce tons of data every minute. To decision making, the voluminous data produced from the different sources can be processed and analyzed. But data analytics are vulnerable to breaches of privacy. One of the data analytics frameworks is recommendation systems commonly used by e-commerce sites such as Amazon, Flip Kart to recommend items to customers based on their purchasing habits that lead to characterized. This paper presents various techniques of privacy conservation, such as data anonymization, data randomization, generalization, data permutation, etc. such techniques which existing researchers use. We also analyze the gap between various processes and privacy preservation methods and illustrate how to overcome such issues with new innovative methods. Finally, our research describes the outcome summary of the entire literature.

Keywords. Data Privacy, Privacy Preservation Techniques, data mining, anonymization, l-diversity, bigdata.

1. Introduction

Privacy and security are essential issues in big data. A significant data protection paradigm isn't recommended in complicated systems because of which it is deactivated by nature. Nonetheless, details will still be quickly corrupted in the absence of it. As such, this segment focuses on topics of privacy and protection. Protection Records Protection is the luxury of getting any power over the processing and usage of confidential details.

_____

[1] Rohit Ravindra Nikam, Research Scholar Computer Science and Engineering Department, Oriental University, Indore; E-mail: rohitnikam3000@gmail.com
[2] Dr. Rekha Shahapurkar, Computer Science and Engineering Department, Oriental University, Indore; E-mail: rekhashahapurkar@orientaluniversity.in

Privacy preservation in data mining is a branch of research that is currently gaining popularity these days. Privacy-preserving is nothing but obtaining data mining algorithm results without compromising the underlying dataset. Data mining algorithms are analyzed whether they are causing any side effects on data privacy. The two-fold is the primary consideration during data mining privacy preservation.

First, all the sensitive raw data in the dataset, which may include contact details, names, addresses, and such demographic details, are identified and changed from the original database. Hence the receiver of this information will not compromise on the authenticity and privacy of other people's personal information. Second, the sensitive information that is being mined as a result of various data mining strategies should also be eliminated since it can also harm an individual's privacy.

The primary focus is on the difficulty of guaranteeing the privacy of information withdrawal results. The meaning of security must fulfil the requirements of clients of a sensible application. Two examples of such application are (1) credit or finance giver, whose clients may comprise of various shops as well as a small business, also who wishes to give them among a classifier that will discriminate which are creditworthy or risky customers, and (2) a medicinal corporation that desires to put out a study identifies the cluster of patients who act in reply another way to a route of action. These information proprietors wish to discharge information mining yield yet be guaranteed that they are not giving the personality of their customers endlessly. Suppose it might be confirmed to free result to withstand limits related to those set by k-anonymity. In that case, the praise donor might free a k-anonymous classifier and dependably declare that the confidentiality of individuals is confined.

The research provides data privacy and security to extensive unstructured data using effective classification and privacy algorithms. The flexibility of the system has measured in a different environment like distributed data systems etc. This work focuses on various existing techniques that used privacy preservation on extensive data with other privacy preservation techniques [1-3].

Existing methodologies
The Fran Casino et al. [1] illustrate blockchain as the Decentralized RS core, seeking to provide it with a wide variety of functionality while protecting consumer privacy. We are implementing a new design focused on decentralized area sensitive hashing classification and a range of suggestion methods, depending on how users handle data. Extensive study findings show the efficiency and efficacy of our methodology relative to cutting-edge approaches.
Chenyang Ma et al. [2] Describes Separating encrypted transactions in the block and calculating only bilinear pairings on node block cypher texts rather than all ciphertexts, which helps to reduce the computing costs that mining operation. Finally, we test the efficacy of our set of rules by conducting statistical analyses and simulator tests on numerical expenses, safety, precision, and time aspects. The results indicate our protocol would produce the right mining outcome and outperform the prior method in terms of performance under a similar protection standard of conditions.

Dongfeng Fang et al. [3] explained IoT network architecture, which contains heterogeneous IoT systems listed. Specific confidence models are developed and evaluated in the IoT framework depending on the confidence relationships between the different actors. We suggest a scalable and robust authentication scheme that considers heterogeneous IoT devices based on a paradigm needed for the least trust. The proposed method offers resource-limited IoT users protection and privacy in a scalable and effective way by using IoT users with improved storage and computational ability.

Anmin Fu et al. [4] Tackle this by introducing a novel Non-negative Matrix Factorization (NMF) outsourced scheme that seeks to reduce the computational pressure of consumers and resolve stable issues posed by NMF outsourcing. O-NMF specifically exploits Paillier homomorphism, focused on two non-collusion servers, to protect data privacy. O-NMF provides a testing system to support clients validate returned findings with a high degree of accuracy probability.

Muhammad Usman et al. [5] this system works in three steps. The first level edge devices affect a lightweight aggregation method to produced data during the first step. This approach limits the size of the data produced and seeks to protect data source privacy. In the 2nd tier, a multi-step process used for linking Level Two Edge Devices (LEDs) with High-Level Edge Devices (LEDs). The validation phase-only valid LEDs can move data to the LEDs, resulting in a reduction in the computational burden on LEDs. In the third level, the LEDs use a convolution neural network to predict the position of touching objects in LED data transmitted.

Jiannan Wei et al. [6] Propose a secure, safe and privacy-conserving IoT Message authentication scheme. Our framework embraces IoT devices with varying cryptographic settings and enables offline and online computing, making them more versatile and powerful than previous systems.

Rong Jiang et al. [7] Big Data processing, delivery, analysis, usage and sharing examined protection and privacy leakage possibilities. They recognized a Clinical Big Data Security and Privacy Leakage Possibility Predictor framework amid four leading indicators and thirty minor indicators. Additionally, weight for every variable was determined using the weight system GI and Entropy. The Fuzzy Method of Structured Analysis was recognized to check the principle of Big Data Medical Protection and Urban Privacy Computing.

Karen R. Sollinset. al. [8] explains requirements and limits and proposes a three-part decomposition of architecture. To arrive at this final analysis, we begin by clarifying the issues in the design space:

There is some agreement about what IoT means, particularly on the security and privacy consequences of various definitions.

1. We then consider the requirements and constraints on Big Data resulting from unique IoT system designs.
2. We examine the industry intricacies in parallel.

In this sense, we can then break down the set of drivers and data protection/privacy and innovation goals into1) the history of regularity and social policy; 2) the economic and industrial history; and 3) the context of technology and architecture.

Ismail Hababehet. al. [9] proposes an advanced approach for classifying and safeguarding extensive data while conducting versatility, duplication, and study. The data classification defines the requirement to protect extensive data accessibility into two categories; confidential and public as per the degree of hazard result of information. The effect of data protection is analyzed and authenticated on the subtle data within the framework of the classification group HDFS cluster.

Si Han et al. [10] proposes a Hidden community exchange key management protocol (SSGK) to avoid unwanted exposure to the contact channel and mutual data. Unlike previous books, the joint data is authenticated with a group key, and a secure distribution mechanism is used to spread the set of keys in SSGK. The detailed protection and efficiency analyses demonstrate that our collection of rules significantly minimizes data contribution's privacy security and hazards in cloud storage and saves on twelve percentage of storage space.

Xiaodan Yan et al. [11] proposed the GAN Model Attackable of deep learning execution; this approach mainly studies the information protection methods under GAN model attack to find a better way to prevent attacks and effectively protect information. Sometimes medical treatment data may be leaked to third-party organizations. When these essential medical data are illegally used by for-profit organizations or obtained by criminals, it will lead to the disclosure of personal privacy information and cause severe economic losses to the victims. However, the victim cannot delete the leaked report by itself or limit the scope and use of the information that has been revealed.

Gamage Dumindu Samaraweera et al. [12] Proposed Protection and Privacy Effects on Big Data Age Database Systems: A study, his paper analyses protection applications in today's essential database models rely primarily on security and privacy attributes. A collection of standard protection measures is defined and tested based on different protection classifications. This offers a thorough overview and study of the sophistication of protection and privacy frameworks in the database models coupled with possible orders/enhancements. Data owners can agree on the maximum appropriate data storage for their data-driven Big data applications.

Yanan Li et al. [13] Propose a data correlation framework effect on privacy leakage, defined as Previous Differential Privacy (PDP), which is recommended to determine information leakage considering the opponent's specific prior awareness. The model uses two methods for evaluating discrete and continuous results, respectively, the weighted hierarchical graph and the multivariate Gaussian model. This further shows the distinct effect of optimistic, negative, and mixed associations on data leakage. A closed structure definition of privacy leak extracted used for unbroken data considering general associations, and a chain law is introduced for separate data.

Yang Liu et al. [14] Preservation of privacy and aggregation through reinforcement learning suggests a Payment-Privacy Protection Level (PPL) game in which each member submit their sensing data along with a given PPL as a system selects the appropriate payment for the participants. In addition to removing the Nash equilibrium (NE) stage of the game. Consider a payment-PPL system is ambiguous; it uses a reinforcement learning strategy, i.e., Q-learning, to get the payment-PPL method in the complicated payment-PPL game. Here it uses a deep Q network (DQN), which combines deep learning and Q-learning to speed up learning.

Gamage Dumindu Samaraweera et al. [15] Proposed defence design review of leading database models of today, with a greater focus on security and privacy attributes. A set of standard protections is specified and assessed based on specific classifications of security. It offers a thorough summary and systematic analysis of the complexity of security and privacy technologies in the database models, together with potential directions/improvements. Data owners may select the most proper data store for their data-driven Big Data technologies models.

David Froelicher et al. [16] proposed Drynx, a Decentralized data management framework Knowledge of mathematical study of distributed databases. The author

relies on a group of computer nodes to allow data, such as standard deviation or severe, to be computed and to educate and test machine-learning models on critical and distributed data. Drynx incorporates collaborative protocols, holomorphic security, zero information evidence of validity, and differential protection to guarantee user confidentiality and service provider safety. This enables a practical and autonomous inspection of the entry data and all computation of the method, thereby ensuring audit ability in a transparent adversarial environment where no individual is involved has to be independently reliable.

DatThanh Dang et al. [17] proposes a trust-based MapReduce system for activities related to extensive data analysis. Precisely, we are first quantifying and proposing to allocate the critical values for map data and trust values than to minimize slots. Then we measure the faith factor of each tool engaged in the tasks of extensive data analysis. Based on the vulnerability level of a job for the data, the role needs a certain degree of faith (i.e., low important data require higher confidence degree servers/slots). The MapReduce scheduling difficulty is developed for the biggest weighted similar problem of a bipartite graph aimed at optimizing the cumulative confidence factor of all available assignments subject to specific confidence needed tasks.

Guangquan Xu et al. [18] suggested an original and versatile framework named So Protector to avoid privacy leakage by examining data stream linking Java and native layers. So, protector discovers a real-time system that detecting malicious features embedded in SO libraries. Placed, we extract the malware features through 3 steps:

1. Current as a grayscale picture binary files in the native family.
2. Using the ARM instructions package to reverse the SO file code using python to find the op-code sequence.
3. IDA Pro converts all files as an assembly language, which contains a .gdl files as an addition.

Dharminder et al. [19] Introduces a sign encryption strategy focused on identification by adding both encryption and signature that offers an answer to safe and authenticate communication in the Big Data location called SFEEC (Safety Mechanism for Energy-Efficient Computing). By providing pairing-free calculation at the end of the customer, SFEEC fulfils the criteria of fewer overhead computation and connectivity. SEC is also proved in the norm under attacks "indistinguishable from chosen-ciphertext" and "stable against chosen post" model.

Ruiyang Xiao et al. [20] Create a mixing scheme with one shared signature protocol without depending on third parties or having a transaction cost. The method utilizes a bargaining mechanism to maintain transparency of the agreement, supervised by the participants. The system also contains a signature procedure focused on the ElGamal signing procedure and key sharing.

2. Problem Description

It might be confirmed so as to free result with stand limits related to those set by k-anonymity, after that the praised on or might free a k-anonymous classifier along with dependably declare that the confidentiality of individuals is confined. Similarly, the author of a medicinal study quoting k-anonymous group centroids might be confident that they fulfil with confidentiality principles, which prohibit the make public of independently identifiable physical data.

Privacy Preservation Techniques

Privacy preservation in data mining is branch of research that is currently gaining lot of popularity these days. Privacy preserving is nothing but obtaining data mining algorithm results without compromising the underlying dataset. Data mining algorithms are analyzed whether they are causing any side-effects on data privacy.

So, the main motive behind preserving privacy during data mining is to build an approach which modifies the original dataset in such a way that even after performing mining process, the privacy of information remains intact.

K-Anonimity: Slice is fundamentally relying upon rows as well as columns parceling. In multiple attribute dividing, we apportioned information as {user_name}, {user_age, user_zip}, {user_disease} and tuple parceling (even segment) as {Att[1], Att[2]......Att[n]}. In trait parceling age and zip are apportioned together on the grounds that they both are exceedingly connected in light of the fact that is nothing but a Quasi identifier (QI). The group of QI should be known to assailant. While tuple dividing framework should check L assorted variety of the Sensitive Attribute (SA) fragment [21]. Algorithm has followed the below procedure.

1. First select the specific attribute set from D Att[] ⟵ D which illustrates the Att set from selected record;
2. While QI not present in record
    On the off chance that iSet⟵ n
    Check L decent variety;
    Else
        iSet++;
        Return Dview *;
3. QSet=QSet-{Dview*+att[i]};
4. Apply pruning stage 2 and 3 with read up to next tuple in Q
5. Dview*= Dview *UA[Datasize]/ Next Anonimized data from table instance

To begin with describe k = point of confinement of information anonymization container measure, number of lines, number of sections, exhibit rundown and database in the queue (step 1). Additionally, process will be done if and just if line isn't void i.e there should to be information in database. Check information for L assorted variety if row count = k = m (stage 2). At first Q= Queue of data. On the off chance that our can information satisfies k obscurity and L decent variety, it returnsDview* i.e anonymized perspective of information. The database information which can't satisfy prerequisite of protection will put away in exhibit list a [i]. Presently information stays in database i.e in Q = Q - {D*+ a[i]} (stage 3). Rehash stage 2 and stage 3. A[D] is anonymization of information in database. Apply above strides for outstanding information and make new anonymization see which is the association of unique examination and new one i.eDview * = Dview *UA [Datasize].

L diversity: L-Diversity is the process of two shows the number of occurrences of specific attribute in view table. Suppose system supports for L=10 it means it can take max L unique records in single virtual view [22].

Step 1: read each (R from Dataset)

Step 2: SpliteDataParts[] ←R.split()
Step 3: check TableViewLcount= L-Diversity count
Step 4: CurrentValue = R.get(Lvalue)
Step 5: If (CurrentValue.notexist (TableView))
TableView.add←CurrentValue
Step 6: else
         Go to step 1;
Step 7: If (TableView.count> L)
      Early stop
      Flag=1;
Step 8: continue
Step 9: if (flag==1)
RetrunTableView

Basically l-diversity illustrates the number of attribute of lenses in specific view. L-Diversity can be holding the privacy scenario during the data distribution. In secure multiparty computation protocol preferred the and privacy as well as hell diversity. In the proposed work system and illustrates the features of L-Diversity. e.g., Algorithm set the specific value for a liability of lenses in particular view in the first time of algorithm system read each row from data table and then data table has a split into the different column values. The desired column considered as a cause identifier for L-Diversity and then distributes the data in network environments. The L-Diversity upper limit checks the condition in algorithm in step 8. L-Diversity sometimes communicates with K- anonymity. Both combinations should be forcing the information breaching issue as well as data leakage.

Random Permutation: The permutation is the process which shuffles data attributes as well as privacy attribute randomly. This algorithm first decides the virtual fuel set value according to data type as well as number of rows. During the data distribution system Shuffle some sensitive information at specific attribute to another instances. The benefit of such random permutation approach, it generates half knowledge two traitors or attacker, which provides security from third party attacks like man-in-the-middle etc [23].

F-score for fitness: Basically F-Score is the function that evaluate the current Virtual Table view fitness spelling with normalize attributes. F-Score works with k anonymity as well as l diversity. Both the values were works with combine approach which generates the run time fitness score based on the occurrences [24].

Constraint C: Basically C-Constraint is the protocol which verify the security you which is generated by privacy approach. The two-approach carried out to achieve the privacy during the data distribution, first slicing and second randomization as well as generalization. C-Constraint is the policy which works for how much data we can publish in a single view. Basically, it works according to column and rows values. The column values called as L-diversity and rows values called as k- anonymity, C-Constraints evaluate policy when distribution mechanism add the data in virtual view. Below is the process to evaluate heat C-Constraint during the data distribution
    1. Define the K-anonimity as well as L-Diversity upper limit by system
    2. Define the privacy constraint minimum score fitness score=1;

3. If (CountVal<C.ConstrainValue)
      Privacy Leakage;
      Then premature stop;
   Else
      Return (Fitness_score);
4. Stop loop

Verification for quality of framework against number of suppliers: For check against number of suppliers, include one more property in anonymized information as a supplier to yield. This confirmation will demonstrate that our procedure of anonymization doesn't rely upon number of suppliers. Existing framework i.e supplier mindful anonymization calculation relies upon database and additionally supplier.

1. Create value of SA by means of Data Generator P= 1…..n
2. ensure for confidentiality constraint also Fscore=1 through number of provider
3. If
      Privacy Leack;
   next premature stop;
   Else
   Return (Fitness_score);
4. Exit

Consider a potential assault on shared information distributing. We utilized cutting calculation for anonymization and L assorted variety and confirm it for security and protection by utilizing double calculation of information security. Cutting calculation is extremely valuable when we are utilizing high dimensional information. It partitions information in both vertical and flat mode. Because of encryption we can expand security.Be that as it may, the restriction is there could be loss of information utility. Above framework can utilized as a part of numerous applications like clinic administration framework, numerous modern zones where we get a kick out of the chance to ensure delicate information like compensation of representative. Pharmaceutical organization where touchy information might be a mix of elements of solutions, in saving money segment where delicate information is account number of clients, our framework can utilize.It can be utilized as a part of military region where information is accumulated from various sources and need to anchor that information from each other to look after security.This projected framework helps to enhance the information protection and security when information is assembled from various sources and yield ought to be in cooperative mode.

We first formally depict our concern setting. At that point, we display our information protection definition as for a security imperative to forestall surmising assaults by information enemy, trailed by properties of this new security idea. Let T = {t1, t2, . . .} be an arrangement of records with similar qualities accumulated from n information suppliers P = {P1, P2, . . . ,Pn}, with the end goal that Ti are records given by Pi. Let AS be a delicate property with an area DS. In the event that the records contain various delicate traits at that point, we regard every one of them as the sole touchy property, while staying ones we incorporate to the semi-identifier. Nonetheless, for our situations we utilize an approach, which saves greater utility without giving up protection.

The objective is to distribute an anonymized T* while keeping any information enemy from surmising AS for any single record.An information foe is a coalition of information clients with n information suppliers participating to rupture protection of anonymized records. At the point when information is accumulated and joined from various information suppliers, primarily two things are done, for anonymization process. To shield information from outer beneficiaries with certain foundation learning BK, I expect a given protection prerequisite C is characterized as a conjunction of security imperatives: C1, C2...Cn.

If a gathering of anonymized records T* fulfils C, we say C(T*) = genuine. By definition C(ø) is valid and Ø is private. Any of the current protection standards can be utilized as a part requirement Ci.We presently formally characterize an idea of information security regarding a protection requirement C, to ensure the anonymized information against information enemies. The thought expressly models the intrinsic information learning of an information foe, the information records they together contribute, and requires that every QI gathering, barring any of those records possessed by an information foe, still fulfils C. It also demonstrates our security framework in which input information is given from various suppliers. Select point for cutting. Watch that information against security imperative C for information protection. Check additionally is cutting is conceivable or not. On the off chance that cutting conceivable at that point do it and if not then show the yield information. Our last yield T* are anonymzed information which will see just by validate client. Any for cannot break security of information. In this framework we are utilizing level and also vertical parcelling over database. Cutting calculation gives better segment parcelling. To comprehend this legitimately we should consider doctor's facility administration framework for explore. Let diverse offices are the suppliers who give information from various sources. We think about ailment as an AS (delicate property) and age and zip code are QI (semi-identifier) [25].

More Privacy approach

(i) **Explicit Identifiers** is an arranged of properties containing data that perceives a record director unequivocally, for instance, name, rate et cetera.

(ii) **Quasi Identifiers** is an arranged of properties that could possibly perceive a record administrator when joined with freely accessible information.

(iii) **Sensitive Attributes** is an arranged of properties that contains tricky individual specific data, for instance, sickness, pay et cetera.

(iv) **Non-Sensitive** Characteristics is an organized of properties that makes no issue if revealed even to plotting social events. Information anonymization empowers the exchange of data over a limit, for instance, between two offices inside an organization or between two offices, while diminishing the danger of unintended exposure, and in specific conditions in a way that empowers assessment and investigation post anonymization. With regards to medicinal information, anonymized information insinuates information from which the patient can't be recognized by the beneficiary of the data. The name, address, and full post code must be evacuated together with whatever other data which, in conjunction with other information held by or unveiled to the beneficiary, could distinguish the patient.

The bottom-up algorithm -

The base up algorithm is like the Top-down algorithm. The principal contrast is in the succession of coalition checks, which is in a base up form beginning from 0-foe, and moving up. The algorithm stops if an infringement by any foe is distinguished (early stop) or all m-foes are checked.

Algorithm displays the general thought of bottom-up speculation strategy. It starts the speculation from the crude information table T. At every emphasis, the algorithm voraciously chooses the best speculation g that limits the data misfortune and augments the protection pick up. This instinct is caught by the data metric ILPG(g) = IL(g)/PG(g). At that point, the algorithm plays out the speculation child (Best) → Best on the table T, and rehashes the cycle until the point that the table T fulfills the given k-anonymity necessity.

Bottom-Up Generalization: Bottom-Up Generalization is a proficient k-anonymity technique. In a k-mysterious informational collection, each record is indistinct from at any rate k-1 different records concerning QID. Fundamentally, Bottom-Up Generalization (BUG) approach of anonymization is an iterative procedure beginning from the most reduced anonymization level. We use the data/security exchange off as the scan metric for our approach, i.e., the Information Loss per Privacy Gain (ILPG). The Advanced BUG comprises of following advances, information segment, run the MRBUG Driver on apportioned informational index, joining the anonymization levels of the parceled informational collection and applying speculation to unique informational index with incorporated anonymization level without abusing the k-anonymity [21-22].

Algorithm Bottom-Up Generalization

1: while T NOT (assure a given k-anonymity necessity)
perform
2: intended for all simplification g do
3: calculate ILPG(g);
4: end for
5: get the Best generalization;
6: generalize T by Best;
7: end while
8: output T;

Let A(QID) and Ag(QID) be the base anonymity tallies in T when the speculation g. Given an information table T, there are numerous conceivable speculations that can be performed. However, most speculations g in reality does not influence the base anonymity tally. As it were, A(QID) = Ag(QID). In this manner, to encourage productively picking a speculation g, there is no compelling reason to think about all speculations. To be sure, we can concentrate just on the "basic speculations" [24].

The Top-Down Specialization (TDS) Privacy Approach

The Top-down algorithm checks the coalitions in a best down manner utilizing descending pruning, beginning from (n-1)-adversaries, and moving down until the point when an infringement by a m-adversary is distinguished or all G m-foes are pruned or checked.

In this technique we break down the adaptability issue of existing TDS approaches when we taking care of huge scale informational indexes on HADOOP platform. TDS is rehashed process which is beginning from the highest space esteems in the course of action trees of traits. Finding the best specialization, performing specialization and

updating estimations of the hunt metric. Such a procedure of TDS is rehashed until the point that k-anonymity is damaged, to depiction for the greatest information will use in that. The exemplary nature of a specialization is estimated by an inquiry metric the distinctive android application authorization is brought from android applications. These authorizations are utilized as dataset for process. In that we acknowledge the data pick up per protection misfortune (IGPL), a tradeoff metric that take as a primary concern both the security and data necessities, as the inquiry metric in our approach. A specialization with the most elevated IGPL esteem is viewed as best one and chose of each round. Whenever answer for Top-Down Specialization User may advance through every specialization to decide a coveted exchange off amongst protection and precision. Client may stop whenever and acquire a summed-up table fulfilling the anonymity necessity. Taking care of both downright and ceaseless traits. Progressively produce scientific categorization tree for constant qualities.

A few miniaturized scale information anonymization methods have been proposed. The most mainstream ones are speculation for k-anonymity and bucketization for 'l-decent variety. In the two methodologies, properties are parceled into three classifications:

1) A few properties are qualifier that can exceptionally recognize personnel, for example, Name or Social Security Digit;

2) A few properties are Quasi Identifiers (QI), that the enemy may definitely know (conceivably from other freely accessible repositories) and which, once considered together, can possibly recognize personnel, e.g., Birth date, Sex, and Zip code;

3) A few characteristics are Sensitive Attributes (SAs), that obscure to the foe and are viewed as touchy, for example, Disease and Salary.

Among speculation and bucketization, one initially expels identifiers from the information and after that segment's tuples into cans. The two methods contrast in the following stage. Speculation changes the QI-values in each container into "less particular yet semantically reliable" qualities with the goal that tuples in a similar can can't be recognized by their QI esteems. In bucketization, one isolates the SAs from the QIs by arbitrarily permuting the SA esteems in each container. The anonymized information comprises of an arrangement of pails with permuted delicate property estimations.

Cutting segments, the informational collection both vertically and on a level plane. Vertical parceling is finished by grouping qualities into segments in view of the relationships among the properties. Every section contains a subset of properties that are profoundly related. Even dividing is finished by grouping tuples into horizontal partitioning. At long last, inside each can, values in every section are haphazardly permutated (or arranged) to break the connecting between various segments. The essential thought of cutting is to break the affiliation cross sections, yet to safeguard the relationship inside every segment. This decreases the dimensionality of the information and jam preferred utility over speculation and bucketization. Cutting jam utility since it groups exceptionally associated properties together, and conserve the relationships between such traits. Slicing ensures protection as this disrupts the relationship among non-related traits, which are rare and consequently distinguishing. Consider that while the informational index comprises QIs and one SA, bucketization needs for disrupting connection; cutting, then again, can merge few QI qualities along through SA, protecting property relationships per delicate characteristic. The primary instinct that cutting gives security assurance is that the slicing procedure guarantees that for any tuple, there are for the most part numerous coordinating containers [22] [23].

Proposed system design

The privacy techniques follow various security techniques like slicing, anonymity, generalization, permutation, etc. When the system deals with a large amount of data, it works with a slicing approach where attributes are suppressed or general awaiting every line is the same through at slightest k-1 extra rows. At this point, the record is supposed to be k-anonymous. It can distinguish between access and non-access details of the Data set. The slicing approach is included in the column separation method. Anonymization of data is yet another approach that eliminates confidential Details to guard user privacy. It is also known as to de-identify. Whenever organizations disclose the data is publicly anonymized. The ideas of k-anonymity, l-diversity, and t-closeness have been used to keep information from re-identification presented.

Privacy Threats in Analytics

Privacy is a person's ability to determine what data can be shared and to use privileged access. If the data is in the public interest, it constitutes a threat to the user's privacy because the data holder owns the information. Data holders can be served for social networking, blogs, mobile phones, e-commerce platform, banks, hospitals, etc. The data holder is responsible for maintaining the protection of consumer data. Besides the data held in the public domain, directly or indirectly, users are contributing to data leakage. For example, most mobile apps seek access to our contacts, flesh, camera, etc., and we agree to all terms of service without reading the privacy policy by leading to the leakage of data. Therefore, there is a need to inform users of smartphones about risks to safety and privacy. Many of the main dangers to privacy include (1) surveillance; (2) Disclosure; (3) discrimination; (4) Personal embracement or abuse.

Surveillance

Many organizations, including retail, e-commerce, etc., study their customer's buying habits and try to develop various offers and value-added services [4]. Based on the opinion data and sentiment analysis, social media sites provide recommendations for new friends, places to be seen, people to be observed, etc. This is possible only when they continuously monitor their customer's transactions. This is a severe privacy threat as no Individuals accept oversight.

Disclosure

Find a hospital keeping data that includes (Zip, gender, age, illness) [5–7]. The data holder has released data for analysis to a third party through anonymization of sensitive data Person shall specify data so as not to identify the person. The third-party data analyst can map this information with the publicly accessible external data sources, including census data, and place a person with a particular disorder. This is how the private data of A person who is deemed to be a severe breach of privacy may be disclosed.

Discrimination

Discrimination is the bias or injustice that may arise when any person's private information is revealed. Statistical analysis of election outcomes, for example, has proved itself that people of one community were entirely against the government-forming party. Now the government may ignore, or discriminate towards, the culture.

Personalized embrace and abuse

This can also lead to intimate embrace or harassment if any person's private information is revealed. A guy, for example, was undergoing treatment privately for some particular problem and regularly bought some medicines from the doctor's shop. The medical shop can submit some of them as part of its daily business model Reminder and offers relating to these medicinal products by the cell when one member of the family has Seen that would result in the personal embrace and even abuse [8].

Activity in data analytics will affect data privacy. In many countries, privacy is enforced on the Lawson Preservation. Lack of awareness is also a big reason for the attacks on privacy. Because Example: Many smartphone users do not know the information stolen from Multiple apps to their phones. Previous research shows that only 17% of Mobile users are Aware of the privacy risks [9].

Few basic things have considered by various privacy techniques during data broadcasting without any existing security approach.

1: Key attributes: Based on the attributes uniquely identifies tuples. Ex: Social security number, pan number, adhar number, voter id, driving license number, etc.

2: Quasi Identifiers: An arrangement of traits that can be conceivably connected with outside data to re-distinguish entities. Ex: Zipcode, date of birth, sex.

3: Sensitive attributes: Some of the features contain sensitive value concerning the data owner. Ex: salary and disease.

4: Non-sensitive attribute (NSA): Disclosing the non-sensitive attributes will not break the secrecy of the user.

K-anonymity: It can be used to prevent record linkage. To preserve privacy, the following Anonymization techniques are applied to the data [9,10, 11].

Suppression: Quasi-identifiers is supplanted or darkened by some steady qualities like 0, *, etc. Ex: some values license number, aadhar number can be invisible using an asterisk.

Generalization: Some values are replaced by parent values [3] [4].

The above techniques used to achieve privacy on heterogeneous datasets when data has broadcast in a vulnerable environment.

3. Results and Discussion

As part of a systematic analysis of literature, it was found that all current privacy protection measures are about structured data. About 80% of the data is unstructured that it is produced today. As such, the following needs to be discussed Challenges ahead. Create practical approaches to protect privacy in structured and unstructured ways. Scalable and reliable techniques to be built for heterogeneous handling of large scales datasets. The data should be allowed to remain in its native form without transformation, and data analytics can be carried out while safeguarding privacy. Advanced innovations must be developed aside from anonymization to ensure protection against critical threats to privacy, including disclosure of identity, discrimination, surveillance, etc. Optimizing the value of data while maintaining the confidentiality. In below table we demonstrated an evaluation of system with numerous existing systems.

Table 1: Comparative analysis between proposed V/S existing approaches with different supplied data size

Method	100 kb	200 kb	300kb	500 kb
Privacy approach using top-down generalization	246	488	723	975
Enhanced Slicing Models for Preserving Privacy in Data Publication	310	602	923	1178
Privacy using Anonimization approach	580	952	1533	2701
privacy protection and fingerprint generation (Proposed)	235	400	650	800

The above Table 1 depicts the time performance to generate the privacy view of proposed system with some existing approaches. The proposed approach is improving the time around 5% over all existing approaches.

4. Conclusion

We have shown in this article that various techniques and methods of privacy preservation, i.e., k-anonymity, l-diversity, t-closeness, provide strong privacy in big data. No practical approach has yet been built on unstructured data. Standard classification and clustering problems can be applied with data mining algorithms but can't protect privacy, mainly when dealing with specific individual information. It's information. It could be used to improve machine learning and soft computing techniques, new and more acceptable solutions to privacy issues, including disclosure of identities that may lead to personal awkwardness and abuse. There are several ways of working in the future. To protect privacy in the future, when interests of the data and data variety enhance, novel analysis is applied.

Acknowledgement

References

[1]. Casino F, Patsakis C. An Efficient Blockchain-Based Privacy-Preserving Collaborative Filtering Architecture.IEEE Transactions on Engineering Management.2019 Oct 22.

[2]. Ma C, Wang B, Jooste K, Zhang Z, Ping Y. Practical Privacy-Preserving Frequent Itemset Mining on Supermarket Transactions. IEEE Systems Journal.2019 Jun 26.

[3]. Fang D, Qian Y, Hu RQ. A Flexible and Efficient Authentication and Secure Data Transmission Scheme for IoT Applications, IEEE Internet of Things Journal.2020 Feb 3.

[4]. Fu A, Chen Z, Mu Y, Susilo W, Sun Y, Wu J. Cloud-based Outsourcing for Enabling Privacy-Preserving Large-scale Non-Negative Matrix Factorization. IEEE Transactions on Services Computing.2019 Aug 28.

[5]. Usman M, Jolfaei A, Jan MA. RaSEC: An Intelligent Framework for Reliable and Secure Multi-Level Edge Computing in Industrial Environments. IEEE Transactions on Industry Applications.2020 Feb 20.

[6]. Wei J, Phuong TV, Yang G. An Efficient Privacy-Preserving Message Authentication Scheme for Internet-of-Things.IEEE Transactions on Industrial Informatics.2020 Feb 10.

[7]. Jiang R, Shi M, Zhou W. A Privacy Security Risk Analysis Method for Medical Big Data in Urban Computing. IEEE Access. 2019 Sep 24; 7:143841-54.

[8]. Sollins KR. IoT big data security and privacy versus innovation. IEEE Internet of Things Journal. 2019 Feb 15;6(2):1628-35.

[9]. Hababeh I, Gharaibeh A, Nofal S, Khalil I. An integrated methodology for big data classification and security for improving cloud systems data mobility. IEEE Access. 2018 Dec 28; 7:9153-63.

[10]. Han S, Han K, Zhang S. A Data Sharing Protocol to Minimize Security and Privacy Risks of Cloud Storage in Big Data Era. IEEE Access. 2019 May 3; 7:60290

[11]. Shuo Zhang, Yaping Liu, Shudong Li, Zhiyuan Tan, Xiaomeng Zhao, Junjie Zhou, "FIMPA: A Fixed Identity Mapping Prediction Algorithm in Edge Computing Environment", Access IEEE, vol. 8, pp. 17356-17365, 2020.

[12]. Samaraweera GD, Chang MJ. Security and Privacy Implications on Database Systems in Big Data Era: A Survey. IEEE Transactions on Knowledge and Data Engineering.2019 Jul 18.

[13]. Y. Li, X. Ren, S. Yang and X. Yang, "Impact of Prior Knowledge, and Data Correlation on Privacy Leakage: A Unified Analysis," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 9, pp. 2342-2357, Sept. 2019.

[14]. Liu Y, Wang H, Peng M, Guan J, Xu J, Wang Y. DeePGA: A Privacy-Preserving Data Aggregation Game in Crowdsensing via Deep Reinforcement Learning. IEEE Internet of Things Journal.2019 Dec 3.

[15]. Samaraweera GD, Chang MJ. Security and Privacy Implications on Database Systems in Big Data Era: A Survey. IEEE Transactions on Knowledge and Data Engineering.2019 Jul 18.

[16]. Froelicher D, Troncoso-Pastoriza JR, Sousa JS, Hubaux JP. Drynx: Decentralized, Secure, Verifiable System for Statistical Queries, and Machine Learning on Distributed Datasets. arXiv preprint arXiv:1902.03785. 2019 Feb 11.

[17]. Dang DT, Hoang D, Nguyen D. Trust-based Scheduling Framework for Big Data Processing with MapReduce. IEEE Transactions on Services Computing.2019 Sep 3.

[18]. Potharaju, S. P. (2018). An Unsupervised Approach For Selection of Candidate Feature Set Using Filter Based Techniques. Gazi University Journal of Science, 31(3), 789-799.

[19]. D. Dharminder, M. S. Obaidat, D. Mishra, and A. K. Das, "SFEEC: Provably Secure Signcryption-Based Big Data Security Framework for Energy-Efficient Computing Environment," in IEEE Systems Journal March 2020.

[20]. Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.

[21]. Arava K, Lingamgunta S. Adaptive k-anonymity approach for privacy preserving in cloud. Arabian Journal for Science and Engineering. 2019 Jul 16:1-8.

[22]. Yao L, Chen Z, Hu H, Wu G, Wu B. Sensitive attribute privacy preservation of trajectory data publishing based on l-diversity. Distributed and Parallel Databases. 2020 Nov 17:1-27.

[23]. Chen J, Liu G, Liu Y. Lightweight privacy-preserving raw data publishing scheme. IEEE Transactions on Emerging Topics in Computing. 2020 Feb 17.

[24]. Yin C, Shi L, Sun R, Wang J. Improved collaborative filtering recommendation algorithm based on differential privacy protection. The Journal of Supercomputing. 2020 Jul;76(7):5161-74.

[25]. Batmaz Z, Kaleli C. Methods of privacy preserving in collaborative filtering. In2017 International Conference on Computer Science and Engineering (UBMK) 2017 Oct 5 (pp. 261-266). IEEE.

MISP: Model for IaaS Security and Privacy

IndraKumar Sahu[a 1], Manisha J Nene[a],

[a]*Dept. of Computer Science & Engineering, Defence Institute of Advanced Technology,*
Pune, Mahaashtra,India

Abstract: Paradigm shift towards cloud computing offers plethora of advantages both for cloud users and Cloud Service Provider (CSP). For cloud users, it offers saving of cost, scaling of resources, pay per use, elastic and on-demand services. On the other hand, it offers centralized resource management and provisioning of operations, safety and security for CSP. By holding multiple virtual IT resources (CPUs, storage servers, network components and software) over the internet, Infrastructure-as-a-Service (IaaS) serves as fundamental layer for all other delivery models. Along with benefits of IaaS, there exists several security and privacy issues and threats to confidentiality, integrity, authentication, access control and availability. In this paper, detailed study of IaaS components, associated security and privacy issues are explored and counter measures for the same are determined. Furthermore, as a result of the study, Model for IaaS Security and Privacy (MISP) is proposed. The model presents a cubical structure and adds more features than the existing models to enhance the security and privacy of data and operations and guide security assessment for safer adoption by enterprises

Keywords. Cloud Computing, Cloud Security, Cloud Deployment Models, Service Level Agreement, Model for IaaS Security and Privacy (MISP), IaaS, Virtualization

1. Introduction

Since the inception in late 1960s, cloud computing became a ubiquitous technology with hardware, software, computational and operational IT resources and services delivered via Internet to the users [1]. Elasticity, scalability, on-demand resources, cheap operational expenses, location and device independence andpay per use business model are the merits for its prime attraction [2]. Cloud computing has provided huge opportunity to migrate from maintaining, securing and operating own standalone, on-premise resources like infrastructure and applications to cloud. Recently, it attracted very considerable attention of academicians, industry people and researchers.

As highlighted by NIST [7], cloud computing has three service models and four deployment models.

---

[1] IndraKumar Sahu,Dept. of Computer Science & Engineering, Defence Institute of Advanced Technology
Email: ikambition.s23@gmail.com

1.1 Service Models

IaaS with resources like data storage servers, computing hardware and networking components provides infrastructures to users to facilitate management of OS and applications.

Platform-as-a-Service (PaaS) where users are provided with an environment to develop, create and use their own tools and software applications.

Software-as-a-Service (SaaS) with readymade application software and tools are delivered to the users with licenses to use remotely without buying them completely.

1.2 Deployment Models

The deployment models define the way cloud may be used. The different models offer varying resources and the cloud users can adopt the one that suit them the best based on their needs and budgets. Four types of deployment models are as under.

Private cloud offers cloud resources and infrastructure to be used as stand-alone resources with greater control over security and data backup facility.

Public cloudoffers shared resources at lower cost but the security and privacy of the data and storage lowered as compared to the private cloud.

Hybrid cloudshows the best of first two models in terms of resources, controls and the cost. The security and privacy are in between that of private and public cloud.

Community cloudoffers shared resources amongst the same types of organizations like banks, hospitals etc.

1.3 IaaS Model

Cloud computing primarily depends upon IaaS delivery model that provides rudimentary operating systems, networking components, security infrastructure and data servers for designing and developing required applications, databases, development tools and services [8]. The Oracle and KPMG Cloud Threat Report 2020 [4] shows the recent adoption trends for cloud computing. As compared to 62% in 2018, in 2020 76% of on-premise business-critical applications migrated to IaaS through 'lift-and-shift' approach. Being the fastest growing sectors amongst all other service model, IaaS is expected to grow to $63 bn in 2021 from $ 23.6 bn in 2017 at a rate of 27.6 % according to Gartner [5]. It also predicted that by 2025, 80% of the enterprises will use IaaS as compared to 65% in 2017.

On demand services and scalable resources with advanced technical capabilities are provided to the users in IaaS model. Hardware comprising of storage servers, networking components and computing hardware (CPUs, RAM, graphic cards etc.) and software like cloud Application Program Interfaces (APIs), Utility Interfaces (UIs), hypervisors, software modules, security and control management modules are two types of components. Quality of Service (QoS) is an important factor and is made part of legal contract [5]. The IaaS model can also be viewed as shown in Figure 1 below [9].

Figure 1: Virtual and Physical layer in IaaS [9]

From rigorous study on security and privacy issues of IaaS delivery, a Model for IaaS Security and Privacy (MISP) is proposed adding various IaaS components to mitigate the threats of the delivery model.

The rest of the paper is presented in four sections. Literature survey is in the second section. Third section is of preliminaries. The fourth section is for the proposed model MISP with details for enhancing the security and privacy in IaaS against existing vulnerabilities and threats. The fifth section concludes the paper along with future speed.

2. Literature Survey

2.1 Related work

The security and privacy issues being faced in IaaS are related to Confidentiality, Integrity, Authentication, Availability and Access control (CIAAA) and a lot of research work has been done to mitigate these issues. Ravi et al. [10] carried out a sincere assessment of threats to security in IaaS along with responsibilities of cloud user and CSP. Their work mainly highlighted the issues in CIA triad and proposed possible solutions. The latest threats and focused malicious approaches are not addressed.

Ahmed et al [9] presented brief of issues in IaaS components and analysed CSA top twelve threats in the model along with possible solutions for them. The threats mentioned in CSA report gets changed from time to time and hence are not very relevant at present time as per CSA report 2020 [4].

Cullum et al [35], in his paper presented host hypervisor security issues in public IaaS and their solutions. The detailed study on hypervisor gives out known attacks that exist in hypervisor shared environment. The solutions are focused mainly on virtualization related issues while other threats are not addressed.

Moutai et al [24], presented a secure architecture-based distributed testing to confirm CAA based on QoS. It is limited to information security. The parameters like security of storage, network and hardware are not tested.

Dawoud et al [8], presented IaaS security model with issues related to components, suggested secure policies along and restriction levels. The security model is limited to some issues only whereas, with the advancement in cyber spectrum, there is need of addition of latest issues and threats.

2.2 Contributions

Our paper presents a comprehensive cubical MISP that comprises of components related issues along with associated threats to IaaS model; each in the common plane of

cloud user and CSP. There are rules and policies to enhance the security and privacy of data and operations in second plane. The third plane of cubical presents levels of rules and policies for implementation varying from lenient level to strictest level. The model summarizes all threats and possible ways out to enhance the security and guides security assessment for safer adoption of IaaS delivery artist.

3. Preliminaries

3.1 Service level agreement (SLA)

SLA is a legal document agreed and signed between CSP and a cloud user to describe the legal responsibilities, liabilities for both of them and define QoS offered by the CSP as part of the agreement [12]. It makes a mention of both the required and the expected level of services to be delivered maintaining availability and security and privacy with review or monitoring of the SLAs, riders and liquidation terms and time span of contract.

3.2 Virtualization of Platform

Virtualization is a process of abstracting and sharing a single hardware that facilitates aggregating multiple stand-alone computing resources like CPUs, memory, storage and network components [8]. A typical example is 'Server virtualization' in which several attributes of physical server is hidden and they are reproduced in a hypervisor in the form of virtual CPU (vCPU), Virtual RAM (vRAM), virtual NIC (vNIC) and virtual disks. It has two important characteristics namely, multi-tenancy and scalability. The virtual and physical layers in the model are illustrated in figure 2 below.

Figure 2: Virtual and Physical layers in IaaS

Two types of virtualization namely OS based in which a software is installed in host OS and hardware based that refers the installation directly on the physical host hardware [14].

3.3 Utility Computing

Utility computing provides multiple resources on demand as per user's request. Various IT resources are packed for metered services and then provided to cloud users at low cost and as pay-per-usage basis with scalability support even if demand reaches to its peak [8].

3.4 Cloud Scalability

Cloud scalability being one of the basis of cloud computing, offers homogenous resources with infinite scalability at linear increase of performance; the answers to when, what and where to scale in multi-tier service-oriented applications in autonomic scaling [14].

4. Model for IaaS Security and Privacy (MISP)

This Security of any service model in the cloud depends on the security of the infrastructure. Various components in IaaS are required to be looked into for user's satisfaction. Multiple agencies undertake works related to threat assessments on privacy and security on cloud computing. Distributed Management Task Force (DMTF), Open Cloud Consortium (OCC) and Cloud Security Alliance (CSA) are some of them that define standards, certifications and practices to ensure a secure cloud environment [18].

We propose a Model for IaaS Security and Privacy (MISP) in cubical form with three planes defined as shown in figure 3. The first plane gives out components of IaaS.
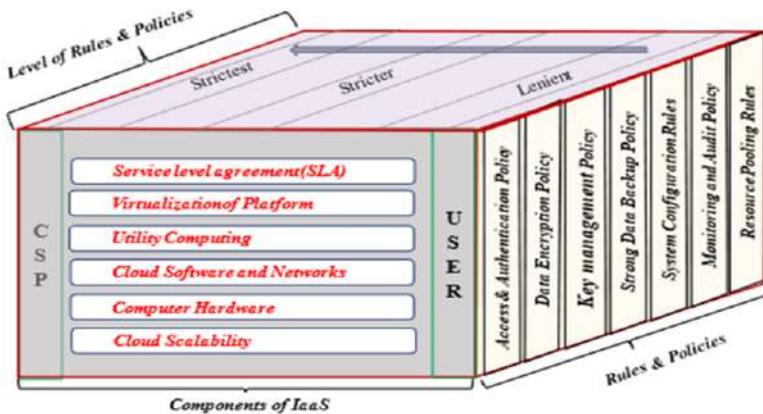
Figure 3. Model for IaaS Security and Privacy (MISP)

The cloud user and CSP are common participants of the plane and they generally share responsibility in maintenance of security and privacy of the model.

4.1 Threats Related to Components of IaaS

The plane consisting of components of IaaS in the proposed model is analyzed and threats associated with the solutions are described.

SLA Related Issues: Lack or non-existence of standardization in creating and performing the SLA between the involved parties creates big loopholes. The leading CSPs like Amazon (AWS), Google (GCP) and SalesForce hide numerous parameters regarding data safety and preservation in their proposed SLAs [6]. SLA may get exposed to vulnerabilities if any misunderstanding amongst the parties arises. So, it becomes imperative to detect user's concerns on priority [25]. The review and study of the environment displays several threats as per CSA classification.

Data breach and usage monitoring of data stored in the cloud is possible through human errors, application vulnerabilities, inadequate security practices or targeted attacks. Strong encryption techniques, prevention of leakage of secret data using neural networks [3], Data Leakage Prevention for Data in Transit using Artificial Intelligence and Encryption Techniques [34], strong backup and retention strategies and use of strong APIs [20] can mitigate this issue.

Insufficient due diligence while transferring responsibility of cloud control and cloud security to the CSP is a threat that is caused due to lack of transparency in security mechanism applied by CSPs [19]. Un-sanctioned application usage and   sanctioned application misuse are the two key challenges in this threat. Strong key management [17], and use of SLA for cloud visibility are solutions for this issue.

Denial of Service (DoS) is a threat mainly due to external agents that can cause unavailability of resources to the cloud users in the form of network, application or bandwidth denial [18]. The threat can be mitigated with regular audits of log and monitoring of services with advanced methods like Software Defined Networks (SDN), EDoS and SEDoS [30].

Many of the researchers argue to propose Web Service Level Agreement (WSLA) that can manage SLAs in IaaS environment [36]. More conveniently, SLA monitoring and enforcement may be delegated to a third party to bridge the trust deficiency between the CSPs and the cloud users.

Virtualization Related Issues: Virtual-aware security is required to face the security issues in IaaS [15]. Three types of possible threats are determined here.

Threats from host Operating System: The host OS being privileged domain can monitor, configure, communicate and modify data or services and hence may cause threats to IaaS model. According to MacAfee Cloud Adoption and Risk Report [21], the average organization has 14 misconfigured IaaS instances at any given time making 2269 instances per month. 5.5% of AWS S3buckets in use are misconfigured. Strong data backup and retention techniques [22] and multi factor authentication can mitigate the threats.

Communication between host and the VM is through virtual network or shared virtual resources and hence vulnerable to threats. An attacker could exploit important features like Clipboard to monitor the activities between them [25]. In case of host being compromised, all the VMs get into risk of any kind of possible attacks. Trusted Virtual Domain (TVD) for infrastructure and security mechanism [29], Trusted Cloud Computing Platforms (TCCP) for confidentiality [31], VLAN for network

virtualization and Identity Based Integrity Verification (IBIV) protocol for data integrity [13] are the solutions for such threats and issues.

Threats from VMs hosted on the same host: CSP provides API to carryout management functions such as provisioning, replication and decommissioning of resources on IaaS. But these insecure ill-designed, broken, exposed or hacked APIs and user interfaces (UIs) may lead to data breach or other security threats.Data Leakage Prevention for Data in Transit using Artificial Intelligence and Encryption Techniques [34], Scarce Attack Datasets and Experimental Dataset Generation [27], multi-factor authentication and robust authentication mechanism [33] can mitigate these issues.

Other possible attack on virtualization platform isVM Escape in which isolation layer between host and VM is broken to get the access of hypervisor's root privileges. As the attacker gets control over the host OS, he can use the compromised OS to manipulate control as per his desires through covert channel for malicious code execution.

Network virtualization partitions or aggregates a collection of network resources and present them a unique and isolated physical view to the users. Communication between VMs is through network virtualization in a direct and efficient manner. To avoid attacks like sniffing, SQL injection and spoofing on virtual network, secure physical channels can be adopted.

Utility Computing Related Issues: The utility computing faces the challenge of complexity in cloud computing. A bigger CSP may lease the services to second level CSP who in turn provides metered service to users. For example, Amazon DevPay5 from Amazon is a second level CSP. In this, the second level CSP might use services and user may be charged for what he has not used. Strong multi-tier passwords and two-factor authentication mechanisms [23] maybe used to mitigate this issue.

Cloud Software and Network Related Issues: In IaaS model, CSP provides cloud software and networks. Open-source cloud software like Eucalyptus and commercially proprietary software are two options but security from vulnerabilities and bugs cannot be ensured in either of the two. Cloud providers either furnish APIs or web service protocols like XML Simple Object Access Protocol or simply SOAP to grant access to cloud users to orchestrate management functions.

Computing Hardware Related Issues: A pool of shared distributed physical resources is provided to cloud users through virtualization in IaaS. Threats and attacks in on-premises hardware scenario occurs internally as a study shows it to be 70% [16]. Threats can be categorized in various ways. Based on type of resources: threats to physical computing resources like CPU, monitor, other physical machines and threat to storage resources where attacker gets access of the data storage.

The other one is based on type of adversary: insider and outsider attackers. Insider attackers have access to the resources of the organization and can cause damage intentionally or otherwise [26]. The outsider may be any hacker or bot to damage the system. Policy Enforcement Points (PEPs) side caching [28], inclusion of human resource management is some of the mitigation techniques.

Management of various changes in internal, system practices and Identity and Access Management (IAM) affects identity, credentials, key and access management. Strong end to end encryption, multi-tier passwords and multi factor authentication, and LDPC decoders [11] are measures to mitigate it.

Cloud Scalability Related Issues: IaaS resources can be scaled as per the user requirements. While doing so, there is a threat of account hijacking and abuses to breach infrastructure through spam mails, social engineering, vishing and phishing. Strong encryption techniques, multi-factor authentication [23] for integrity and strict monitoring of unauthorized activities may help to tackle this issue.

4.2 Rules and Policies for IaaS

The rules and policies for security and privacy are presented in the second plane in a vertical axis that implicates their presence through all components of IaaS. They are as mentioned below.

1. *Access and Authentication Policy:* to restrict any unwanted and unwarranted users to get access and verify the authorized users of the IaaS delivery model.
2. *Data Encryption Policy*: to ensure confidentiality, integrity and authentication in IaaS model using strong encryption techniques.
3. *Key Management Policy:* to enforce no loss and misuse of keys used in the IaaS for various purposes.
4. *Strong Data Backup Policy:* to avoid loss, deletion, tampering or theft of data in event of any unprecedented natural disaster, data corruption or cyber-attack.
5. *System Configuration Rules:* to avoid system misconfiguration, system bugs and internal or external attacks through exploitation.
6. *Monitoring and Auditing Policy*: to prevent any intrusion, system failure, status of software, untoward event and possible security breaches.
7. *Resource Pooling Rules:* To utilize the resources available with CSP for users as per demand optimally and judiciously.

4.3 Levels of Rules and Policies

The third plane is level of rules and policies. The level of rules and policies implementation need to be based on judicious scrutiny of security of data and operation on IaaS infrastructure, expertise of the user and the environment. If the data and operation are of critical in nature, the strictest level to be followed. In case of normal or low value data and operations, lenient level may be implemented. Since the strictest level might be slow and time consuming, the levels may be decided accordingly.

The proposed model is an attempt to standardise the IaaS layers, various components present in the model that are threatened and rules and policies to mitigate the threats, issues and challenges. Level of rules and policies implementation suggest degree to enhance the privacy and security accurately traded off between operational time and required security.

5. Conclusion and Future Scope

IaaS delivery model provides the basis for all other models and faces the security issues across hardware and software. Virtualization is core of the IaaS model for isolation. The security and privacy issues arise due to numerous reasons like lack of adequate knowledge, complex policies, technical glitches, system errors, standardization, certification and violation of established policies and practices. In this paper, security issues associated with IaaS components are investigated. The security

issues related to security of each IaaS components and proposed countermeasures are provided. The proposed MISP summarises all the issues and possible ways out to secure IaaS model to enhance the security and guide security assessment for safer adoption by enterprises. Cryptography and the best available techniques-based solutions are proposed to mitigate the threats to manage and secure the cloud in an optimal manner

Due to phenomenal rise in computing capabilities, the existing issues and challenges may get aggravated to unimaginable level of difficulties. New technologies like Network as a service (NaaS), Cloud of Things (CoT) etc. may pose different challenges. Timely review of the issues with the changes in policies and procedures will be warranted.

Another imminent threat is from quantum computing that possess extremely high computing capabilities. So, the security and privacy concerns of IaaS are required to be seen in the prism of quantum threats. The future work may be carried out to find quantum solutions in cloud computing for the post quantum era.

References

[1]   Madni, S. H. H., Abd Latiff, M. S., & Coulibaly, Y. (2016). Resource scheduling for infrastructure as a service (IaaS) in cloud computing: Challenges and opportunities. Journal of Network and Computer Applications, 68, 173-200.
[2]   Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. Journal of network and computer applications, 34(1), 1-11.
[3]   Ghouse, M., & Nene, M. J. (2020, June). Graph Neural Networks for Prevention of Leakage of Secret Data. In 2020 5th International Conference on Communication and Electronics Systems (ICCES) (pp. 994-999). IEEE.
[4]   Oracle and KPMG "Cloud Threat Report 2020", https://www.oracle.com/in/a/ocom/docs/cloud/oracle-cloud-threat-report-2020.pdf, assessed on 17 Dec 2020.
[5]   Gartner, Infrastructure-as-a-Service Adoption and Risk Report.[Online]Available:https://www.gartner.com/en/newsroom/press-releases, accessed on 20Mar 2021.
[6]   Modi, C., Patel, D., Borisaniya, B., Patel, A., &Rajarajan, M. (2012). A survey on security issues and solutions at different layers of Cloud computing, 1-32.
[7]   Mell, P., &Grance, T. (2011). The NIST definition of cloud computing.
[8]   Dawoud, W., Takouna, I., &Meinel, C. (2010, March). Infrastructure as a service security: Challenges and solutions. In 2010 the 7th International Conference on Informatics and Systems (INFOS) (pp. 1-8). IEEE.
[9]   Ahmed, A., &Zakariae, T. (2018). IaaS cloud model security issues on behalf cloud provider and user security behaviors. *Procedia computer science*, *134*, 328-333.
[10]   Kumar, P. R., Raj, P. H., &Jelciana, P. (2018). Exploring data security issues and solutions in cloud computing. Procedia Computer Science, 125, 691-697.
[11]   Rao, R. G., & Nene, M. J. (2019, March). CNR: A Technique for Data Replication Organization in BigData. In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) (pp. 1-6). IEEE.
[12]   Remesh Babu, K. R., & Samuel, P. (2019). Service-level agreement–aware scheduling and load balancing of tasks in cloud. Software: Practice and Experience, 49(6), 995-1012.
[13]   Sahu, I. K., & Nene, M. J. (2021, February). Identity-Based Integrity Verification (IBIV) Protocol for Cloud Data Storage. In 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) (pp. 1-6). IEEE.
[14]   Lloyd, W., Pallickara, S., David, O., Arabi, M., & Rojas, K. (2014, March). Dynamic scaling for service oriented applications: implications of virtual machine placement on IaaS clouds. In 2014 IEEE International Conference on Cloud Engineering (pp. 271-276). IEEE.
[15]   Amato, F., Moscato, F., Moscato, V., &Colace, F. (2018). Improving security in cloud by formal modeling of IaaS resources. Future Generation Computer Systems, 87, 754-764.

[16] Markatos, E. (2008). Large Scale Attacks on the Internet Lessons learned from the LOBSTER. Crete, Greece.[Online]. Available: http://www. ist-lobster. org/publications/presentations/markatosattacks. pdf.

[17] Seitz, L., Pierson, J. M., &Brunie, L. (2003, October). Key management for encrypted data storage in distributed systems. In Second IEEE International Security in Storage Workshop (pp. 20-20). IEEE.

[18] Sen, J. (2015). Security and privacy issues in cloud computing. In Cloud technology: concepts, methodologies, tools, and applications (pp. 1585-1630). IGI global.

[19] Alliance, C. S. (2020). Top Threats to Cloud Computing. [online] Available:https://cloudsecurityalliance.org/artifacts/top-threats-to-cloud-computing-egregious-eleven; accessed on Mar 16, 2021.

[20] Amara, N., Zhiqui, H., & Ali, A. (2017, October). Cloud computing security threats and attacks with their mitigation techniques. In 2017 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC) (pp. 244-251). IEEE.

[21] MacAfee, Cloud Adoption and Risk Report. [online] Available:https://www.mcafee.com/enterprise/en-us/assets/reports/restricted/rp-cloud-adoption-risk-report-iaas.pdf assessed on 4 Feb 2021.

[22] Singh, S., & Thokchom, S. (2018). Public integrity auditing for shared dynamic cloud data. Procedia Computer Science, 125, 698-708.

[23] Sharma, M. K., & Nene, M. J. (2020). Two-factor authentication using biometric based quantum operations. Security and Privacy, 3(3), e102.

[24] Moutai, F. Z., Hsaini, S., Azzouzi, S., &Charaf, M. E. H. (2019, March). Security Testing Approach for IaaS Infrastructure. In Proceedings of the 2nd International Conference on Networking, Information Systems & Security (pp. 1-5).

[25] Lawal, B. O., Ogude, C., & Abdullah, K. K. A. (2013). Security management of infrastructure as a service in cloud computing.

[26] Cho, Y., Qu, G., & Wu, Y. (2012, May). Insider threats against trust mechanism with watchdog and defending approaches in wireless sensor networks. In 2012 IEEE symposium on security and privacy workshops (pp. 134-141). IEEE.

[27] Devi, M. G., & Nene, M. J. (2018, March). Scarce Attack Datasets and Experimental Dataset Generation. In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1112-1116). IEEE.

[28] Yaseen, Q, Jararweh, Y., Panda, B., &Althebyan, Q. (2017). An insider threat aware access control for cloud relational databases. Cluster Computing, 20(3), 2669-2685.

[29] Shishir, K. C. DATA CENTER SECURITY & VIRTUALIZATION. Diss. Touro College, 2018.

[30] Rao, R. G., & Nene, M. J. (2017, March). SEDoS-7: a proactive mitigation approach against EDoS attacks in cloud computing. In 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET) (pp. 965-970). IEEE.

[31] Shaikh, A. H., &Meshram, B. B. (2021). Security issues in cloud computing. In Intelligent Computing and Networking (pp. 63-77). Springer, Singapore.

[32] Vaezi, M., & Zhang, Y. (2017). Cloud mobile networks (Vol. 5, No. 3, pp. 23-37). Springer.

[33] Dharmakeerthi, T. D. A Study on Cloud Security Concerns and Resolutions (September 2019).

[34] Ghouse, M., Nene, M. J., &Vembuselvi, C. (2019, December). Data Leakage Prevention for Data in Transit using Artificial Intelligence and Encryption Techniques. In 2019 International Conference on Advances in Computing, Communication and Control (ICAC3) (pp. 1-6). IEEE.

Cullum, p. A survey of the host hypervisor security issues presented in public iaas environments and their solutions.

A Study on Image Segmentation Method for Image Processing

S. PRABU[a 1] and J.M. GNANASEKAR[b]

[a] *Associate Professor Department of Computer Science & Engineering, VI Institute of Technology Sirunkundram-603108, Chengleput Distirct, TamilNadu, India*
[b] *Professor, Department of Computer Science & Engineering, SriVenkateswara College of Engineering, Sriperumbadur, TamilNadu, India*

Abstract. Image processing techniques are essential part of the current computer technologies and that it plays vital role in various applications like medical field, object detection, video surveillance system, computer vision etc. The important process of Image processing is Image Segmentation. Image Segmentation is the process of splitting the images into various tiny parts called segments. Image processing makes to simplify the image representation in order to analyze the images. So many algorithms are developed for segmenting images, based on the certain feature of the pixel. In this paper different algorithms of segmentation can be reviewed, analyzed and finally list out the comparison for all the algorithms. This comparison study is useful for increasing accuracy and performance of segmentation methods in various image processing domains.

Keywords. Image Segmentation, Digital Image Processing, K-Means Clustering, Edge detection, histogram.

1. Introduction

Image is binary representation of visual information by means of pixels. It contains lot of information to perform some useful operations on different area of applications like medical field, object detection, video surveillance system, computer vision, pattern recognition, remote sensing etc. With the help of computer algorithms, we can manipulate the pixels either enhance the quality of image or extract the useful information from it. Digital image processing has different stages, in which segmentation is the essential and challenging part in the operation of image processing. Image segmentation should be segregating the images into meaningful parts that are having similar features and behaviors. The purpose of segmentation is, to make the image representation easily, classified meaningfully, and analyzed properly. Using the image segmentation, we can localize the objects and identifying its boundaries in an image. Then assign the labels to each pixel in the entire image. Some pixels having the same labels means they have common characteristics. This paper analyzes the different methods of segmentation and their algorithms. First, we analyze the segmentation carried out by means of pixel properties [1] to analyze the local and global properties, and to tell how speed up the segmentation process by modifying the existing ACM model along with local and global properties. Expectation of maximization algorithm [2] needs more iteration to segment the image. Modification can be based on the

---

[1] S. PRABU, Associate Professor Department of Computer Science & Engineering, VI Institute of Technology Sirunkundram, India
Email: sprabumkm@gmail.com

likelihood properties we can achieve the minimum number of calculations. EM based localization and 3D U-Net and EM based attention mechanism will provide the optimized segmentation results. Color images can be segmented by RGB Histogram [3] along with firefly algorithm. Combination of SCT-I and SCT-V algorithm and SAMFO-TH [17] algorithm solves the multilevel threshold problem of RGB color image. Color hymnographies [6] can be used to identify the color object detection. Image restoration algorithm [4] can segment the image based on the sample blocks. This method utilizes the advantages of both Criminisi algorithm & Watershed image segmentation algorithm. So many factors are affecting while extracting text information from the image [5]. Noise is more, when the image can be captured from live video camera. The Encoder – Decoder framework resolves the issues in segmentation for retrieving text information. To speed up the text retrieval process, Instance Segmentation Network technique can be used. This paper analyze the different methods and compare it to give the clear idea about image segmentation process.

2.Related Work
2.1 Segmentation by means of properties of pixel
Segmentation operation can be also performed by means of global and local properties of pixels [1]. It states that, the global properties are calculated by mean values of various pixels and successive edges of objects, and the local properties are characterized by interaction of successive pixels and the boundary of the image. Image entropy can be done by improved weighting function that can be adapt the weight between local term and global term, so the segmentation speed can be improved significantly.

2.1.1 Thresholding Method
Based on the pixel's intensity, this method partitions the image, into two parts, they are intensity of pixel value lower than threshold and intensity of pixel value greater than threshold. Multi stage thresholding approach is necessary for color image segmentation in many applications [10]. To segment the color components Red, Green and Blue, more than two optimal thresholds can be needed. In future a greater number of thresholds is necessary. But a greater number of thresholds can decrease the performance of segmentation process. The main drawback of multi-level thresholding leads long execution time [11]. During the processing of low-quality images, cause the wrong prediction of threshold values. The improved salp swarm algorithm can minimize the number of thresholds to ensure effective segmentation process. Modified grasshopper optimization algorithm [16], metaheuristic algorithm, firefly algorithm, novel population-based bee foraging algorithm [18] are also used for optimized threshold values.

2.1.2 Edge detection
An edge is a set of connected pixels that creates the boundary between the two successive regions. The edge detection is basically a process of segmenting the image into discontinuity of regions. The perfectness of the image-by-image processing and computer vision can be depending on identifying meaningful edges. Out of all the edge detection techniques, 'Canny Edge Detector' produces good results than other edge detection methods. The basic limitations of edge detection are edge connectivity and edge thickness [7]. Multiple values of Threshold approaches need to solve the above two issues. To find the optimized thresholding values Genetic algorithm can be

applied which optimize the coefficient of filter [8]. This kind of optimization improves the quality of segmenting edges in MR scan images.

2.1.3 Clustering methods

Clustering is the process of image information can be replaced by clusters. Cluster is the collection of similar attribute data points. Pixels, which have common attributes like same color, same texture or any related attribute. Main issues in Clustering are identify the correct inter cluster distance and identify the total number of Clusters. Robust self-sparse fuzzy clustering algorithm improves the clustering result by reducing the noisy features. The main drawback of K means clustering method is image segments are disconnected and disseminated with wide distance. To find the optimal number of clusters is very hard without knowing the initial parameters. The unsupervised K-means clustering technique is needed to achieve this [9]. To solve the initialization problem, UK-means algorithm uses number of points to determine the initial number of clusters. To avoid overlapping, kernel K-means clustering can be proposed [13]. In which kernel functions are transformed into feature space, to estimate the correct number of clusters. The combination of depth and semantic information of images [20] improve the accurate identification of initial center value.

2.2 Optimized Segmentation

Segmentation operation can be improved effectively by applying small changes in the existing segmentation algorithm called Expectation of maximization; it has a greater number of iterations and takes more computing power [2]. Compression based segmentation methods express the approximation of actual pixel values with some of the sample pixels, so that the size of the image can be tremendously reduced. To improve the efficiency of compression, the original image can be pre quantized with higher bits [15]. But sometimes the predictions may not work properly when using high bit rates.

2.3 Color Image Segmentation

Segmentation operation can be achieved in a color images by RGB Histogram [3]. By applying firefly algorithm, the optimal multi-level image segmentation can be achieved. The two new algorithms called SCT-I and SCT-V for image and video input data. For each frame SCT-V algorithm locate the target of interest (TOI) for object tracking and SCT-I algorithm maintain the original color in the target of interest. SAMFO-TH algorithm [17] based on moth flame optimization. Histogram is developed by dividing the range of the data into same sized classes. Then for each class, those data set points which present into the class are calculated. The key parameter of this method is the selection of the threshold value, which can be computed manually or automatically by some algorithms. The basic concept of this method is, to identify the mean or median value, so that pixels of the object are brighter than the background. By applying color hymnographies to increase the color fidelity and improve the color object detection [6].

2.4 Segmentation based on the sample blocks

Segmentation operation can be achieved by improved image restoration algorithm based on the sample blocks [4]. It resolves the image repairing problems occurred when the image restoration process. Criminisi algorithm and watershed image algorithm is applied to the large amount of image set. Then identify the matching pixel

blocks in image segmentation. So that excessive extension of texture blocks for the process of restoration of images can be avoided.

2.4.1 Region-based methods

This kind of segmentation groups the pixels that have similar properties. It segregates the pixel, those having the similar characteristics and dissimilar characteristics. It compares the properties of neighboring pixels and produces the result. The main objective is to differentiate the homogeneity of the image. That can be achieved by improved image restoration algorithm based on the sample blocks [4]. A new method using region of interest for segmenting images, having less computation complexity, to preprocess the training sets so that to reduce the redundant information. For color region segmentation, hue division based selective color transfer algorithm can be used. In which HSV color model transferred into luminance and saturation integrity.

2.5 Text-Based Segmentation

Segmentation operation can be achieved only in a text-based images, to retrieve the necessary text information from the whole image [5]. The information may be single or multiple line, words, or even one or more characters. This paper proposes different methodologies at the various segmentation stages. It first justifies the segmentation process in the text context based on the information retrieval. This paper also discusses different factors affecting the process of segmentation. The Captured images from camera have more background noise [12]. The main issues are symbols having several separate primitives with complex background and distortions from camera. To extract the text information from natural scenes are also complex process. The encoder – decoder framework [17] is proposed. It is used by the combination of attention mechanism and connection time classification. To reduce the more processing time the new technique called Instance Segmentation Network (ISNet) to detect the text content by generating prototype masks simultaneously.

2.6 Motion & Interactive Segmentation

Motion Segmentation means pixels are grouping together in a particular movement of the object. The main objective of this technique is to segment the images for object that are in moving condition. To make the image analysis, and understanding the image properties, to analyze the motion sequence is important. First Identify the background and foreground objects, and then make analysis these objects independently, this model is commonly called 'variation Model' which takes two successive video frames, evaluate the movement ranges between each frame. The main aim of this method is to identify the object boundaries with minimal user interventions.

3.Survey Discussion

Various segmentation algorithms are discussed and each method has its own advantages and disadvantages within a particular context. Some methods need the modification and improvement. The following table 1 compares the various factors of different segmentation methods.

Table 1: Comparison of different segmentation techniques

Method	Description	Characteristics	Advantages	Issues
Thresholding Method	Segmentation done by value of pixel Intensity	Types: Local & Global Thresholding	Simple & Effective Processing.	Single Threshold value not give accurate result. More Threshold values leads more processing time
Edge detection	Segmentation based on discontinuity of regions	Steps Filtering Enhancement Detection	Using probability for finding error rate, Improving signal to noise ratio.	Highly sensitive to noise. More possible for inaccuracy.
Clustering methods	Segmentation by means of collection of similar attribute data points	They are unsupervised algorithms	Applicable for real time problems	1.Identify the correct inter cluster distance 2. Identify the total number of Clusters. Not suitable for wide distances.
K-means Clustering	This method is used when you have unlabeled data (i.e. data without defined any category)	It clusters, he given data into K-clusters or parts based on the K-centroids.	Simple to implement. scalable for large data sets	image segments are disconnected and disseminated with wide distance dependant on initial values.
Region-based methods	Segmentation based on similar properties of pixels	Partitioning an image into homogeneous regions.	Accurately segment the regions that have the same properties we define. It works well with respect to noise	Expensive Computation. Variation of intensity will affect the result.
Motion & Interactive Segmentation	pixels are grouping together in a particular movement of the object	Criminisi algorithm & Watershed algorithm	The pixels that have high intensity variation can be clearly classified as moving objects.	If registration of background pixels are not perfect, leads false prediction of moving objects. More computation cost.
Compression-based methods	approximation of actual pixel values with some of the sample pixels	Types: Lossless, Compression Lossy Compression	Best suited in storage and transmission of images	More computation complexity.
Histogram-based methods	each region's pixels are having similar properties, like intensity, color values, texture.	Best suited for image enhancement	Directly process the color images	Requires high number of operations on each pixel. Computationally slow. Over segmentation occurred since no spatial information.

4. Conclusion

This paper discusses about various image segmentation techniques and different image segmentation approaches are illustrated and correlated. All these methodologies are suitable for many applications in medical field, object detection, computer vision, surveillance system, computer vision etc. From this study, each segmentation methods are desirable for specific image types. So, the combination of multiple segmentation methods is needed to get good result. To improve the accuracy of the segmentation result, need to apply machine learning techniques. The important parameters of image segmentation are accuracy, complexity, efficiency and

interactivity. No particular methods are comfortable for all the types of images. For this reason, segmentation of images in various applications has high demand in future.

References

[1] ZhenZhou Wang, "Image segmentation by combining the global and local properties", Elsevier, Expert Systems With Applications (2017), Vol-87, PP- 30-40.
[2] Lahouaoui Lalaoui, Tayeb. Mohamadi and Abdelhak Djaalab, "New Method for Image Segmentation", Elsevier, Procedia - Social and Behavioral Sciences (2015), Vol-195, PP- 1971–1980.
[3] V. Rajinikanth, and M. S. Couceiro, "RGB Histogram based Color Image Segmentation Using Firefly Algorithm", Elsevier, Procedia Computer Science(2015), Vol-46, PP- 1449–1457.
[4] Huang Ying, Li Kai, and Yang Ming, "An Improved Image Inpainting Algorithm based on Image Segmentation", Elsevier, Procedia Computer Science(2017), Vol-107, PP- 796–801.
[5] Gupta Mehul, Patel Ankita, Dave Namrata, Goradia Rahul and Saurin Sheth, "Text-Based Image Segmentation Methodology", Elsevier, Procedia Computer Science (2014), Vol-14, PP- 465–472.
[6] Gupta Mehul, Patel Ankita, Dave Namrata, Goradia Rahul and Saurin Sheth, "Text-Based Image Segmentation Methodology", Elsevier, Procedia Computer Science (2014), Vol-14, PP- 465–472.
[7] Mamta Mittal, Amit Verma, Iqbaldeep Kaur, Bhavneet Kaur, Meenakshi Sharma, Lalit Mohan Goyal, Sudipta Roy and Tai-Hoon Kim, "An Efficient Edge Detection Approach to Provide Better Edge Connectivity for Image Analysis", IEEE Access (2019), Vol-7, PP- 33240–33255.
[8] Ahmed H. Abdel-Gawad, Lobna A. Said, Dave Namrata, and Ahmed G. Radwan, "Optimized Edge Detection Technique for Brain Tumor Detection in MR Images", IEEE Access (2020), Vol-8, PP- 136243–136259.
[9] Kristina P. Sinaga and Miin-Shen Yang, "Unsupervised K-Means Clustering Algorithm", IEEE Access (2020), Vol-8, PP- 80716–80727.
[10] Heming Jia Jun Ma and Wenlong Song, "Multilevel Thresholding Segmentation for Color Image Using Modified Moth-Flame Optimization", IEEE Access (2019), Vol-7, PP- 44097–44134.
[11] Heming Jia Jun Ma and Wenlong Song, "Multilevel Thresholding Segmentation for Color Image Using Modified Moth-Flame Optimization", IEEE Access (2019), Vol-7, PP- 44097–44134.
[12] Ahmed A. Ewees, Mohamed Abd Elaziz, Mohammed A. A. Al-Qaness, Hassan A. Khalil and Sunghwan Kim," Two-Step CNN Framework for Text Line Recognition in Camera-Captured Images", IEEE Access (2020), Vol-8, PP- 26304–26315.
[13] Soosan Beheshti, Edward Nidoy, and Faizan Rahman, "K-MACE and Kernel K-MACE Clustering",IEEE Access (2020), Vol-8, PP- 17390–17403.
[14] Dan Wang, Guoqing Hu, Qianbo Liu, Chengzhi Lyu, and Md Mojahidul Islam, "Region-Based Nonparametric Model for Interactive Image Segmentation", IEEE Access (2019), Vol-7, PP- 111124–111134.
[15] Jianyu Lin,"A New Perspective on Improving the Lossless Compression Efficiency for Initially Acquired Images", IEEE Access (2019), Vol-7, PP- 144895–144906.
[16] Hongnan Liang, Heming Jia, Zhikai Xing, Jun Ma, And Xiaoxu Peng, "Modified Grasshopper Algorithm-Based Multilevel Thresholding for Color Image Segmentation", IEEE Access (2019), Vol-7, PP- 11258–11295.
[17] Heming Jia, Jun Ma, And Wenlong Song, "Multilevel Thresholding Segmentation for Color Image Using Modified Moth-Flame Optimization", IEEE Access (2019), Vol-7, PP- 44097–44134.
[18] Zhicheng Zhang And Jianqin Yin, "Bee Foraging Algorithm Based Multi-Level Thresholding For Image Segmentation", IEEE Access (2020), Vol-8, PP- 16269–16280.
[19] Zhen Zheng , Bingting Zha, Hailu Yuan, Youshi Xuchen, Yanliang Gao, And He Zhang, "Adaptive Edge Detection Algorithm Based on Improved Grey Prediction Model", IEEE Access (2020), Vol-8, PP- 102165–102176.
[20] Hanxiao Rong, Alex Ramirez-Serrano, Lianwu Guan, And Yanbin Gao, "Image Object Extraction Based on Semantic Detection and Improved K-Means Algorithm", IEEE Access (2020), Vol-8, PP- 171129–171139.

Shielding Software Defined Network Using Hidden Authentication Technique

B. Srivani [a 1], S. Renu Deepti [a], and Qualid Unnisa [a]

[a]*Department of Information Technology, VNR VJIET, Hyderabad, India*

Abstract. Software defined networking (SDN) permits community feature program ability intended to facilitate about design along with renovation, as well as permit community directors toward adapt congestion guidelines. Nevertheless, denial of provider (DoS) assailants causes productivity issues upon centralized consolidate aircraft about SDN. Even through shipping layer safety (TLS) be able to assist comfy manage plane, that far analytically extensive and composite design. Within the document, we plan light-weight validate compound, known as Hidden Authentication (HiAuth), toward guard the SDN through battering specifications about redirecting devices to control packets thru effective bitwise functioning. HiAuth be that initially toward incorporate records battering methods for Open Flow toward offer safety in opposition to DoS attacks. HiAuth utilizes IP identification field about IPv4 as well as proceedings recognition area about OpenFlow within two attestation methods. The investigational outcomes display that HiAuth able to efficiently alleviate trespasser DoS assaults as well as supply excessive unnoticeable toward assailants.

Keywords. Software defined networking (SDN), OpenFlow, Denial of service (DoS) attacks, Information hiding

1.   Introduction

Software defined networking (SDN) disassociate the system manipulate through redirecting gadgets as a result simplifies and complements community maintenance [1]. Within fashionable, SDN shape contains three layers: software plane, manage plane, and data aircraft. Structure and the attachments a few of the three layers abouts SDN. The northbound utility programming interfaces (APIs) permit SDN packages toward grant plant as well as protection guidelines to command aircraft. Later, SDN manipulate aircraft enforce one's regulations at forward gadgets through the southbound APIs. OpenFlow is the first SDN popular which defines open southbound interfaces for controlling network flows. The SDN centralization of community intelligence with the supply about worldwide view of the whole community improved programmability and scalability for destiny network and service manage. As an end result, many new SDN packages are proposed to optimize the network performance from incredible factors, which includes throughput maximization [2][3], deterministic delay functionality to influence the network behavior via software program from a logically centralized manage brings numerous benefits.

However, software vulnerabilities grow to be a project [7]. More critically, the centralization of the control aircraft can motive an unmarried aspect of nonsuccess.

---

[1] B. Srivani,Department of Information Technology, VNR VJIET, Hyderabad
Email: srivaanib@gmail.com.

Regrettably, transport layer protection (TLS) be optionally available with OpenFlow protocol.

Because of its configuration complexity, TLS isn't adopted by means of way of a few OpenFlow-enabled switches and controllers [8]. As a stop result, the legitimacy of the forwarding gadgets can't be verified. Hence, malicious assults be able to without issues release denial of provider (DoS) attacks closer to SDN controller. Toward solve the DoS problem, we generalize danger worries about DoS attacks towards the SDN controller. In unique, we observe feasible DoS assault eventualities, which include the outsider and the insider attack lessons. Then, we compare DoS affects at the SDN controller. Eventually, designed a light-weight attestation scheme, Hidden Authentication (HiAuth), toward reduce attatrespasser DoS assults. HiAuth have essential capabilities: 1) difficult to understand, 2) light-weight. Initially, intended to obscurity, HiAuth mimics the original statistical distributions of the values cause through the going for walks frameworks towad say unpredicted through attackers. Then having mild-weight, HiAuth simplest is based totally on easy bitwise operations for computations.
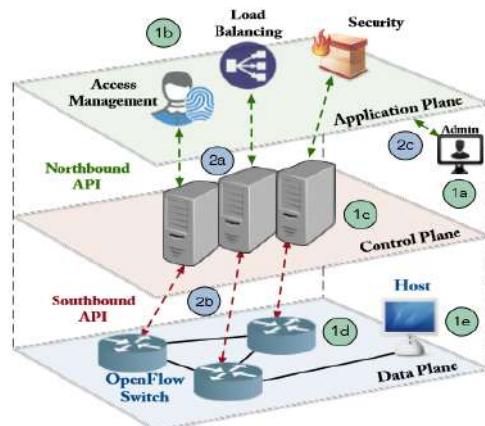
Figure 1. SDN architecture and its security challenges hiding

Thus, it does no longer require specialized hardware. Toward excellent about information, HiAuth be initiate toward consist of records battering models for OpenFlow rule toward gives safety in case of DoS assults on rule degree. There, recommend two HiAuth schemes: IP identification (IPID) based totally definitely HiAuth hids recognition about redirecting things inside IPID province about the IPv4 header. Transaction identification (XID) base absolutely HiAuth what hide the recognition about the redirecting things of XID area about OpenFlow header. IPID-primarily base completely HiAuth presents attestation on the community layer. Nevertheless, within IPv4 turns into out of date in the furtherer else total change toward IPv6 takes place, XIDbased HiAuth may use toward equal stage of attestation.

2.   Research Methodology

SDN has two major blessings: community application potential and centralized community control. Initially, through seperate about manage plane as well as statistics

plane, program of SDN lets in system rules toward changed with the aid about software program rather the guide composition correlated toward conventional systems. Toward renovate community model in conventional systems, every device must be manually configured, which may additionally reason as a result toward safety susceptibility. Next, the consolidated manipulate common sense able to ease system maintenance due to supply about system worldwide view. Protectivity within SDN have aspect, particularly security via SDN. On the only hand, safety thru SDN makes a specialty of utilizing SDN features to solve conventional community protection problems. SDN protectivity packages able to look at packets via manipulate aircraft. Then protectivity analysis, those application be able to drop or else redirected traffic toward safety center boxes. Upon alternative, protection about SDN offers through safety challenges resulting from SDN, e.G., factor about failure. To similarly difficult the differences among these elements, we took DoS elimination for instance. With safety over SDN, DoS assault upon classical systems may prevent through SDN applications.

Defense4All plays two important responsibilities, particularly (a) conduct observing through using studying traffic records, (b) traffic redirected through the server underneath DoS assault inside community can be covered through such SDN programs. Within case safety about SDN, DoS attack in opposition to manipulate plane may be prevented. TLS charge methods be the solution encouraged by using the current SDN requirements. The planned HiAuth procedure belongs toward the latter issue wherein targets toward a lightweight packet stage attestation toward save you DoS assaults towards manipulate aircraft.
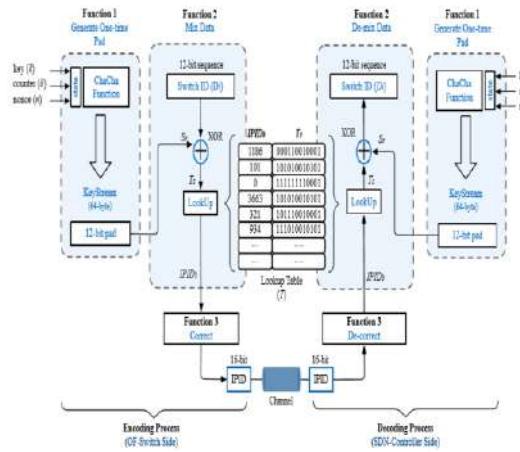
Figure 2. IPID-based HiAuth

HiAuth be packet-degree attestation procedure focus toward mitigate elegance about external DoS assaults within opposition to manage aircraft. The most effective required a bitwise operation and a easy mapping characteristic to cover identification facts about gadget for control packet header. Upon opposite hand, HiAuth design be instead honest and easy at the same time as compare toward TLS layout. Within phase, we introduced two HiAuth methods specifically, IPID-primarily base HiAuth along with XIDbased HiAuth.

IPID-base totally HiAuth be plan toward used for inside systems so as to be put into effect chance of IPID generation. In spite of paintings posted inside the writing upon hide records savvy IPID, it far cleans so to be more about these strategies forget apparent adjustment what to be brought the issuing about IPID. Moreover, the total techniques be useful best whilst community MTU be set on. Consequently, suggest IPID-based HiAuth toward offer data attestation as well as triumph over the deficiency. The IPID-primarily base HiAuth encoding along with deciphering methods be regularity along contains 3 consequences: (1) one-time pad era, (2) information mining, (3) distribution corrective model. HiAuth encoding technique calls for subsequent inputs: a) a 256-bit key, a sixty-four-bit block counter, and a 64bit for use in the one-time pad generator, b) a 12-bit tool recognition for use within the information blending characteristic. The output of those abilities be the mapped right into 12-bit IPID base cost. Eventually, a distribution corrective model changes the IPID base cost right to issue compliant sixteen-bit IPID if you want to be used for the manage packet transmission.

3.   Results and Analysis

In this section, we demonstrate the results of our computations, which gain the security by hiding its identity of the network without leaving the traces behind as per the proposed technique, hence it mitigates the intruder. Also, it describes and demonstrates the computational analysis precisely with minimal efforts and good security comparatively.
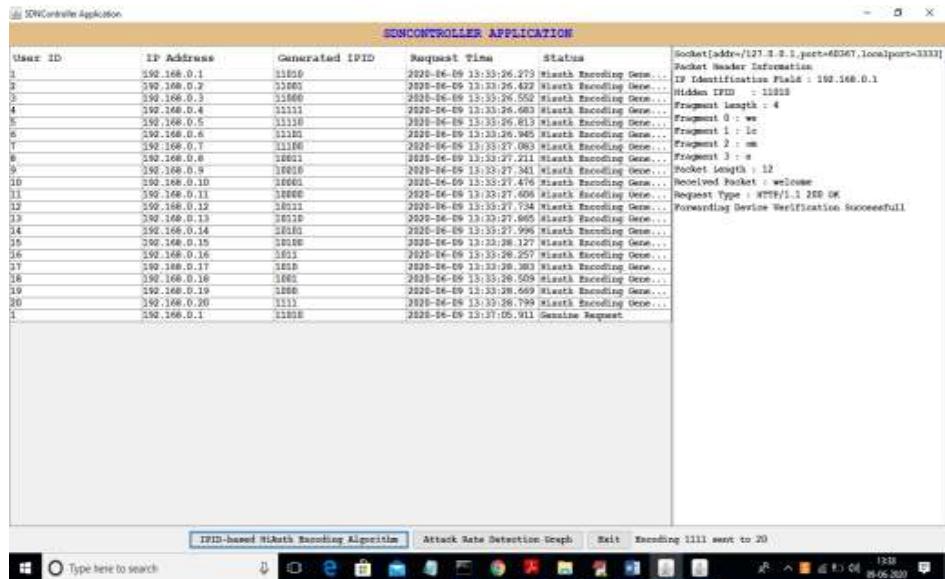
Figure 3. Hiding the IPID of the request

In Figure 3 we can see that the received request in text area has all the details of its header, wrapped with the hidden authentication mechanism with all the request details

in text area we can see number of fragments received and data in each fragment and hidden IPID and then verification message.
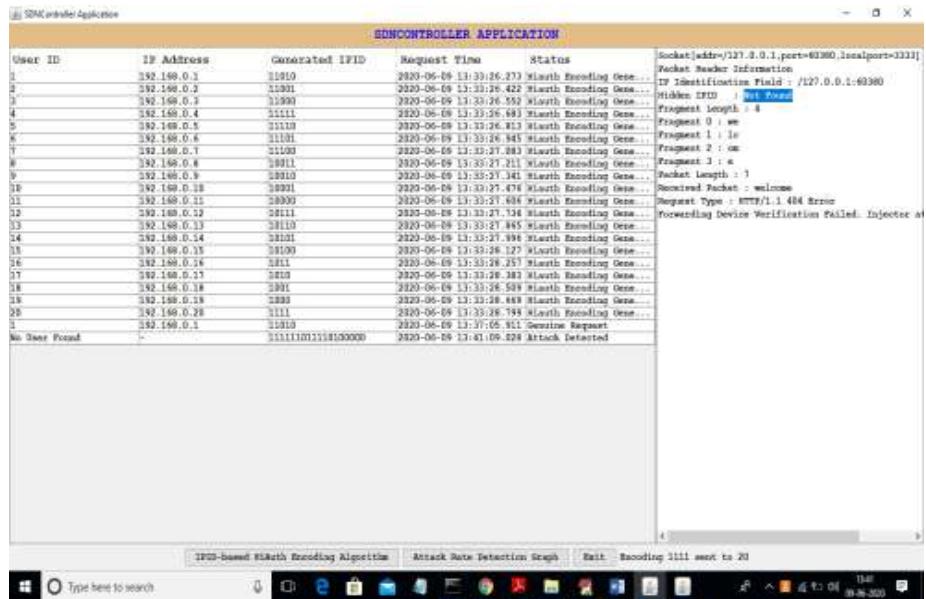
Figure 4. Rejecting the unsecure request

In Figure   4 the controller did not find the ID to be secured in the text area and hence rejected the request as the verification failed.
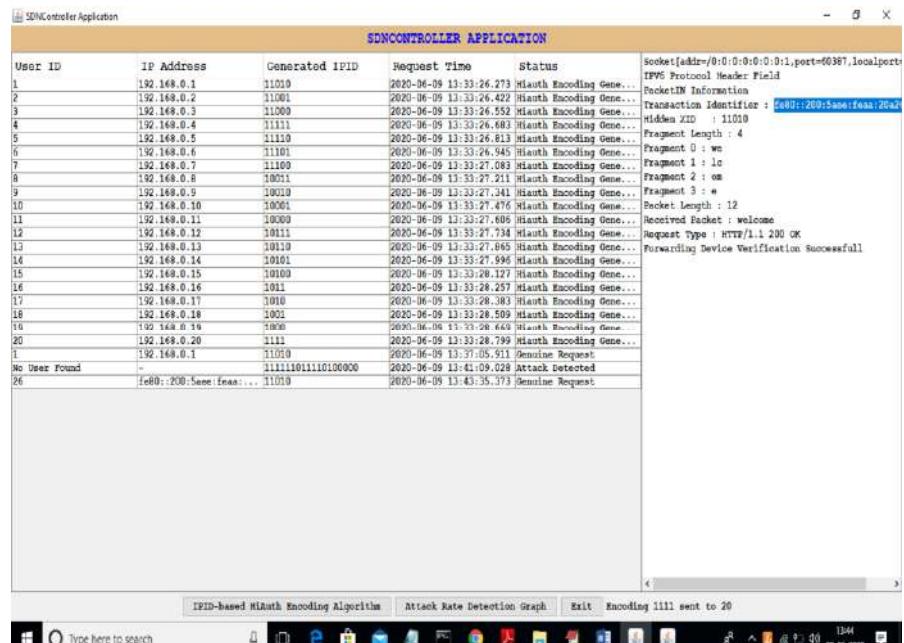
Figure 5. Approving the authenticated request

In Figure 5 the text area this time we can see at selected text that this request process as IPV6 address and if SDN found correct XID then device request will be authenticated. If request not authenticated then user not found in the look table or not authenticated with XID process and display information as Attack detected.

4.   Conclusion

In this research paper, we introduce packet authentications concept called IPID and XID. IPID used IPV4 protocol to hide device id in packets along with data and then send packet to SDN controller where IPID will get authenticated by extracting details from packet. Device ID will be encrypted using CHA-CHA algorithm by generating random number and then perform XOR operation between CHA-CHA random number and device id to get secure device ID with the help of DATA MIXING. Secure device id will get exchange between genuine device and SDN controller and get authenticated for each request. After assigning secure ID SDN controller will used lookup table to check whether forwarding device id exists or not. If exists device will be considered as genuine else malicious. In future if protocol changed from IPV4 to IPV6 then IPID packet fragmentation will not work so author using XID transaction-based header to hide device id. XID will hide device details in transaction header and to ease computation look up table will not be used and XID will dynamically compute device ID for authentication upon each request.

References

[1]   M. Huang, W. Liang, Z. Xu, and S. Guo, "Efficient algorithms for throughput maximization in software-defined networks with consolidated middleboxes," IEEE Transactions on Network and Service Management, vol. 14, no. 3, pp. 631–645, 2017.

[2]   Y.-J. Chen, L.-C. Wang, F.-Y. Lin, and B.-S. P. Lin, "Deterministic quality of service guarantee for dynamic service chaining in software defined networking," IEEE Transactions on Network and Service Management, vol. 14, no. 4, pp. 991–1002, 2017.

[3]   H. Hamed and E. Al-Shaer, "Taxonomy of conflicts in network security policies," IEEE Communications Magazine, vol. 44, no. 3, pp. 134–141, 2006.

[4]   S. J. Vaughan-Nichols, "OpenFlow: The next generation of the network?," IEEE Computer, vol. 44, no. 8, pp. 13–15, 2011.

[5]   I. Ahmad, S. Namal, M. Ylianttila, and A. Gurtov, "Security in software defined networks: A survey," IEEE Communications Surveys & Tutorials, vol. 17, no. 4, pp. 2317–2346, 2015.Yin B., Shen W., Cheng Y., Cai L.X. and Li Q, "Distributed resource sharing in fog-assisted big data streaming", IEEE international conference on communications (ICC), pp. 1-6, May 2017.

[6]   W. Mazurczyk and L. Caviglione, "Steganography in modern smartphones and mitigation techniques," IEEE Communications Surveys & Tutorials, vol. 17, no. 1, pp. 334–357, 2015.

[7]   J.-M. Guo, G.-H. Lai, K. Wong, and L.-C. Chang, "Progressive halftone watermarking using multi layer table lookup strategy," IEEE Transactions on Image Processing, vol. 24, no. 7, pp. 2009–2024, 2015

[8]   B. Jankowski, W. Mazurczyk, and K. Szczypiorski, "Padsteg: Introducing inter-protocol steganography," Telecommunication Systems, vol. 52, no. 2, pp. 1101–1111, 2013.

431

A Study on Speech Emotion Recognitions on Machine Learning Algorithms

Shanmuk Srinivas Amiripalli [a] [1], Potnuru Likhitha [a], Sisankita Patnaik [a], Suresh Babu K [a], Rampay Venkatarao [b]

[a] *Dept of CSE, GIT, GITAM University,Visakhapatnam,Andhrapradesh,India*
[b] *Dept of CSE, Wollega University, Nekemte, Eithopia*

Abstract. Speech emotion detection has been extremely relevant in today's digital culture in recent years. RAVDESS, TESS, and SAVEE Datasets were used to train the model in our project. To determine the precision of each algorithm with each dataset, we looked at ten separate Machine Learning Algorithms. Following that, we cleaned the datasets by using the mask feature to eliminate unnecessary background noise, and then we applied all 10 algorithms to this clean speech dataset to improve accuracy. Then we look at the accuracies of all ten algorithms and see which one is the greatest. Finally, by using the algorithm, we could calculate the number of sound files correlated with each of the emotions described in those datasets.

Keywords. Scikit-learn, MLPClassifier, Logistic Regression, Naïve Bayes, XGB, LightGBM, Stochastic Gradient Descent, Support Vector Machine.

1. Introduction

In classification, a set of data is categorized into classes, and it can be performed not only on structured data but also on unstructured data. Each data point of datasets is predicted into which class it falls under. These classes can be referred as targets, labels or categories. The task of the Classification predictive model is to approximate the mapping function from input variables to discrete output variables. Example for Binary Classification: While flipping a coin, the chances of getting head or tail can be categorized into two classes. Example for Multiclass Classification: There may have a 3-class classification problem of an animal set to classify as lion, tiger or leopard with a total of 100 instances. The classifier, in this case, needs training data to understand how the given input variables are related to the class. Once the classifier is trained accurately, it can detect the output of the particular testing data. This concept was used to get the accuracy of the dataset. Speech is one of the most natural means for us to communicate ourselves as humans. We depend on it so often that we can deduce its significance from other forms of contact, such as emails and instant messaging, in which we often use sentiment to convey the messages' contents. Since feelings are so central in conversation, sensing and interpreting them is crucial in today's world of remote communication.

---

[1] Shanmuk Srinivas Amiripalli, Dept of CSE, GIT, GITAM University,Visakhapatnam, India

Email: shanmuk39@gmail.com

It is tough to detect feelings because they are emotional. There is no agreed-upon procedure for quantifying or categorizing them. An SER paradigm is a collection of methodologies for interpreting and categorizing speech signals to detect emotions. This type of interface can be used for some items, including interactive voice-based assistants and caller-agent contact research. We examine audio files' acoustic properties in this study to detect real feelings in recorded expressions. The method of recognizing human emotion through expression and affective states is known as Speech Emotion Recognition or SER. This takes advantage of the idea that voice always reflects natural feeling by tone and pitch. It involves defining people's feelings dependent on the tone of their voice in their expressions. People cannot all talk at the same volume. It varies based on their moods and circumstances. This is the same phenomenon that allows creatures like dogs and horses to understand human feelings. It's popular in call centres. If you note, call centre workers do not communicate to each customer in the same way; their method of speaking to them differs based on the customer. Speech emotion recognition systems enable workers to understand consumer emotions through speech. In order for them to strengthen and deliver services to their customers.

2. Literature Survey

In [1-3] compares the speech emotion classification accuracy of speaker-based and the time to construct the model between Support Vector Machine(SVM) and Multi-Layer Perceptron(MLP) classifiers. The classification was performed with the WEKA unit, and the features were extracted with PRAAT. A basic SER module structure was chosen to compare the described classifiers. Confusion matrix, classification precision, and construct time are used to test supervised learning algorithms' efficiency. Despite the fact that MLP outperforms SVM in total emotion classification, SVM's preparation was quicker.MLP and SVM had acceptance rates of 78.69 and 76.82, respectively. In MLP, the highest emotion identification was for depression (89%), with pleasure and anxiety being the most perplexing emotions, while in SVM, the highest emotion recognition was for indignation (87.4%), with disgust and fear being the most perplexing emotions. In research [4-7], wavelet packet techniques were used to recognize speech sentiment. The wavelet packet coefficients were examined at five decomposition stages, analyzed, and used as inputs to Support Vector Machine (SVM) classifiers. The findings showed that using these features on seven emotional states in two languages, German and Chinese, this wavelet packet strategy increased efficiency by 4.5 percent and 16.9 percent, respectively, as opposed to a single one without these features. These two datasets have a final success average of 61.9 percent and 62.2 percent, respectively. As a result, it was concluded that wavelet packet coefficient features outperform Mel-Frequency Cepstral Coefficient (MFCC) features. The ANN-based decision fusion for SER was introduced in [7-9]. SVM, k-NN, Gaussian Mixture Model, HMM, ANN, and other sequence classification methods were used to determine which was the most efficient tool for classifying speech emotions. SVM was said to have the highest results out of all of them. Some suspected that the ANN could achieve ideal results, but they didn't know which kind was best for SER. To identify various feelings, four separate ANNs were used: Probabilistic Neural Network (PNN), Radial Base Function (RBF) network, Back Propagation (BP) network, and Elman Network. At the judgment stage, voting systems were used to fuse the recognitions utilizing Statistical and Spectral characteristics. Principal Factor Analysis decreases the

dimensionality of super vectors built from spectral features (PCA). However, since PCA was used as a dimensionality reduction method rather than a pattern recognition method, it caused some issues. Proposed decision fusion was discussed as a way to escape them. The proposed decision fusion was successful, and the dimensionality reduction was probable, according to the results. In research [10-12] proposes a speaker-independent approach for categorizing emotional vocal sounds. The treatment divided the mechanism of recognizing emotions into two sections. The first phase entails a coarse encoding and grouping of six emotional states to determine which pair of emotions has the highest likelihood. Low-level encoding procedures were suggested at this time, and the extracted features were combined to produce the best emotional state descriptive acoustic vectors. Second, utilizing the Sequential Floating Forward Selection (SFFS) algorithm, modern encoding strategies were used to define a special collection of acoustic features for each pair of emotions that can be used to discriminate between them. There are a total of 72 high-level acoustic features.

3. Proposed Classification Algorithm

The classification is a supervised learning principle of Machine Learning that separates a dataset into groups. Speech Expression Recognition, Face Identification, Handwriting Recognition, Text Classification, and other classification problems are some of the most important. It may also be a multiclass problem or a binary classification problem. In Machine Learning, there are many classification algorithms. On the RAVDESS, TESS, and SAVEE datasets, the following algorithms were used. First of all, we give some audio datasets as input. Extracted features from those speech files. Declared one dictionary for emotions in the dataset and another dictionary for emotions that we want to observe. Loaded the dataset and split into two subsets i.e. 75% of data for training and 25% of the data for testing. Initialized a classifier and trained the model using the dataset to predict the emotions of each of the speech files. Finally, it gives the emotion as output. We took 3 different datasets, namely RAVDESS, SAVEE and TESS, which consist of different emotions. We found the accuracies of the ten different classifiers Logistic Regression, Naïve Baye's, Stochastic Gradient Descent, KNN, Decision Tree, Random Forest, Support Vector Machine, MLPC, XG Boost and Light GBM, for each dataset and compared them to know which classifier have more accuracy [13-16]. The accuracy was calculated before and after masking of the datasets. Depending on the accuracy, we got to know that MLP Classifier has more accuracy compared to others. Before masking, the accuracy for RAVDESS, SAVEE, and TESS in MLPC were 70%, 100% and 80%, respectively and after masking, the accuracy for RAVDESS, SAVEE, and TESS in MLPC were 75.60%, 100% and 84%, respectively [17-21].
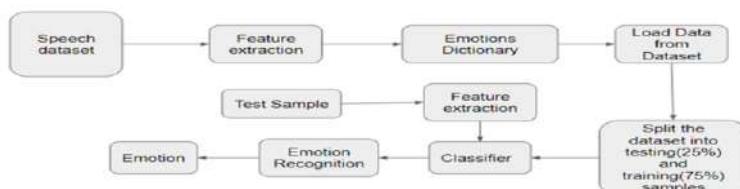
Figure 1. Block Diagram of SER
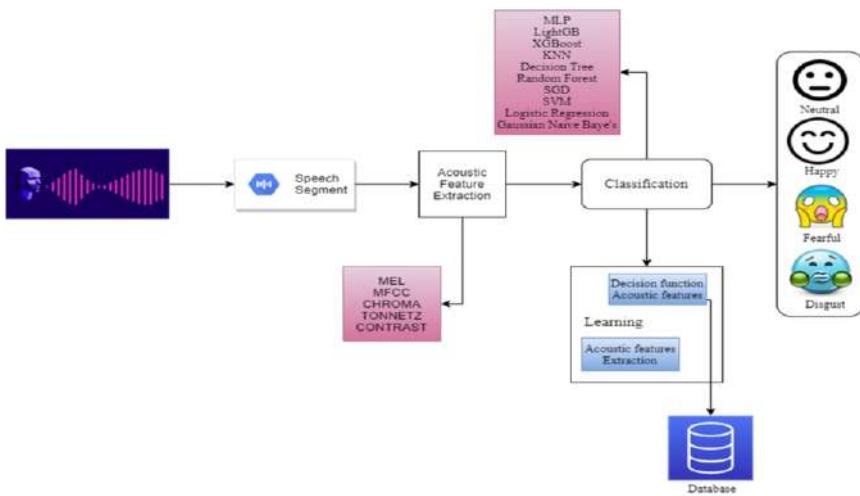
Figure 2. Architecture of SER

Algorithm
Step1: *Start*
Step2: *Imported all required packages, libraries and modules.*
Step3: *Considering five standard features of any audio file and declared a dictionary that contains the emotions in the dataset and a list with the emotions observed.*
Step 4: *Input datasets are considered in this research is RAVDESS, TESS, SAVEE.*
Step 5: *Configuring the experiment with considered 75% of data for training and 25% of data for testing.*
Step 6: *Initialized a classifier model and fit the model.*
Step 7: *Accuracy for ten different classifiers and picked one classifier with more accuracy.*
Step 8: *Using that classifier, we found a count of files of each of the emotions in each dataset.*
Step 9: *Classification reports and confusion matrix for each of them are drawn.*
Step10: *Stop*

4. Results and Discussion

After considering four emotions only as the observed emotions increases, there may be a chance to decrease accuracy, so we did not consider all emotions. RAVDESS contains eight emotions, whereas TESS and SAVEE contain seven emotions, so we took four common emotions from these three datasets. We took emotions happiness, fear, disgust and neutral in our observation. The below are the results for the count of files of each emotion we considered in each dataset.

(a)   (b)   (c)
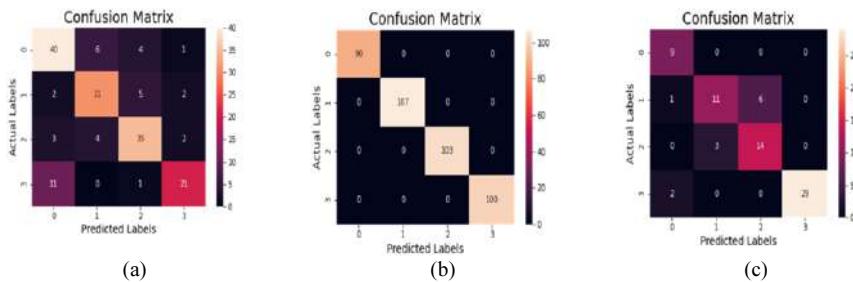
Figure 3. Confusion Matrix of MLP Classifier using (a) RAVDESS Dataset (b) TESS Dataset (c) SAVEE Datasets

Table 1. Comparison of Modulation schemes

Emotion	Number of files in RAVDESS	Number of files in TESS	Number of files in SAVEE
HAPPY	45	103	20
FEARFUL	41	107	14
DISGUST	56	90	12
NEUTRAL	26	100	29
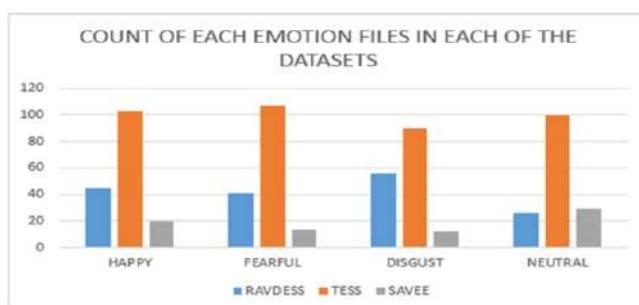
Figure 4. Count of files of each emotion in each dataset

5. Conclusion

In the above research, the major observation is that MLP Classifier is the best classifier compared to any other classifier in Machine Learning. However, to improve this model's accuracy, we need to clean the noise in our dataset. We can improve the efficiency of the model from 71% to 76% for RAVDESS, accuracy for TESS before, and masking remained constant, i.e. 100%, 80% to 84% for SAVEE by considering four emotions such as neutral, happy, fearful and disgust. We can use the model to predict the emotions of the audio files in the datasets and any other sample audio files with the extension .wav.

References

[1]. Idris, I., & Salam, M. S. H. (2014, December). Emotion detection with hybrid voice quality and prosodic features using neural network. In 2014 4th World Congress on Information and Communication Technologies (WICT 2014) (pp. 205-210). IEEE.

[2]. Wang, K., An, N., & Li, L. (2014, September). Speech emotion recognition based on wavelet packet coefficient model. In The 9th International Symposium on Chinese Spoken Language Processing (pp. 478-482). IEEE.

[3]. Xu, L., Xu, M., & Yang, D. (2009). ANN based decision fusion for speech emotion recognition. In Tenth Annual Conference of the International Speech Communication Association.

[4]. Atassi, H., & Esposito, A. (2008, November). A speaker independent approach to the classification of emotional vocal expressions. In 2008 20th IEEE international conference on tools with artificial intelligence (Vol. 2, pp. 147-152). IEEE.

[5]. Huang, C., Jin, Y., Zhao, Y., Yu, Y., & Zhao, L. (2009, September). Speech emotion recognition based on re-composition of two-class classifiers. In 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (pp. 1-3). IEEE.

[6]. Zhang, J., Yin, Z., Chen, P., & Nichele, S. (2020). Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. Information Fusion, 59, 103-126.

[7]. Kerkeni, L., Serrestou, Y., Mbarki, M., Raoof, K., & Mahjoub, M. A. (2018, January). Speech Emotion Recognition: Methods and Cases Study. In ICAART (2) (pp. 175-182).

[8]. Basu, S., Chakraborty, J., Bag, A., & Aftabuddin, M. (2017, March). A review on emotion recognition using speech. In 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT) (pp. 109-114). IEEE.

[9]. Tarunika, K., Pradeeba, R. B., & Aruna, P. (2018, July). Applying machine learning techniques for speech emotion recognition. In 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.

[10]. Kerkeni, L., Serrestou, Y., Mbarki, M., Raoof, K., Mahjoub, M. A., & Cleder, C. (2019). Automatic speech emotion recognition using machine learning. In Social media and machine learning. IntechOpen.

[11]. Amiripalli, S. S., Bobba, V., & Potharaju, S. P.: A novel trimet graph optimization (TGO) topology for wireless networks, (2019) doi:10.1007/978-981-13-0617-4_8.

[12]. Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.

[13]. Amiripalli, S. S., & Bobba, V.: Research on network design and analysis of TGO topology. International Journal of Networking and Virtual Organisations, 19(1), 72-86, (2018).

[14]. Potharaju, S. P., & Sreedevi, M. (2017). A Novel Clustering Based Candidate Feature Selection Framework Using Correlation Coefficient for Improving Classification Performance. Journal of Engineering Science & Technology Review, 10(6).

[15]. Amiripalli, S. S., & Bobba, V. (2019). An Optimal TGO Topology Method for a Scalable and Survivable Network in IOT Communication Technology. Wireless Personal Communications, 107(2), 1019-1040.z

[16]. Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.

[17]. Amiripalli, S. S, V. Bobba, "A Fibonacci based TGO methodology for survivability in ZigBee topologies". INTERNATIONAL JOURNAL OF SCIENTIFIC &TECHNOLOGY RESEARCH, 9(2), pp. 878-881. (2020).

[18]. Ramiah Chowdary, P., Challa, Y., Jitendra, M.S.N.V.: "Identification of MITM Attack by Utilizing Artificial Intelligence Mechanism in Cloud Environments" Journal of Physics: Conference Series, 1228 (1),012044, (2019).

[19]. Thota, J.R., Kothuru, M., Shanmuk Srinivas, A., Jitendra M, S.N.V.: Monitoring diabetes occurrence probability using classification technique with a UI, International Journal of Scientific and Technology Research, 9 (4), pp. 38-41, (2020).

[20]. Potharaju, S. P., Sreedevi, M., Ande, V. K., & Tirandasu, R. K. (2019). Data mining approach for accelerating the classification accuracy of cardiotocography. Clinical Epidemiology and Global Health, 7(2), 160-164.

[21]. Jitendra, M.S.N.V., Radhika, Y.: A review: Music feature extraction from an audio signal, International Journal of Advanced Trends in Computer Science and Engineering, 9 (2), pp. 973-980, (2020).

Content Filtering of Social Media Sites Using Machine Learning Techniques

U. S. Tambe [a 1], N. R. Kakad [a], S. J. Suryawanshi [a] and S. S. Bhamre [a]

[a]*Dept of IT, MVPS's KBT College of Engineering, Nashik, Maharashtra, India*

Abstract. To build a social network or social relations between people, we use social networking platforms like Facebook, Twitter, apps, etc. Using this media, users can share their views and opinions about a particular thing. Many people use their media for personal interests, entertainment, the market stocks, or business purposes. Nowadays, user security is the major concern for social networking sites. Online social networks give a little bit of support regarding content filtering. In this article, we proposed a system that provides security regarding malicious content that is posted on their social networking sites. To filter the content that might be unwanted messages, labeled images, or vulgar images, we proposed three level architecture. The user can use the auto-blocking facility as well.

1. Introduction

In today's life, social media plays a very important role. People spend most of their time on social media for chat and sharing their ideas. Using these media, people can share their information or exchange several types of content such as images, videos, textual messages or audio messages. Many people comment on this shared content. The peoples receive feedback on whatever material he or she has shared on the wall. Sometimes it gives a positive response, a negative response, or suggestions which are very useful for improvement. According to Face book, users make up 90 bits of content each month, though; more than 30 billion bits of content (web links, news stories, blog posts, notes, photo albums, etc.) are shared each month [1].

Users can post any type of content on social media. Unwanted text messages, branded photographs, obscene, pornographic images, personal mischievous comments, and so on are examples. Other users can see these posts and comment on them. The user's social image can be harmed as a result of this message. As a result, the protection of such a user's wall is critical. Up to a point, Face book provides protection. Only a selected group of people on Face book have access to other people's walls, such as friends, friends of friends, or established groups of friends. The user has the option of locking his profile picture. However, because content-based filtering is not supported, such unwanted messages cannot be prevented. The goal of present system is to protect user walls by filtering out unwanted content and user social

---

[1] U. S. Tambe, Assistant Professor, Department of IT, MVPS's KBT College of Engineering, Nashik
Email: tambe.ujwala@kbtcoe.org.

media images. Users can modify the filtering rules as per his/her choice. The user has control over who can send messages to their wall. For text filtering, a short text classification technique is used [2]. For training, a vocabulary list is used. The vocabulary list contains words and its class. For labeled image filtering, we used the OCR algorithm. OCR can extract the text from an image and save that text into a text file. Then, we match the extracted data with the dataset. If a match is found, then that image will be filtered by the system. For image filtering, we used the skin sheriff technique, which consists of skin detection and pornography detection algorithm. The skin detection algorithm can extract all the skin areas from the image and label the pixel that classified as skin as grey and non-skin as white. A pornography detection algorithm is used to calculate the maximum area of non-skin pixel. If it exceeds the maximum value, the image will be vulgar.

2.   Literature Survey

2. 1 Content-based Filtering:

 In a Content-based filtering system, a document is approved by comparing the document profile with the user profile using information retrieval techniques like Term Frequency and Inverse Document Frequency (TF-IDF) [3]. As per the user's prior feedback and choices, user characteristics were gathered and profiled. The system needs an item-to-item relationship which helps to advise the document to the user. The system begins by assembling content information about the item, such as behaviors, indications, and so on, for the disease-related item. The system will then put the user to the test in order to rate the item. The system then compares the unrated item to the user profile item and assigns a score to the unrated item, resulting in the user being presented with items that are graded consistent with the score assigned

2. 2 Collaborative Filtering:

Collaborative filtering is a technique which will filter information supporting the user's interest (i. e. history) and therefore the ranking of other users with an equivalent interest. This filtering technique works on an outsized group of individuals to look and find a smaller set of users with related interest. It creates a ranked list of suggestions. It follows the things which are liked by users and combines them to make a ranked list of suggestions. It's widely utilized in many filtering systems or recommender systems, especially in e-commerce applications. Samples of such programs are Amazon. com, YouTube, eBay, etc. where recommendations for new products are given to a user supported by the user's past shopping history, likes and dislikes of comparable users.

2. 3 Policy-based Personalization of OSN Contents:

In OSN, there have recently been some operator classification frameworks to personalize entry. Several classification methods for short text messages were proposed to a large number of users of micro blogging services [4]. Twitter, the device focuses on a community of categories and provides content updates. The consumer

would then see those types of tweets that support his or her interests. Kuter and Golbeck, on the other hand, have created a film trust application that uses OSN relationships of trust and origin information to customize location access. However, these systems lack a policy layer of filtering that allows the user to see the results of the classification process and choose how to filter out the unwanted data.

Table 1. Related Work

Related Work	Objective
J. Golbeck and kuter[5]	The Film Trust application claims that the OSN title deed relationship and provenance information make it simple to access the website in this post. Such systems lack a filtering policy that allows the user to use the results of the sorting process to decide how to filter and how much unnecessary data to exclude.
P. J. Hayes, P. M. Andersen, I. B. Nirenburg and L. M. Schmandt [9].	The text categorization approach is used in this article. It means that a TCS run-time framework and a set of rules are included in the TCS text sharing programme. The legal structure establishes what classifications you should give documents in the text, as well as the rules that govern how you classify texts into different categories.
N. J. Belkin and W. B. Croft[10].	Unlike database applications, which use highly structured data such as an email, this method of data filtering uses informal or partially structured data. The email format varies depending on where it is sent, but the email body contains random data.
P. W. Foltz and S. T. Dumais[11]	The method of analyzing structured data is discussed in this paper. By evaluating the user's response and keywords given by the user, this approach determines the user's interest. Latent Semantic Indexing and keyword matching methods are used. The same choices are considered for this person.
Zelikovitz and Hirsh [12]	The author of this paper intends to enhance the separation of short text strings by enhancing sub-reading, a technique that combines labeled training data with a second collection of unwritten but longer linked texts.
Bobicevand Sokolova[13]	This paper focuses on using a mathematical learning approach that works well without feature engineering to avoid the problem of error-making functionality. The method used to develop and generate a language model used in classifiers of hard text and not easily integrate soft, multi-membership paradigms is described in this paper.

3.   Proposed System:

In this section, we introduce a Proposed System Architecture to filter unwanted posts on the user wall. There are three levels in the current system, as shown in figure 1. The user interface is provided by the Social Network Manager Layer. External

activities are supported by the Social Network Applications layer. The third layer is graphical user interfaces (GUIs), which are used to display the output. Users communicate with the system through a graphical user interface (GUI) to set up and manage their FRs/BLs (Blacklists).
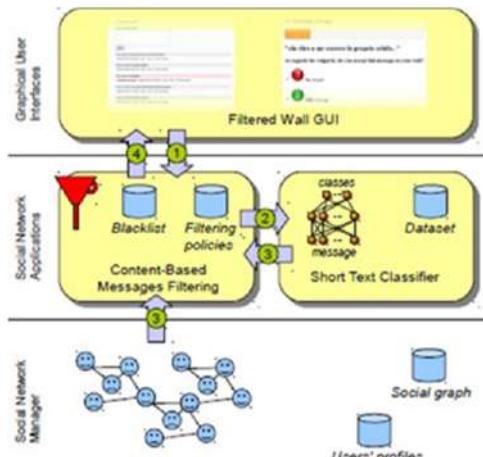
Figure 1. Proposed System
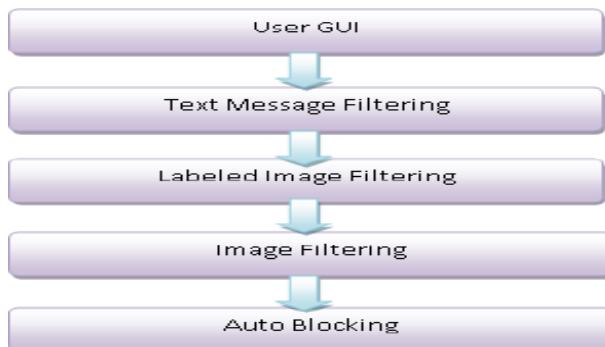
Figure 2. Layers of proposed system architecture

As shown in figure 2, it describes the layers of the proposed system. The User GUI is used to post messages which can be in any form like text message, labeled image or image. An Auto-blocking facility is provided for an authenticated user.
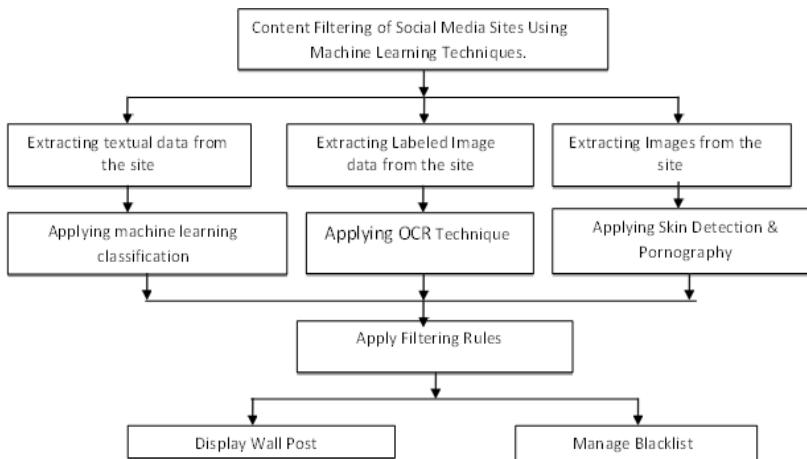
Figure 3. Flow of the Proposed System

•To post a message on the wall, a user must first navigate to the wall, which is divided into three sections: text filtering, labeled image filtering, and image filtering.

•The data of the content is retrieved using a text-based classifier for text filtering. It corresponds to the data collection. If a match is found, the material will be hidden from view on the wall.

•Image Filtering: With the help of the Skin detection algorithm and pornography detection algorithm, we filtered unwanted images from the wall. If an image is found in the dataset, then it is not displayed on the wall.

•In labeled image filtering, we used both previous methods. We can extract the text from an image using OCR techniques and apply message filtering rules.

•According to this logic, the output will be filtered and published on the wall.

4. Blacklist Management and Filtering Rules

We describe blacklist management and filtering rules here.

4.1 Auto blocking

If a particular user continuously posted unwanted messages on the wall, we can block that person.

4.2 *Filtering rules*

Users can specify which content should be blocked or displayed on the filtered wall using filter rules. Filter rules are specified according to the user profile and use of social media. An author is a person who defines the rules. OSN users are denoted by Creator Spec, and Content Spec is a Boolean expression

5.   Result

Figure 4 shows the Post Comment Page. The user can select the friend's wall and post the content, such as text messages, labeled images or images.
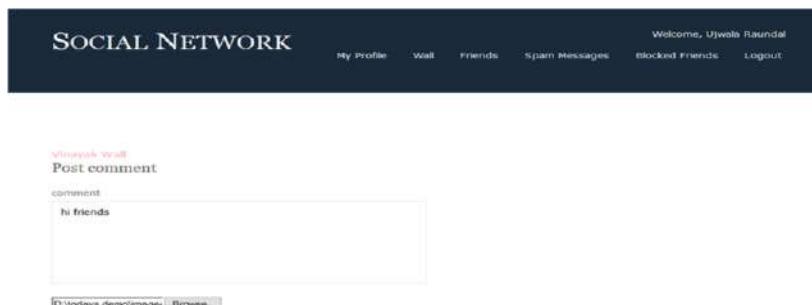
Figure4. Post Comment Page

Figure 5. Shows the list of unwanted messages and blacklisted users.

Figure 6. Spam Messages and Blacklisted users

Figure 7. shows the vulgar image which is filtered by the system. It should not be posted on the wall.

Figure 7. Image Filtering

6. Conclusion

In this paper, we presented a system to filter out unwanted messages, labeled images, and unwanted images from the user wall. The system uses text classifiers and machine learning techniques to filter text messages. For filtering labeled images, we have used OCR techniques. With the help of a pornography detection algorithm, we can detect vulgar images. Our system also provides an auto-blocking system. So only authorized users can post the messages on the user wall.

References

[1]. Vanetti, ElisabettaBinaghi, Elena Ferrari, Barbara Carminati, Moreno Carullo Department ofComputer and Communication, University of Insubria "a system to filter out unwanted messages walls OSN user" IEEE Transactions on Knowledge And Engineering Flight Data: 25 the Year 2013

[2]. F. Sebastiani, "Machine learning in automated text categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002. J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg, "Predicting Positive and negative links in online social networks," inProc. 19th Int. Conf. World Wide Web, 2010, pp. 641-650.

[3]. Vinaitheerthan Renganathan1*, Ajit N Babu2 and SuptendraNath Sarbadhikari3 " A Tutorial on Information Filtering Concepts and Methods for Bio-medical Searching".

[4]. B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in Twitter to improve information filtering," in Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2010, 2010, pp. 841–842.

[5]. J. Golbeck, "Combining provenance with trust in social networks for semantic web content filtering," in Provenance and Annotation of Data, ser. Lecture Notes in Computer Science, L. Moreau and I. Foster, Eds. Springer Berlin / Heidelberg, 2006, vol. 4145, pp. 101–108.

[6]. Pennock DM, Horvitz E, Lawrence S, Giles CL (2000) Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach. Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence (UAI-2000), Morgan Kaufmann Publishers Inc 473-480.

[7]. Boger Z, Kuflik T, Shapira B, Shoval P (2000) Information filtering and automatic keyword identification by artificial neural networks. Proceedings of the 8th European Conference on Information Systems.

[8]. Jennings A, Higuchi H (1993) A user model neural network for a personal news service. User Modeling and User- Adapted Interaction 3: 1-25.

[9]. P. J. Hayes, P. M. Andersen, I. B. Nirenburg, and L. M. Schmandt, "Tcs: a shell for content-based text categorization," in Proceedings of 6th IEEE Conference on Artificial Intelligence Applications (CAIA-90) IEEE Computer Society Press, Los Alamitos, US, 1990, pp. 320–326.

[10]. N. J. Belkin and W. B. Croft, "Information filtering and information retrieval: Two sides of the same coin?" Communications of the ACM, vol. 35, no. 12, pp. 29–38, 1992.

[11]. P. W. Foltz and S. T. Dumais, "Personalized information delivery: An analysis of information filtering methods," Communications of the ACM, vol. 35, no. 12, pp. 51–60, 1992.

[12]. S. Zelikovitz and H. Hirsh, "Improving short text classification using unlabeled background knowledge," in Proceedings of 17th International Conference on Machine Learning (ICML-00), P. Langley, Ed. Stanford, US: Morgan Kaufmann Publishers, San Francisco, US,2000, pp. 1183–1190.

[13]. V. Bobicev and M. Sokolova, "An effective and robust method for short text classification," in AAAI, D. Fox, and C. P. Gomes, Eds. AAAI Press, 2008, pp. 1444–1445.

[14]. http://www. myvocabulary. com/word-list/violencevocabulary

[15]. https://archive. ics. uci. edu/ml/datasets/SMS+Spam+Collection

[16]. http://www. bannedwordlist. com/lists/swearWords. txt

445

Identification of Fake Video Using Smart Contracts and SHA Algorithm

SwapnaliTambe [a][1], Dr. Anil Pawar [b], Dr. S K Yadav [a]
[a]Research Scholar, JJTU University, Rajasthan, India
[b]Dean Academic Sanjivani COE, Kopargaon, India

Abstract. Deepfake is as a matter of fact a medium where one individual is supplanted by another who appears as though him. The profound bogus demonstration has been continuing for quite a long while. Profound phony uses incredible strategies, for example, AI and man-made consciousness to create and control visual and sound substance with high potential for the gadget. Profound misrepresentation relies upon the sort of impartial association called and the programmed encoder. These are essential for an encoder, which lessens a picture to a lower dimensional ideal and an ideal introduction picture. I examined various answers on various advances via web-based media stages like twitter and face book. From these examinations we are roused to extend this objective. In our proposed framework, we centre around identifying profound phony recordings utilizing blockchains, keen agreements, and secure hashing calculations. We utilize a few calculations to relieve the issue, for example, the SHA string

Keywords: Block Chain, Smart Contracts, SHA Algorithm

1. Introduction

Deepfakes are horrible for security, the confirmation of society and well-known government. At the point when this peril was introduced, methods for recognizing deepfake were proposed. Early undertakings relied upon made features got from abnormalities and sham video association relics. In relationship, continuous procedures have applied significant sorting out some way to thusly isolate critical and one-sided characteristics to recognize significant disfigurements [1]. A large portion of the present warmth exchangers are changing to a just progressed arrangement with no paper support being saved. This has ideas for dependability, check and provenance in various zones, like the arraignment, where the two players should be happy with the trustworthiness of the high level test, or occupied with security, where cases can be productive or come up short. In especially dependable terms and conditions; you may have to know precisely what the terms and conditions were at the hour of the arrangement. You in addition need to check the conditions of a game plan that were applied when the arrangement was concurred and set the essential creation dates when copyright issues emerge concerning mechanized substance. Moreover, there are ensured necessities for setting up earlier information prior to consenting to secret courses of action [09, 10]. An immense piece of this is that the crude information

---

[1] swapnali N Tambe, Research Scholar, JJTU University, Rajasthan, India
Email: swapnali4231@gmail.com

substance of a chronicle can be checked in any case, when the metadata is changed because of moving the record between working designs or contraptions. Summed up and feasibly open contraptions have gotten fundamental for video blend. Guaranteeing the authenticity of video occupations, for instance in court, acknowledgment offices, advancements, and the universe of diversion is basic. Thusly, there is a ton of investigation in the space of video approval and position adjustment systems. Moreover, research is in progress utilizing Darwin's information structure configuration to guarantee the unwavering quality of cutting-edge appropriation applications [2, 11]. While there are various open apparatuses for gathering, encoding and isolating data, there are respectably not many open in the space of unwavering quality, check, and provenance confirmation. For example, you make and issue a verification that contains a SHA 256 hash assessment of the sent media close by customer nuances and a timestamp. This confirmation, hence, is affirmed by a high-level presentation gave by the essential Comodo authentication authority. Regardless, the issue here is that the principal report and the attestation are discrete components and could be easily detached while scattering or flowing the record [3, 4 and 18].

2. Review of Literature

In [01], display that a modernized underwriting relating to an image record can be set in that image archive close by the attached metadata containing references to the capable association. Despite the assortments among devices and between working structures and applications, a JPEG record holds its plan well in general. Right when changes occur, they by and large occur in the metadata locale and don't impact the data of the encoded picture, which is the center of the record and the part that ought to be sure. References to the capable association can be implanted in the archive's metadata. There is the advantage of having the high-level confirmation as a key piece of the archive it applies to and travel with the record suitably. Finally, we show that the metadata inside an archive offers the likelihood to join data that can be used to exhibit the uprightness, believability and provenance of the high-level substance inside the record.

The paper [02], clarify how deepfakes are a genuine danger to society, the type of government and organizations since they put focus on columnists who battle to channel the genuine news from the phony, they compromise security spreading promulgation that meddles with decisions, thwarting residents' trust in data given by specialists, and disposing of digital protection issues for people and associations. This examination territory expresses that there are numerous elements included that decide the idea of deepfakes. Contentions can be made, both for and against, however they may be basic if improvement and admittance to deepfake age apparatuses are appropriately administered. On the off chance that firmly checked, deepfakes can be utilized to help current mankind as opposed to making it fall.

FIPS [03], this article explains the Secure Hash Signature standard, four secure hash computations are shown in this standard: SHA-1, SHA-256, SHA-384 and SHA-512, to enlist a thick depiction of electronic data (message). Exactly when a message of any length <264 bits (for SHA-1 and SHA-256) or <2128 bits (for SHA-384 and SHA-512) is gone into a computation, the result is a yield called the message digest. Message digests change long from 160 to 512 pieces, dependent upon the estimation. Secure hashing computations are conventionally used with other cryptographic estimations, as

modernized mark computations and keyed hash message check codes, or in the time of sporadic numbers (bits). The four hashes showed in this standard are called protected considering the way that, for a given estimation, it is computationally hard to 1) find a message that facilitates with a given audit of the message, or 2) find two special messages that produce a comparative synopsis of the message. Any movements in a message will, with a particularly high probability, achieve another message digest. This will achieve an affirmation goof when the ensured hashing computation is used with a high-level imprint estimation or a keyed hash message approval computation.

The article [04, 08, 12], presented a blockchain-based response for modernized video validity testing in which secure and trustworthy perceptibility to the primary video. This print course of action can help fight deepfake video and sound by helping customers with choosing if video or progressed substance is unmistakable to a genuine and reliable source. If a video or mechanized substance can't be followed, the high-level substance can't be trusted. Our sagacious agreement-based game plan gives discretionary skilled workers a strong strategy to request approval from the main specialist to copy and modify chronicles.

This article [05], focused in on the entertainment and replacement of human deepfakes. This article gives a start to finish assessment of how these headways work, the divergences between their designs, and how is being managed recognize them. We believe this information will be helpful to the neighborhood cognizance and thwarting malicious deepfakes.

The paper [07, 15], profound learning has been viably applied to deal with a couple of complex issues going from tremendous data examination to PC vision to human level control. Regardless, moves in significant learning have moreover been used to create programming that can make risks security, dominant part rule government and public wellbeing. One such significant learning application that has actually emerged is deepfake. Deepfake's computations can make fake pictures and accounts those individuals can't perceive from real ones. Thusly, proposing propels that can normally recognize and survey the uprightness of mechanized visual media is basic. This article presents an examination of the computations used to make deepfakes and, even more essentially, the techniques proposed for distinguishing deepfakes to date.

3. Proposed System

Our proposed framework appeared in Figure 1 fundamentally our framework completed in three distinct advances: administrator login, client login, programmer login. In this paper our emphasis is on the client login and programmer login, where the framework accepts the video as information and cycles that video for required yield. Through client login we are transferring credible video utilizing secure hash calculation. 'Secure Hash Algorithm' (SHA) is a gathering of cryptographic hash limits made by the US 'Public safety Agency' (NSA) and disseminated as a standard by the US 'Public Institute of Science and Technology' (NIST). It is the crucial estimation for secure applications used by US government associations. A critical component of SHA estimations is that they do an intensifying wonder, which infers that a little change in the data achieves a tremendous change in the yield regard [16].
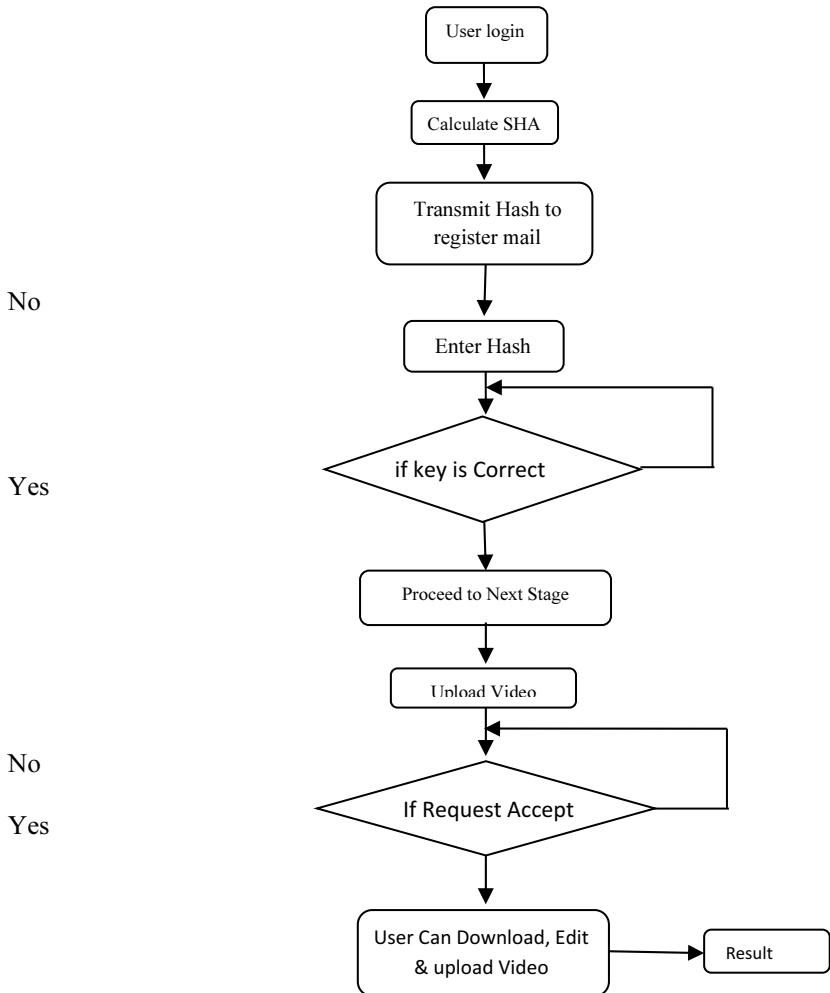
Figure 1. Proposed system for the smart contract

The main variation, SHA-1, makes a 160-piece regard and was conveyed in 1993; regardless, it was taken out not long after appropriation due to an undisclosed shortcoming and a changed structure was conveyed two years afterward. In 2002, NIST conveyed the SHA-2 gathering of limits. Not in any way like SHA-1 with a fixed 160-piece hash size, SHA-2 is offered in six structures that produce an extent of yield sizes from 224 to 512 pieces, the most ordinarily used being SHA-256 and SHA-512. Like MD5 and SHA-1, SHA-2 limits rely upon the Merkle-Damgård advancement. The computation used for SHA-256 is:

1. A 256-bit data support is made, included 8 to 32-digit words which are taken care of with the underlying 32 bits of the incomplete bits of the square hidden establishments of the underlying 8 primes [01, 03 and 17].

2. A 64 segment table of constants is prepared using the first 32bits of the fractional bits of the 3D shape establishments of the underlying 64 Primes [01, 03 and 17].

The data is padded with a primary piece "1" and the length of the main data is conveyed as a 64-cycle number, segregated by the number of zeros expected to make the message length, including padding, an alternate of 512 – little [01,03 and 17].

3. Each 512-cycle block is taken care of through 64 rounds where each round incorporates a movement of exercises included bitwise undertakings and estimated extension [01, 03 and 08].

4. The assessment of the pad on fulfilment of each square is the fundamental impetus for the going with square; around the completion of the last square, the support contains the hash regard [01, 03 and 17].

While transferring a credible video as demonstrated in the flowchart, we should initially utilize the client login through which a SHA is determined and a hash key is created which is shipped off a valid email ID which is a brilliant agreement. The primary expert sharp arrangement is made using attributes, for instance, the owner having the Ethereum address of the main specialist and mappings containing video detail records reliant on the circumstance with assents permitted or denied. Moreover, regardless of sales are put something for reference and history following. In like manner, a critical once-over that guides in conspicuousness is the overview of designated authentication accounts which are seen as assistant chronicles of the principal understanding. Savvy contracts are robotized programs that encode conditional auto-execution rationale and are implemented utilizing decentralized encryption strategies. Notwithstanding their name or legitimate status, brilliant agreements produce extraordinary premium and venture since they can fundamentally change how the gatherings cooperate. Savy Contract makes the first craftsman video. In this arrangement, an understanding is made for the other individual who wishes to acquire consent to change, adjust and even circle these recordings identifying with the T and Cs gave in the arrangement. This solicitation for understanding is put away on the interplanetary record framework worker. Hash planning is accessible in the savvy contract. Someone else demands a solicitation to alter, share or alter this video content. At the point when an auxiliary craftsman demands the proprietor, it implies that they affirm the terms and states of the arrangement [13,14]. The solicitation is gotten by the proprietor and afterward the yield is shown.This arrangement not exclusively can take various consent at time yet in addition can taking different solicitation sent from same proprietor. At the time proprietor affirms the application, they made the youngster contract thought about at the underlying arrangement and they update parent's information. Now, the another proprietor, request a consideration of their as of late maded arrangement through beginning proprietor by contact of the main video. At same time, beginning proprietor point supports and grants a validation at that point noticing an as of late made SC. An adequately affirmed shrewd agreement would by then be incorporated the kid inside a remarkable SC. They address the Ethereum of the other segment of their qualities appeared in Figure 2 stream graph.
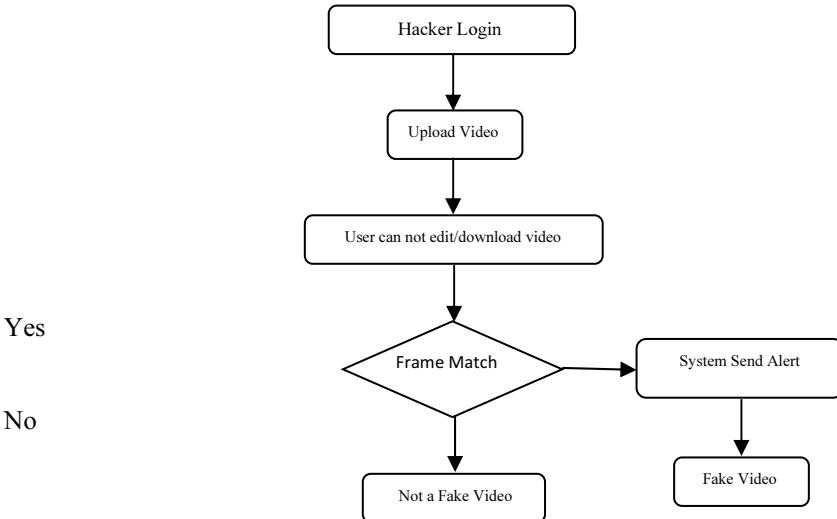
Figure 2. System for the identification of video

4. Testing

Shrewd Contract makes the first craftsman video. In this understanding, the arrangement is made for the other craftsman for whom you wish to get consent to adjust, alter and even course these T&C related recordings gave in the arrangement. This solicitation for understanding is put away on the interplanetary record framework worker. Hash planning is accessible in the keen agreement. Another proprietor demands a solicitation to alter, share or alter this video content. At the point when an optional craftsman demands the ownerit suggests that they confirm the terms and conditions of the plan. The solicitation is gotten by the proprietor and afterward the yield is shown. Not exclusively would this be able to contract demand various consents simultaneously, it can likewise acknowledge different solicitations sent by a similar proprietor. Right now the proprietor affirms the solicitation, has gone into the agreement for the youngster as for the underlying agreement and updates the information of the guardians. Now, the other proprietor demands a consideration of their recently closed agreement through the underlying proprietor by means of the contact of the main video. Simultaneously, the underlying proprietor point backs up and ensures validation, at that point takes a gander at a recently made SC. A keen agreement really affirmed around then would incorporate the youngster inside a solitary SC. They go to Ethereum on the opposite side of your traits.Smart contracts make a connection between unmistakable substances, the SC is the main proprietor is a proprietor who has that Ethereum address in the underlying proprietor and tallies until he keeps the video records. Additionally, all solicitations are put something aside for history following and following. In the extension, a fundamental rundown has an effect in the discoverability rundown of permitted affirmation enlistments that are viewed as optional enrollments of the underlying agreement. Each agreement comprises of a video. In this manner, 1: 1 is the connection between the arrangement and the substance of the video. Besides, every video is connected as though it were an

Ethereum address with a proprietor. Moreover, a tricky arrangement can have a few kid contracts dependent on effective verification. Thusly, a 1: N association between the mischievous proprietor of the underlying understanding. At last, the Interplanetary File System is likewise a substance with a 1: 1 association with any SC, every video is moved to the Interplanetary File System workers, and its Interplanetary File System hash is a property inside the SC. Moreover, the T&C understanding structure of each agreement is passed notwithstanding the interplanetary document framework worker and its hash is a quality inside the keen agreements made for enlistments. Moreover, the complete code available to all inconspicuous components was additionally made. Following pictures are some Hash kay produced test appeared in Figure 3,
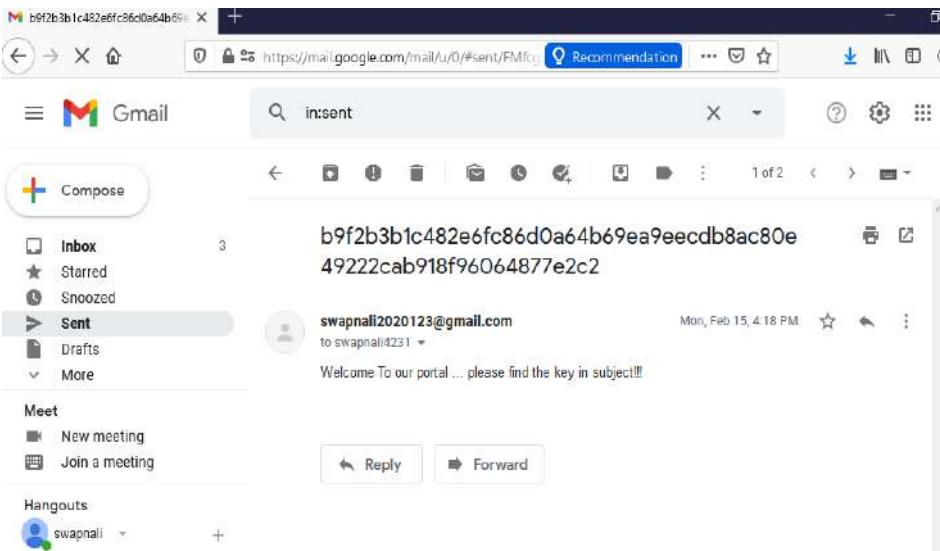
Figure 3. Hash Kay send to email

5. Conclusion

Free admittance to make and share data that has no realities behind it via online media stages like WhatsApp and other advanced stages has uncovered another issue of bogus data, which has produced bits of hearsay all throughout the planet. In this paper, we have introduced a climate arrangement of an Ethereum blockchain-based solution for confirmation of verify of advanced recordings where secure and dependable discernibility can be set up to the maker or wellspring of the first video, in a totally decentralized. Our answer permits web-based media clients to approach believed information from computerized content so they can follow the information and have certainty that the information is genuine. The arrangement utilizes Ethereum shrewd agreements and IPFS decentralized capacity framework. The Ethereum wallet executes keen agreements for recordings and IPFS is utilized to store the metadata of the recordings and furthermore creates an extraordinary hash of the recordings to find the documents in IPFS. The development of our proposed plan, system arrangement, progression diagrams and execution nuances can be applied to any automated substance like video and pictures. This Smart Contract-based game plan gives

discretionary experts a strong strategy to request assent from the primary specialist to copy, change, share and adjust chronicles. Work is in progress to plan and do a totally valuable and operational decentralized standing system. At this moment, the arrangement for the Ethereum framework, savvy arrangements and private chain has been successfully made.

References

[1] Martin Harran, William Farrelly, Kevin Curran, "A method for verifying integrity &authenticating digital media", Applied Computing and Informatics, 2017

[2] Hrisha Yagnik1, Akshit Kurani2, Prakruti Joshi3, "A Brief Study on Deepfakes", IRJET, Volume-7 Issue-12, DEC 2020.

[3] FIPS, "Secure Hash Standard",Federal Information Processing Standards Publication 180-2, 01 Aug 2002

[4] Haya R. Hasan And Khaled Salah, "Combating DeepfakeVideos Using Blockchainand Smart Contracts"IEEE Access April 12, 2019, Digital Object Identifier 10.1109/ACCESS.2019.2905689.

[5] YisroelMirsky,Wenke Lee, "The Creation and Detection Of Deepfakes: A Survey", ACMComputing Surveys, Vol. 1, No. 1, January 2020

[6] Alok Chauhan, Amit Kumar,"Establishing Environment Setup for Preventing DeepfakesusingBlockchain Technology",MuktShabd Journal, Volume IX Issue V, MAY/2020.

[7] Thanh Thi Nguyen, Cuong M. Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, SaeidNahavandi, "Deep Learning for Deepfakes Creation andDetection: A Survey", aXiv: 2019.11573v2 [cs.CV].Jul 2020

[8] Paula Fraga-Lamas, Tiago M. Fern´andez-Caram´es. "A Review on the Use of Blockchain for the Internet of Things",IEEE Access, Electronic ISSN: 2169-3536, 31 May 2018.

[9] H. Allcott, "Social Media and Fake News in the 2016 Election",Journal of Economic Perspectives, Vol. 31, No.2 Spring 2017.

[10] R. Chesney and D. K. Citron., "Deep Fakes: A Looming Challenge for Prives: A Looming Challenge for Privacy, Democr , Democracy, and National Security", California Law Review, Vol. 107:1753 2018

[11] Potharaju, S. P. (2018). An Unsupervised Approach For Selection of Candidate Feature Set Using Filter Based Techniques. Gazi University Journal of Science, 31(3), 789-799.

[12] P. Treleaven, R. G. Brown, and D. Yang, "Blockchain technology in finance," Computer, no. 9, pp. 14–17, 2017.

[13] F. Tian, "A supply chain traceability system for food safety based on haccp, blockchain& internet of things," in Service SystemsandServiceManagement(ICSSSM), 2017InternationalConferenceon. IEEE,2017, pp. 1–6.

[14] Potharaju, S. P., & Sreedevi, M. (2017). A Novel Clustering Based Candidate Feature Selection Framework Using Correlation Coefficient for Improving Classification Performance. Journal of Engineering Science & Technology Review, 10(6).

[15] M. A. Khan and K. Salah, "IoT security: Review, block chain solutions, and open challenges," Future Generation Computer Systems, vol. 82, pp. 395–411, 2018.

[16] K. Biswas and V. Muthukkumarasamy, "Securing smart cities using block chain technology," in 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Dec 2016, pp. 1392–1393.

[17] K.Christidis and M. Devetsikiotis, "Block chains and smart contracts for the internet of things," IEEE Access, vol. 4, pp. 2292–2303, 2016.

[18] Swapnali N. Tambe, A.B.Pawar, "Detecting fake Videos Using Block Chain and Smart Contracts", IJRTE, ISSN: 2277-3878, Volume-8 Issue-4S5, December 2019.

Blockchain Technology: Rising Trend in Various Applications

Dr. Zalte S. S [a,1], Patil P. N [b], Deshmukh S. N [b], Patil M. G [b], Katkar P. S [a]

[a] *Department of Computer Science, Shivaji University, Kolhapur, India*
[b] *Department of Computer Science, Vishwakarma College of ACS, Pune, India*

Abstract. In this tech world, size of block chain is exponentially growing in this covid pandemic year which was not predicted. This pandemic environment causes digital transformation drive in various areas, particularly via the utilization of block chain or distributed ledger technology. To enhance and accelerate business process in various organization and industry showed a growing interest to adopt this technology. This paper summarized various block chain applications which are widely used in number of sectors and also focus on some challenges becomes hurdle while adopting block chain technology.

Keywords: Blockchain, Challenges, Covid, Distributed Ledger, Bitcoin.

1. Introduction

Blockchain is a technology that permits dispensed public ledger through immutable facts in an encrypted and secured way. This ensures a safe transaction that cannot be altered. Blockchain is a decentralized transaction, and data management technology [1]. Latest hyper-linked global, or the "global village", is challenged with the upward thrust of human sickness epidemics and pandemics because of complex interactions among a couple of factors which encompass ecological, environmental, and socioeconomic elements. The enormously infectious severe Acute respiratory Syndrome Corona virus2(SARS-CoV-2), which reasons then oval corona virus disorder (cited here after as COVID-19), has been wreaking havoc around the globe. Number of blockchain solutions had been carried out all through the COVID-19 pandemic. Public health strategists and policymakers explore distinct procedures to lessen the devastating consequence so novel corona virus disorder (COVID-19).

Blockchain generation has emerged as a useful resource that may be utilized in several tactics. Many blockchain technologies had been proposed or implemented during the COVID-19 pandemic [2].

---

[1] Patil M.G. Dept. of Computer Science, Shri S.M. Mahavidyalaya, Pune.
E-mail: minalpatil8491@gmail.com

2. Challenges in Block chain technology

As per our discussion in above paragraph, the block chain technology offers very comfort solutions like transparency, decentralization, integrity, immutability, and security without requiring any centralized trusted authority. However, some challenging issues are to be addressed for various applications.

Energy Consumption: - To validate the transaction for insertion into network Proof-of-Work mechanism is used, which requires lot of computational power for the processing of complex mathematical puzzles.

Scalability: - Ability to handle plenty number of nodes at a time is challenging task for this technology. It also involves complication calculation for single transaction which may slow down performance.

Privacy: - As it is open ledger, anyone can view the contents, which might be beneficial in many applications, but in case of sensitive application it becomes liability.

Lack of Talent: - Currently, there are few employees to build decentralized block chain. Educating employees to work with block chain will be lengthy process.

Security: - Security is another crucial topic here. We have a tendency to all skills each block chain technology boasts regarding its security. However, like every different technology, block chain conjointly comes with a couple of security loop holes. There are various attacks on the network is one among the protection flaws of the network. While imposing these attacks, hackers will take over the network and exploit it in their method. They will even modify method and prohibit others from making a block.

3. Pillars of Blockchain Architecture:

The main four pillars of block chain architecture are shown in Figure 1. The blockchain architecture consists of a few essential ideas like decentralization, digital signature, mining, and data integrity [3].

Decentralization: -Rather than to give preference to one vital authority overpowering others in the environment, blockchain explicitly distributes control among all participants inside the transaction chain.

Digital signature: - Block chain allows a change of transactional price the usage of public keys through the mechanism of unique digital anatomy.

Mining: - In a dispensed device, every person mine and digs deep into the statistics that are then evaluated in line with the cryptographic rules and it additionally recognizes miners for confirmation and verification of the transactions.

Data integrity: - Complicated algorithms and agreement among customers guarantees that transaction information, as soon as agreed upon, can't be tampered with and hence remains unaffected. Information saved on block chain acts as an unmarried version of reality for all parties worried as a result reducing the hazard of fraud [4].
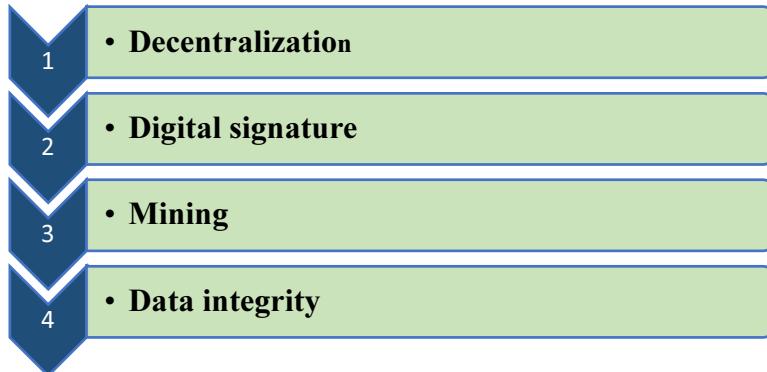
Figure 1. PillarsArchitecture of Blockchain

4. Block chain Applications:

Blockchain applications go so much on the far side crypto-currency and bitcoin, with its ability to form a lot of transparency and fairness. Here, we have summarized some applications of blockchain in various sectors as shown in Figure 2.

Figure 2. Application of Blockchain

4.1. Land Registry

A future-evidence, Block chain the actual estate market is building up new commercial enterprise models of connecting potential shoppers and sellers. In blockchain solution, there's no need of third-party, lawyers and brokers. The ubiquity joined with the real property signs up a workplace in Brazil and advanced the undertaking which pursuits to reduce a fee even as improving, a protection and a transparency. The ubiquity seeks to great immutability to belongings possession data treated through the land an information office. Facts stored on Blockchain are immutable. Block chain-based a totally method to register assets titles could appreciably boom the performance of convincing or even prevent a fraud [5].

Merits:- Use of block chain in land registry process can able to transfer land ownership legally to the buyer without involvement of third party. It can also bring transparency and reduce frauds with the help of smart contract. You can generate unique digital id for buyer and seller which accelerate process of ownership transfer. Once registration complete, smart contract update land title for every new buyer and information regarding to this is permanently stored in block chain.

4.2. Voting:

Existing voting system faces lot of issues like fraud, long delay time, accessibility, affordability and privacy. Voting strategies stay in many countries a controversial topic, as incident of electoral fraud (invalid or faulty vote, a couple of registration) and the large percentage of abstention frequently shape the end result.

The variation of block chain technology promise trust, security in transaction and transparency. The voters may want to connect with a pc-primarily based system through their laptop, computer or telephone, the use of open-supply code this is open to modify the use of a kind of authentication (biometric, written) prove their identity to this system. Then, they enter their private key to access their proper to vote and the usage of their public key to pick their choice and affirm it. Thus far, three tasks had been founded that sell voting through blockchain systems. The first is, "Bit Congress" that makes use of the Ethereum platform to develop its idea primarily based on the scenario that every voter has get entry to one "vote coin" that enhances him to vote only one time, and his vote will be recorded at the blockchain after the machine verifies it. Normally, the transformation of the voting gadget from paper-primarily based to digital will increase its reliability, and the benefit that gives to the voters [6].

Merits: -The voters can submit their vote from anywhere without burden of geographical barrier and without revealing their original identity. Fake counting of vote can be reduced by counting one ID for one vote. Voter's hash can keep voter's vote confidential. Voter can able to tally whether his/her vote was included or not.

4.3. Healthcare

Healthcare systems control and monitor the health of patients with the assist of superior technologies. The advancement of these structures desires to contain an unequivocal spotlight on making these systems efficient [7]. Blockchain generation should play a pivotal position in the healthcare enterprise with numerous programs in areas like public healthcare control, longitudinal healthcare information, automatic fitness claims adjudication, online affected person get entry to, and person oriented medical studies, drug counterfeiting, medical trial, and precision medicine. Especially, blockchain era and using SCs could resolve issues of clinical credibility of findings (lacking data, endpoint switching, facts dredging, and selective guide) in clinical trials as well as problems of patients' knowledgeable consent [8]. Blockchain carried out to the fitness sector can provide new and powerful possibilities to enhance numerous activities related to the prevention and manipulate of pathologies and, therefore, higher medical danger management within the context of a virulent disease emergency consisting of the current one. The sudden appearance and the speedy and out of control diffusion around the world of Corona virus has shown us no longer only the failure of existent healthcare surveillance systems in directly dealing with the public health emergency, however additionally an obvious lack of advanced predictive systems

based totally on the sharing of scientific records on a large scale, capable of saves you or as a minimum reduce emergencies of such value [9].

Merits: -Automatic update of patient's health information is possible and information can be sharable with other hospitals or organization through sharable private key. Block chain help to securely tracking of pharmaceutical supplies with full of transparency. It also helps in health insurance process to present medical events in timely manner and helps to reduce fraud.

4.4. Bitcoin: The first block chain

A bitcoin is a peer to see digital currency that permits online payment from one party to any other without a want for third party. It turned into proposed to remedy the shortcomings of fiat cash and banks. A bitcoin uses blockchain to report the transaction. (Nakamoto, 2008) The bitcoin protocol become brought in 2008, by using pseudonym Satoshi Nakamoto in a white paper named, "Bitcoin: A Peer-to-Peer digital cash gadget". The bitcoin is the primary and the most popular cryptocurrency. The current marketplace capitalization of bitcoin is $1186.36 billion (Statista 2021). Blockchain is the era in the back of the bitcoin. Blockchain can be meant as a running system and bitcoin as one of many programs. Blockchain gives disbursed ledger to report bitcoin transactions and has many makes use of beyond bitcoin [10].

Merits: Users can transfer money to anyone in the network without dealing with the bank or government or third party, so transaction fees are also low for international payments. Anonymity is also preserved because bit coin purchases are discrete. It is easily accessible through computer, laptop etc.

4.5. Banking:

Block chain technology is a center, underlying era with promising utility potentialities in the banking industry [11]. Banking and technology are very intently related and innovations have modified banking considerably over the time frame. The virtual improvements within the banking sector commenced with the creation of cash that changed the barter machine, and then the sluggish substitute of wax seal with digital signatures. Such innovation that is changing the banking sector globally is Block chain technology (BCT). Blockchain is shared allotted ledger which stores business transaction to an everlasting unbreakable chain which may be viewed through the parties in a transaction. Blockchain generation has the ability to disrupt the financial enterprise applications as it affords permanent and tamper evidence recording of transactions in a disbursed network. It could be extensively applied in digital currency, trade finance, KYC and go border remittances, and so on [12].

Merits: Customer's identity verification process is accelerated using block chain and also this verification can reuse for different services to reduce time. Peer-to peer payment transfer is possible without involvement of third trusted authority.

4.6. Online Music

Block chain technology is often seen as "the next thing" that can change the structure of different industries. The music industry is no exception in that respect as a lot has been already said regarding the implications of block chain for the industry as a

whole, including an increase in the value captured by artists, a shift of market power, disintermediation, more transparency and more efficient systems [13].

Block chain technology has stand up as a technique to bring secure, reliable, and scalable distributed transaction processing to music licensing. Significant investment and technical talent have entered the music trade through block chain technology. A block chain is a distributed ledger – a database of transactions, of which each and every party to each group action incorporates a complete copy at all times. Block chains take advantage of low-cost storage, reliable internet connection, and progressive security to switch massive, monolithic, proprietary systems with open, light-weight, distributed systems [14].

Currently, music will be streamed and downloaded at the clicking of a button, however, payments to those that build that music can be slow and opaque. Block chain technology offers transparency through the worth chain, permitting musicians and their managers to examine specifically how much cash they're owed, as against a culture of non -disclosure agreements and black boxes. Though some of the claims created for block chain technology are premature, it seems to possess a minimum of the potential to transform the music business [15].

MEDIACHAIN, UJO, CHOON, OPEN MUSIC INITIATIVE, MUSICOIN and many more music companies started with adopting this methodology to ensure that the artist are paid fairly, their royalty stipulations without the hassle of confusing third parties or contingencies, eliminates the confusion of music ownership and facilitates crowd funding for up-and-coming artists and rewards listeners for creating personalized playlists [16].

Merits: - Intellectual property rights of creative items can be preserved by distributed ledger by keeping evidence of ownership in block chain. Artists also aware of how much money made by publishing music album, so they can received fair amount within time.

5. Conclusion

In this era, of the blockchain, although in case of huge scale implementation engross some major serious problems that can't be unmarked. These challenges cause the sluggish adoption of block chain technology by numerous industries and organizations. The acceptance of blockchain's potential to act as a completely unique disruptor, across multiple industries and businesses could presently cause mass adoption. However, blockchain technology is going through a hype phase. To get rid of some hurdles and power the techno world we have to wait. In the future, career opportunities are growing to extend at an alarming rate. Integration with new-age technology like IoT can facilitate in building secure infrastructure. Enterprise block chain can still mature and develop, resulting in high job prospects and sensible pay.

References

[1]. Yli-Huumo, J., Ko, D., Choi, S., Park, S., &Smolander, K. (2016). Where is current research on blockchain technology?-a systematic review. *PloS one*, *11*(10), e0163477.

[2]. Abd-alrazaq, A. A., Alajlani, M., Alhuwail, D., Erbad, A., Giannicchi, A., Shah, Z., ... &Househ, M. (2020). Blockchain technologies to mitigate COVID-19 challenges: A scoping review. *Computer Methods and Programs in Biomedicine Update*, 100001.

[3].   Panicker, S., Patil, V., & Kulkarni, D. (2016). An overview of blockchain architecture and it's applications. *Int. J. Innov. Res. Sci., Eng. Technol.*, *5*(11), 20074-20084.

[4].   Shah, T., & Jani, S. (2018). Applications of blockchain technology in banking & finance. *ParulCUniversity, Vadodara, India.*

[5].   Barbieri, M., &Gassen, D. (2017, March). Blockchain–can this new technology really revolutionize the land registry system. In *Responsible Land Governance: Towards an Evidence Based Approach: Proceedings of the Annual World Bank Conference on Land and Poverty* (pp. 1-13).

[6].   Foroglou, G., &Tsilidou, A. L. (2015, May). Further applications of the blockchain. In *12th student conference on managerial science and technology* (pp. 1-8)..

[7].   Singh, P., & Singh, N. (2020). Blockchain With IoT and AI: A Review of Agriculture and Healthcare. *International Journal of Applied Evolutionary Computation (IJAEC)*, *11*(4), 13-27.

[8].   Shen, C., & Pena-Mora, F. (2018). Blockchain for cities—a systematic literature review. Ieee Access, 6, 76787-76819.

[9].   Fusco, A., Dicuonzo, G., Dell'Atti, V., &Tatullo, M. (2020). Blockchain in healthcare: Insights on COVID-19. *International Journal of Environmental Research and Public Health*, *17*(19), 7167.

[10]. Khadka, R. (2020). The impact of blockchain technology in banking: How can blockchain revolutionize the banking industry?,*Thesis in CENTRIA UNIVERSITY OF APPLIED SCIENCES Business Management October 2020.*

[11]. Guo, Y., & Liang, C. (2016). Blockchain application and outlook in the banking industry. *Financial Innovation*, *2*(1), 1-12.

[12]. Gupta, A., & Gupta, S. (2018). Blockchain Technology: Application in Indian Banking Sector. *Delhi Business Review*, *19*(2), 75-84.

[13]. Sitonio, Camila&Nucciarelli, Alberto. (2018). The impact of blockchain on the music industry, *29th European Regional ITS Conference, Trento 2018 184968, International Telecommunications Society (ITS)*.

[14]. Rosenblatt, B. (2017). Watermarking technology and blockchains in the music industry. *Retrieved from Digimarc(https://www.digimarc.com/docs/defaultsource/digimarcresources/whitepaper-blockchain-in-musicindustry.pdf.*

[15]. De León, I. L., & Gupta, R. (2017). The impact of digital innovation and blockchain on the music industry. *Inter-American Development Bank.(Nov 2017). Available online: https://publications.iadb. org/en/impact-digital-innovation-and-blockchain-music-industry (accessed on 23 June 2020)*.

[16]. Daley, S. (17). Blockchain Music Companies Reshaping a Troubled Industry. Built In. *Retrieved on November*, *3*, 2019.

Deployment of Computer Assisted Instruction in Higher Educational Organization

N. P. Chougale [a][1], K. G. Kharade [b], S. K.Kharade [c], S.R.Ghatage [d],
Mallappa G. Mendagudli [e] , S.Yuvaraj [f], K.Vengatesan [g]

[a] *KIT's Institute of Management Education and Research, Kolhapur, MH, India*
[b] *Department of Computer Science, Shivaji University, Kolhapur, MH, India*
[c] *Department of Mathematics, Shivaji University, Kolhapur, MH, India*
[d] *Gopal Krishna Gokhale College, Kolhapur, MH, India*
[e] *BLDEA's A.S.Patil College of Commerce, Vijayapur, KA, India*
[f] *Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Kattankulathur, TN, India*
[g] *Computer Engineering, Sanjivani College of Engineering, Kopargaon, MH, India*

Abstract. The current study looked at how computers help students of different reading abilities. Studies have been conducted on the effect of computers on educational achievement and attitudinal levels. During this study, we have gone through the impact of Bloom's taxonomy over the conventional system and then focused on integrating CAI in higher education. This research is branching out to encompass the use of computers in the education system. The paper aims to outline and elaborate on the way computers aid every department of life, including education. Information is much more valuable when presented within the context of education or training. Teaching to reflect well is known as "advance" teaching. This research paper is concerned with the educational faculties, facilities, and the organizational structure of the educational process. This study aims to focus on CAI to improve the quality of higher education in India. We have also discussed the areas where we can implement this technology

Keywords. CAI, Classroom lecture, Higher Education, Learning Process

1. Introduction

A person who uses a computer to help present the instructional material and track student progress is called instructional assisted interaction (IA). In Computer-Assisted Instructional (CAI), techniques are used to give instructions and test performance on a computer. As one moves through the learning process, one can move through various computer technologies that support it[3]. A variety of self-paced learning self-directed learning techniques include standard input into CAI linear programs, introduced elementary schools to collaboration with International Business Machines, which occurred in the mid-1950s and early 1960s. This CAI is a treasure trove of information that everyone who needs college and/or department information should have. CAI is relevant for preschool through to post-secondary[2].

---

[1] N. P. Chougale, KIT's Institute of Management Education and Research, Kolhapur
Email. navni48@gmail.com.

It brings about an overall improvement in computer-based instruction because it allows educators to engage in CAI with their students using these packages. The use of computers refers to being a tool to enhance and promote education. The topics are covered through examples, exercises, simulations, and problems to test the student's comprehension[7]. Computer-assisted instruction had been in existence for 40 years before the personal computer arrived; there are various job training topics that CAI could open up for advancement and new abilities in incorporating technology available[20]. Initially, CAI was used to improve traditional teaching modalities; as time went on, however, educational pressure imposed on all levels impelled time-efficient, high-quality teaching modalities. Several research studies on microcomputer applications have looked at their impact on learning outcomes and student attitudes. Most researchers agree that CAI tends to create a more positive effect on student attitudes than instruction. At this point, CAI probably has a positive impact on students[4].

2. Need of CAI Education Organization

Because we value knowledge, we have adopted Bloom's Taxonomy. Since the Taxonomy of Knowledge & Learning Outcomes was published in the late 1960s, it has created digital tasks, evaluated apps, and determined their success. A group of cognitive psychologists, educational researchers, educational theory developers, and assessment specialists put out a report in 2001. They argued that a single intelligence is needed to excel in this industry[5]. Bloom's Taxonomy can be helpful to students as well as teachers. The Bloom Taxonomy can make class assignments, identify and structure lessons, and devise curricula. Educational organization research will often have to do all three things. Gather new data from primary or first-hand sources, use existing data in an untested way or new model, or fulfill an original goal[6].

On the whole, no one wants to know how things like education and colleges work, except students and employees who need to follow the rules of those who don't know about configuration and fees. The essential requirement is the amount of flow in the department of servers[19]. For colleges to properly utilize Creative Applications of Information is for instructional purposes, they must become aware of university functionality. Many researchers and reviewers found that computer-based learning alone did not show an appreciable effect on achievement. This paper focuses on implementation needs and their advantages over traditional systems[8].

3. Specified areas of CAI

Tutorial work is of information and practice of various application methods, including drill and practice and games. The discovery approach offers a massive amount of relevant information about a specific course or subject and challenges students to probe, compare, and speculate on that data[9]. It's cheaper and safer to use simulation software to calculate and predict events than to engage in real-life activities. The problem-solving approach boosts a child's specific problem-solving abilities when presented with new skills; students benefit from repetition. Still, for them to be mastered, repeated practice is essential [10]. CAI is associated with other benefits, including a greater sense of self-control, improved attendance, motivation, and higher

performance on tasks and teamwork. Conventional instruction combined with CAI increases academic achievement over the use of either of the two alone[11]. However, many people have not been able to find any evidence supporting the use of conventional teaching alone, or CAI, as a strategy for improving students' creative development. There is evidence that computer-based learning improves performance better than traditional approaches. Conventional approaches lead to more favorable attitudes to computers, instruction, student learning, and self-learning [1].

4. Advantages of CAI

Several advantages of integrating CAI over conventional education system are;
a) Allows learning new skills and refining existing ones.
b) The students are excited and interested in real-time feedback.
c) Audiovisual, text, images, or a combination can be used in the program[12].
d) Data mining (specifically machine learning) optimizes information presented systematically and comprehensively, making it accessible and entertaining.
e) It looks for ways to help our teachers and authors grow their capabilities[15].
f) Interactive illustrations let you extract more visuals from a book than text alone.
g) Feedback and progress reporting facilitated with the use of computer technologies. Thus, the learner has the opportunity to work on his or her weaknesses over time
h) Your computer will patiently and calmly allow for an incorrect response[13].
i) To avoid student embarrassment, you can alter the style of instruction and the pace of delivery according to each individual's physical abilities.
j) As a novel technique, the novel process helps with enrichment because it adds variety to the program[14].
k) Many CAI training programs offer only a handful of options regarding how the material is presented and what learning strategies are included. Other programs are called learner-controlled because they are flexible enough to respond to students' needs and interests [16].

5. How it works

Using computers in this context will help a school better understand its instructional process. The use of computers to carry out operations in which creative thought is combined with instructional design. Interactivity is incorporated into the use of software and hardware programs. This software is highly user-friendly; everything is generally applicable in this educational context. When you have started the search, it will provide details on entering and applying [17]. The CAI is intended to assist staff in compiling information about team and work. Besides, it helps keep track of where they work and the number of teachers. This software tells you where admission forms are and what makes them available, so it's excellent for procrastinators[18]. Paying fees is beneficial in itself, but also, it pays for additional benefits as well. Choose the best for you. You can pay your fees in various locations or pick the one that works for you. It's also beneficial to include the exact geographic location of the server [21].
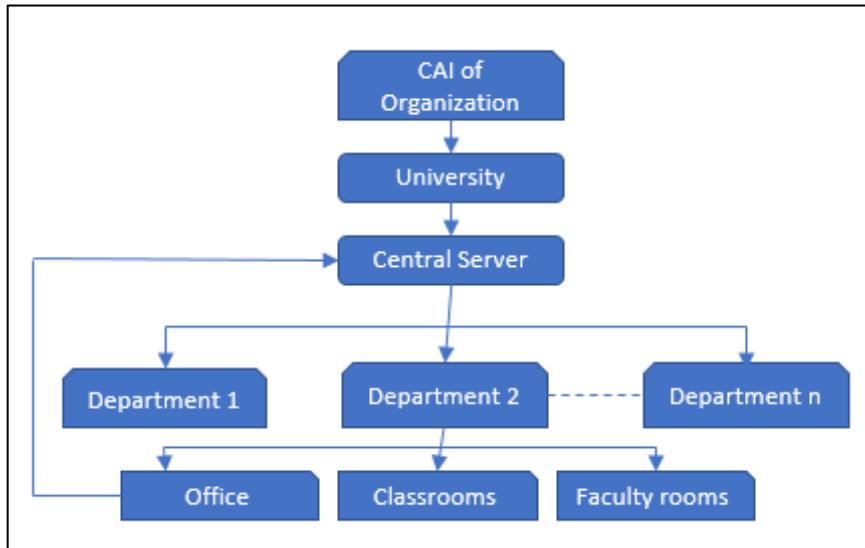
Figure 1. Implementation structure of CAI

Figure 1 shows the implementation structure of CAI by including various components. CAI of the organization is the root component. It transfers the data to the leaf components. University/Institute/College is another component of this system as it receives the information from the parent organization. Received information will be stored on the central server of the university/Institute/College, and it will be shared with all the sub-departments. Again departments will share their data to sub-sections related to the department. The advantage of this hierarchy is, the same data will be shared with all the stakeholders. In case of modification of data, revised data will be shared with them. This technique is used to minimize the resources with its maximum utilization.

6.   Recommendations

a) CAI could be effective as a teaching strategy.
b) Initially, we can only use CAI on a few subjects, and even so, the implementation of CAI can be carried out regardless of the resources'
c) Language and culture differences between teacher and student can be lessened by using class-based assignments and assessments.
d) it may be used as an extension of classroom instruction for particular students
e) Cognitive ability can be enhanced through CI.

7.   Conclusion

Researchers in several fields have tried to determine whether this hypothesis is true and have reached different conclusions. A majority of studies indicate CAI is beneficial to younger students. Research shows that CAI can benefit students of all three grade levels. The effect is less significant from elementary to secondary school to post-

secondary levels. The findings show that, for suitable careers, colleges need to devote a broad focus on educating students. While helping them find opportunities and provide guidance and assistance with career advancement, there needs to be regular and coordinated staff and programs specific to each field of study. It has collected accurate, intelligent data that demonstrated students' understanding and performance. Computer-Assisted instruction has become an effective and efficient means of delivering information. Worldwide research conducted on the Use of the CAI indicates that it is successful in most operating systems and computer platforms. If an educational organization is supposed to be involved in the educational process, it can be very time-consuming. However, it is available to everyone. It is well understood, as well, because of how straightforward it is. Using this CAI, you can get students to follow the rules.

References

[1] *Computer-Assisted instruction -> Advantages—Dphilwork*. (n.d.). Retrieved March 16, 2021, from https://sites.google.com/site/phdworkedu/computer-assisted-instruction-module-6/advantages-of-cai-method

[2] *Computer-Assisted Instruction (CAI)—WikiEducator*. (n.d.). Retrieved March 16, 2021, from https://wikieducator.org/Computer_Assisted_Instruction_(CAI)

[3] Cotton, K. (2008). Computer-Assisted Instruction. In C. R. Reynolds & E. Fletcher-Janzen (Eds.), *Encyclopedia of Special Education* (p. speced0481). John Wiley & Sons, Inc.

[4] Dede, C. (1986). A review and synthesis of recent research in intelligent computer-assisted instruction. *International Journal of Man-Machine Studies*, *24*(4), 329–353.

[5] Dede, Y., Doğan, M. F., &Gürbüz, R. (2018). The Role of Computer-Assisted Instruction in the Teaching of Probability. *Hacettepe University Journal of Education*, 1–18.

[6] Guo, H. (2018). Application of a Computer-Assisted Instruction System Based on Constructivism. *International Journal of Emerging Technologies in Learning*, *13*(4), 33–44.

[7] Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore

[8] Dongale, T. D., Kharade, K. G., Naik, G. M., & Kamat, R. K. (2017). Artificial Neural Network Modeling of NixMnxOx based Thermistor for Predicative Synthesis and Characterization. Journal of Nano And Electronic Physics, 03042(1)-03042(4).

[9] Katkar, S. V., Kharade, K. G., & Kamat, R. K. (2017). Role of WhatsApp in Farmer's Welfare. National Conference on Rural Livelihoods Models for Sustainable Development organized by Yashvantrao Chavan School of Rural Development, Shivaji University, Kolhapur.

[10] Katkar, S. V., Kharade, K. G., Kharade, S. K., & Kamat, R. K. (2020). An Intelligent Way of Modeling and Simulation of WO3 for Supercapacitor. In N. Thapa, Recent Studies in Mathematics and Computer Science (pp. 109-117). Hooghly, West Bengal, India. Book Publisher International.

[11] Kharade, K. G., Kamat, R. K., & Kharade, S. K. (2017). Review On Microsoft Windows Azure Technology. National conference on "Innovative Trends in Electronics and Allied Technology (ITEAT-2017), (p. 33).

[12] Kharade, K. G., Kamat, R. K., & Kharade, S. K. (2017). Utilization of Copy Logger Software Package to Protect Sensitive Data from Malicious Individuals. National conference on "Innovative Trends in Electronics and Allied Technology (ITEAT-2017)" organized by Yashvantrao Chavan Institute of Science, Satara, (p. 19).

[13] Kharade, K. G., Kamat, R. K., Kharade, S. K., &Katkar, S. V. (2019). Automation of Paper Setting Process to Improve Effectiveness of The Examination System of The University. Journal of Emerging Technologies and Innovative Research, 490-493.

[14] Kharade, K. G., Kamat, R. K., Mudholkar, R. R., & Kharade, S. K. (2018). Removable Drive Blocker Application for Virus Detection. International Journal of Research Culture Society, 233-234.

[15] Kharade, K. G., Kharade, S. K., &Kumbhar, V. S. (2018). Impact of Digital India on Various Sectors. Indian Journal of Innovation in Management and Excellence In Research (IJIMER), 37-40.

[16] Kharade, K. G., Kharade, S. K., &Oza, K. S. (2018). Security Threats In E-commerce. Indian Journal of Innovation In Management And Excellence In Research (IJIMER), 14-17.

[17] Kulik, C.-L. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction. An updated analysis. *Computers in Human Behavior*, *7*(1–2), 75–94. Ramani, P., &Patadia, H. (2012). The Effectiveness of Computer-Assisted Instruction in Teaching Arithmetic. *International Journal of Scientific and Research Publications*, *2*(11), 1–9.

[18] Rosali, L. J. D. (2020). Effect of Computer-Assisted Instruction (CAI) on the Academic Achievement in Secondary Physics. *OALib*, *07*(05), 1–11.

[19] Tomesko, J., Touger-Decker, R., Dreker, M., Zelig, R., & Parrott, J. S. (2017). The Effectiveness of Computer-Assisted Instruction to Teach Physical Examination to Students and Trainees in the Health Sciences Professions. A Systematic Review and Meta-Analysis. *Journal of Medical Education and Curricular Development*, *4*, 238212051772042.

[20] Potharaju, S. P. (2018). An Unsupervised Approach For Selection of Candidate Feature Set Using Filter Based Techniques. Gazi University Journal of Science, 31(3), 789-799..

[21] Patil, B. P., Kharade, K. G., &Kamat, R. K. (2020). Investigation on Data Security Threats & Solutions. International Journal of Innovative Science and Research Technology, 5(1), 79–83.

[22] Rao, G., Kumari, K., Shankar, D., & Kharade, K. G. (2021). A comparative study of augmented reality-based head-worn display devices. Materials Today. Proceedings. 1-9.

A Systematic Review of Blockchain Technology and Its Applications

Bharati H Naikwadi [a] [1], K. G. Kharade [b], S. Yuvaraj [c], K. Vengatesan [d]

[a] *BLDEA's A.S.Patil College of Commerce, Vijayapur, KA, India*
[b] *Department of Computer Science, Shivaji University, Kolhapur, MH, India*
[c]*Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Kattankulathur, TN, India*
[d] *Computer Engineering, Sanjivani College of Engineering, Kopargaon, MH, India*

Abstract. Blockchain technology could radically transform our business environment in the decades to come. It has the power to redefine our views of business processes and may even impact our overall economy. The works comprehensively document the implementation of blockchain applications in various sectors. Our goal is to determine what state blockchain technology is in and what kinds of applications it has. A large majority of the current blockchain-related research is dedicated to crypto currencies, such as Bitcoin. Only a small percentage of work is dedicated to exploring the wide range of potential blockchain technology applications. Blockchain technology has multiple applications across multiple industries. This study attempts to determine the opportunities and threats presented by Blockchain Technology for current or future applications. The number of published studies that were studied carefully and critically and added to the Block chain's body of knowledge was large.

Keywords. Applications of Blockchain, Blockchain, Electronic Health Records, IoT, Security

1. Introduction

Blockchains disrupted traditional business processes by introducing decentralized applications and transactions, which had previously relied on centralized architectures or trusted third parties to verify them. Blockchain technology provides characteristics like transparency, robustness, suitability, and security, all of which are inherent to the system [3]. Despite the revolutionary possibilities for building the future Internet system via blockchain technology, many technical challenges are currently confronted. The first issue is scalability. At present, the maximum block size in Bitcoin is one megabyte, and blocks are mined about every ten minutes. The system then limits the number of transactions the network can perform per second to seven, unable to deal with high-frequency trading [23]. In a public ledger, each transaction is verified by the agreement of most participants in the system. In other words, once data is entered, it can never be erased. Every transaction ever made is included in the Blockchain's immutable record [17]. An easy analogy to use is to say that it is easier to steal a cookie from a cookie jar, where the cookie jar is hidden away in a secluded location rather than to steal a cookie from a cookie jar, where there are numerous people around to witness the theft [5].

Distributed ledger technology is commonly referred to as blockchain technology. The Blockchain is the underlying technology that makes bitcoin possible, devised by an anonymous individual known as Satoshi Nakamoto in 2008. It is best to understand how the bitcoin mechanism works before learning about Blockchain. Bitcoin's digital currency is a peer-to-peer, distributed, and decentralized payment system [1]. Like distributed ledger technology (DLT), Blockchain does not require a central authority to manage operations. No one knows who controls it, except for those directly involved. Blockchain data is organized in terms of blocks, which is why it is called a blockchain. This complex network of interlocked blocks is protected with encryption. The data added to the Blockchain is immutable, i.e., once added, it cannot be altered or deleted. It is one of the distinguishing features of this database compared to conventional databases. Every transaction exists in history because you cannot remove or alter even a single block of data [2].

2. Literature Review

(Xu et.al ,2019) identified that, many people see blockchain as a core technology that will disrupt an industry. A great deal of work has been done in the advancement of blockchain, but the amount of investigation remains limited. in order to thoroughly research block chain's current academic research, it incorporates the topics of business and economics [24]. (Junyao Wang et.al, 2019) focused on Blockchain development and application are being explored in both academic and commercial environments. Blockchain is already proving itself in various fields, particularly in the financial, sales, medical, and other fields [25]. (Kitsantas, 2019) revealed that, at the beginning of the blockchain era, where the technology will lead to significant changes in business environment, and that will have a profound impact on the decades to come. It can have a radical impact on the way we understand business processes, and can fundamentally transform our economy. The main purpose of blockchain technology is to provide transparency, data security, and integrity. This is because blockchain technology is impervious to tampering or fabrication [26]. (Liang Zhou, 2020) stated that, blockchain has recently emerged as a distributed and fast-evolving technology, and this has led to it influencing crypto currency and e-commerce, which have drawn the interests of governments, businesses, and research institutions. This suggests that institutions should plan their research effectively in order to get a sense of the current state of blockchain research [27].

3. Research Methodology

The present study employs a systematic literature review. A systematic literature review aims to find previous research relevant to the topic and then outline research gaps and future research paths. It also focuses on the benefits and applications of Blockchain and the areas where Blockchain has been implemented. During this research we have worked on areas where block chain can be implemented. More than 100 papers have been studied during this paper. It is observed that, block chain technology has many advantages but lack of awareness and government policies affecting on the use of this technology. Health care is the most common area where this technology is being implemented. With the rapid growth in all the industry, block chain

is being used by many industries. We have identified few industries where it is being used.

4. Benefits of Blockchain

Block chain technology has various benefits; some of the benefits are;

a) **Better Transparency:** With the advancement of blockchain technology, data and money transactions are becoming easier. All participants share the document in the networks instead of being distributed to each of them. Even if the participants wanted to change the documentation, they would not do so without changing the documentation blocks that came before. It is because, in the context of heavy processes, blockchain technology has proven to be more accurate, consistent, and transparent [21]

b) **Enhanced Security:** Blockchain technology is superior to other platforms or record-keeping systems when it comes to security. All transactions, even the smallest ones, need to be approved through consensus. The process of generating and applying a cryptographic hash function, then making that hash transaction-based linked to the prior transaction, occurs every time a transaction is performed [13]. Network security is also strengthened because each node keeps a record of all the transactions. Thus, malicious actors would be unable to alter transactions since others in the network will reject their attempts to do so [8].

c) **Reduced Costs:** For the vast majority of businesses, cost reduction is of prime importance. Because Blockchain eliminates the need for third parties and mediators, you don't need to trust your trading partner. A single, immutable version will be available to everyone, so you will not have to review much documentation [6]

d) **Data Ownership**: Blockchain provides this vital benefit. Blockchain is widely distributed instead of being controlled from a single point. Thus, even if a third party or a government were to attempt to control their plan and your process, they could not succeed. Blockchain is simply a more descriptive term for a fair distribution system [18].

e) **Accurate Traceability:** The buyer is already familiar with how difficult it is to trace a product back to its source because you deal with traded products through a complex supply chain. You end up with an audit trail showing where each asset started its journey and where it stopped on the way. The transaction data from this historic trade could be used to verify the authenticity of support and help prevent [9].

f) **Improved Speed and Highly Efficient:** Trading anything is a time-consuming process prone to human error and often requires third-party mediation, regardless of whether it's done through paper-heavy or digital means. The main benefit of using a single ledger shared by participants is that there is no need to reconcile multiple ledgers. As a result, you don't end up with as much clutter [10].

5. Advantages of Blockchain Technology

Blockchain technology has many advantages over traditional technologies.

- ✓ You can trust your business process's quality because it is protected with a high level of security thanks to Blockchain.
- ✓ Transactions done are transparent, which allows them to be tracked easily
- ✓ Enterprise blockchain technology gives organizations the ability to control accessibility using a wide range of accessibility levels [11].
- ✓ When organizations use Blockchain, they can perform transactions up to five times faster.
- ✓ The threats to your business will be more significantly reduced because of hacking.
- ✓ Blockchain has eliminated the need to pay for centralized entities or intermediaries' services, as the network is decentralized
- ✓ Automated account reconciliation is possible [12].

6. Applications of Blockchain for Businesses

Typical applications of the Blockchain technology are;

a) **Blockchain for Communications:** The connectivity, infrastructures, resources, end devices, and applications in communications and networking systems are becoming much more complex and heterogeneous due to the rapid advancement of information and communication technologies [15]. Additionally, there is a significant volume of data and a large number of end devices that will bring substantial security, privacy, services, and network management complications. The combined consideration of Blockchain and machine learning (ML) may lead to significant benefits and have sparked a great deal of interest in both academia and industry [16]

b) **Blockchain in Music Industry:** A longstanding problem has increased in urgency: a central database that catalogs music metadata does not exist, making it challenging to find the owners or those with rights and ensure they get paid. A decentralized ledger, an unalterable, secure, efficient, and cost-effective rights management system, could be built in the music industry; GRD, the centralized Global Repertoire Database, has failed so far [4].

c) **Blockchain in IoT:** For an IoT solution to succeed, it must possess reliability, scalability, and long-term durability. And we know IoT is useful. In contrast to Blockchain, Blockchain is much more experimental and episodic. It has not

been scientifically proven. Blockchain flares up but then quickly levels off. Meanwhile, IoT solutions are continuously put into use [19]

d) **Blockchain Empowered Cloud Gaming System:** Cloud gaming offers cloud-rendered game scenes and real-time video footage delivered to end devices via the Internet. It allows players to access game services anytime and anywhere, even if their hardware capabilities are limited. Still, since the cloud gaming market is a commercial service, the latest payment models for cloud gaming are in the early development stages [22].

e) **Blockchain in Healthcare:** This idea comes out of the desire for healthcare security and interoperability. Much medical data is collected and transmitted daily thanks to IoT and the increasing abundance of health devices and mobile healthcare applications. This data traffic requires privacy and security management. With blockchain technology, it is possible to secure the recording and sharing of medical records while assuring each patient's data privacy. Furthermore, Blockchain has the advantage of making healthcare management more straightforward, but it also presents challenges that should be anticipated [14].

f) **Blockchain In Energy Sector:** For governments, one of the most critical aspects of energy distribution, allocation, and production has always been present. In the absence of a well-thought-out energy strategy, no government can provide necessary growth to its economy. Similarly, private players also contribute to the Blockchain, both materially and immaterially.

g) **Blockchain in Banking:** A Corda-based implementation is created through a merger of financial institutions and adheres to industry standards. Corda is a blockchain-based ledger platform, which has been appropriated. Hyper ledger Fabric blockchain platforms are widely accepted as secure and private blockchains. Banks find security assurance helpful. Ethereum is a type of platform that enables people to build decentralized applications [20].

h) **Blockchain in Real Estate:** For the past few years, the real estate sector's conditions have been subpar. Since house prices have increased and it has become more challenging to buy a property that is one reason. With the introduction of Blockchain, the real estate sector is looking at its operations in a whole new light. Blockchain will allow tokens to be issued in connection with particular actions. In other words, properties can be leased out using pre-defined programming. Also, tokens enable businesses to add any necessary business logic, such as combat fraud [7].

7. Conclusion

This paper demonstrated how blockchain technology is being applied in various industries, including finance and healthcare. While the first blockchain research was carried out in 2012, more research was conducted on the subject in the years since. This paper studies how blockchain technology can encourage open science, including reviewing the latest developments and identifying and describing the science's advantages and challenges. To test whether the requirements for an open scientific

ecosystem fit with blockchain technology properties, we first identified them and compared them to see if they matched. Finally, we proved that the technology was reliable and appropriate for open science.

References

[1]   Alla, S., &Soltanisehat, L. (2018, May). Blockchain Technology in Electronic Healthcare System. https://www.researchgate.net/publication/325542378_Blockchain_Technology_in_Electronic_Healthcare_System

[2]   Anwar, H. (2019, January 6). Blockchain vs. Distributed Ledger Technology: A Detailed Guide. https://101blockchains.com/blockchain-vs-distributed-ledger-technology/#prettyPhoto

[3]   Casino, F., Dasaklis, T. K., &Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification, and open issues. Telematics and Informatics, 36, 55–81. https://doi.org/10.1016/j.tele.2018.11.006

[4]   Celebucki, P., Atkins, T., Gutierrez, J., McKone, M., & Steinberg, A. (n.d.). Blockchain + the Music Industry. Retrieved February 23, 2021, from https://onprem.com/wp-content/uploads/2020/08/wp_blockchain-1.pdf

[5]   Crosby, M., Nachiappan, Pradhan, P., Varma, S., &Kalyanaraman, V. (2015). Sutardja Center for Entrepreneurship & Technology Technical Report (p. 35). chrome-extension://oemmndcbldboiebfnladdacbdfmadadm/https://scet.berkeley.edu/wp-content/uploads/BlockchainPaper.pdf

[6]   Dongale, T. D., Kharade, K. G., Naik, G. M., & Kamat, R. K. (2017). Artificial Neural Network Modeling of NixMnxOx based Thermistor for Predicative Synthesis and Characterization. Journal of Nano And Electronic Physics, 03042(1)-03042(4).

[7]   Geroni, D. (2021, January 27). Top 5 Benefits of Blockchain Technology. https://101blockchains.com/benefits-of-blockchain-technology/

[8]   Hooper, M. (2018, February 22). Top five blockchain benefits transforming your industry—Blockchain Pulse: IBM Blockchain Blog. https://www.ibm.com/blogs/blockchain/2018/02/top-five-blockchain-benefits-transforming-your-industry/

[9]   Kharade, K. G., Kharade, S. K., & Kumbhar, V. S. (2018). Impact of Digital India on Various Sectors. Indian Journal of Innovation in Management and Excellence In Research (IJIMER), 37-40.

[10] Kharade, K. G., Kharade, S. K., & Oza, K. S. (2018). Security Threats In E-commerce. Indian Journal of Innovation In Management And Excellence In Research (IJIMER), 14-17.

[11] Kharade, S. K., Kamat, R. K., & Kharade, K. G. (2019). Simulation of Dye Synthesized Solar Cell using Artificial Neural Network. International Journal of Engineering and Advanced Technology (IJEAT), 1316-1322.

[12] Kharade, K. G., Kamat, R. K., Kharade, S. K., & Katkar, S. V. (2019). Automation of Paper Setting Process to Improve Effectiveness of The Examination System of The University. Journal of Emerging Technologies and Innovative Research, 490-493.

[13] Kharade, K. G., Kumbhar, V. S., & Kharade, S. K. (2017). A Comparative Study of Traditional Server and Azure Server. Journal of Advances in Science and Technology (JAST), 329 – 331.

[14] Kharade, K. G., Mudholakr, R. R., Kamat, R. K., & Kharade, S. K. (2019). Perovskite Solar Cell Simulation using Artificial Neural Network. International Journal of Emerging Technologies and Innovative Research, 336-340.

[15] Kharade, K. G., Mudholkar, R. R., Kamat, R. K., & Kharade, S. K. (2019). Artificial Neural Network Modeling of CZTS Solar Cell For Predicative Synthesis And Characterization. International Journal of Research And Analytical Reviews, 997-1006.

[16] Potharaju, S. P., & Sreedevi, M. (2017). A Novel Clustering Based Candidate Feature Selection Framework Using Correlation Coefficient for Improving Classification Performance. Journal of Engineering Science & Technology Review, 10(6).

[17] Patil, B. P., Kharade, K. G., & Kamat, R. K. (2020). Investigation on Data Security Threats & Solutions. International Journal of Innovative Science and Research Technology, 5(1), 79–83.

[18] Rao, G. N., Kumari, K. A., Shankar, D. R., & Kharade, K. G. (2021). A comparative study of augmented reality-based head-worn display devices. Materials Today: Proceedings.

[19] Schaefer, J. (2019, August 2). Blockchain Applications in IoT: Good and Bad | IoT For All. https://www.iotforall.com/blockchain-applications-in-iot

[20] Sharma, D. (2020). Application of Blockchain in an Indian Banking Sector. Global Scientific Journal, 8(5), 526–530.

[21] Techfunnel. (2020, May 13). 14 Amazing Benefits of Blockchain Technology | Techfunnel Guide. https://www.techfunnel.com/information-technology/benefits-blockchain-technology/

[22] Zhao, J., Chi, Y., Wang, Z., Leung, V. C. M., & Cai, W. (2020). CloudArcade: A Blockchain Empowered Cloud Gaming System. Proceedings of the 2nd ACM International Symposium on Blockchain and Secure Critical Infrastructure, 31–40. https://doi.org/10.1145/3384943.3409420

[23] Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017). An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. 2017 IEEE International Congress on Big Data (BigData Congress), 557–564. https://doi.org/10.1109/BigDataCongress.2017.85.

[24] Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.

[25] Wang, J., Wang, S., Guo, J., Du, Y., Cheng, S., & Li, X. (2019). A Summary of Research on Blockchain in the Field of Intellectual Property. *Procedia Computer Science*, *147*, 191–197. https://doi.org/10.1016/j.procs.2019.01.220

[26] Kitsantas, Thomas & Vazakidis, Athanasios & Chytis, Evangelos. (2019). A Review of Blockchain Technology and Its Applications in the Business Environment.

[27] Zhou, L., Zhang, L., Zhao, Y., Zheng, R., & Song, K. (2020). A scientometric review of blockchain research. *Information Systems and E-Business Management*. https://doi.org/10.1007/s10257-020-00461-9

EyePhone Technology: A Smart Wearable Device

K. G. Kharade[a] [1], S.K.Kharade[b], S. R. Ghatage[c], V.S.Kumbhar[a],
T. Nandana Ravishankar[d], K.Vengatesan[e]

[a]Department of Computer Science, Shivaji University, Kolhapur, MH
[b]Department of Mathematics, Shivaji University, Kolhapur, MH
[c]Gopal Krishna Gokhale College, Kolhapur, MH
[d]Department of Computer Science & Engineering, Vel Tech High Tech, Avadi, Chennai, TN
[e]Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon, MH

Abstract. In the past few decades, wearable sensors and devices have evolved into central technologies that have significantly impacted next-generation healthcare solutions in the previous decade. It is an age of incredibly cut-throat competition, even the youth. This technology is Hand-free, eyephone-operated telephone technology. It measures where the user's eyes are located on the phone's display employing a camera attached to the device. The typical work the eye bones assigned to eye tracking, eye blinking, detecting, etc., by way of eye-tracking, users can take care of their email, calendar, phone, etc. The leading technologies in use are a Human-Computer Interface (HCI) and a Human Phone Interface (HPI). It helps those with disabilities greatly. The IT Index is an expanded form of HCI. The use of mobile devices like smartphones and tablets is on the rise, and, to some extent, it could displace the use of desktops and laptops. Human-computer interaction is interested in the interaction between people and the computer system and between software and applications. Our focus is on this revolutionary new type of technology called the eye-phone.

Keywords. Eye Phone, E-Healthcare, HCI, HPI, Mobile Sensing Systems, Mobile Phones

1. Introduction

A new generation of smartphones has revolutionized touch screens, gyroscopic, 3D cameras, and other technology advancements. The software change and improved hardware performance enable a new way of employing these devices and techniques. Some recent research projects show how communication with mobile devices enhanced the creative process. An important avenue of communication has been opening up to the blind or visually impaired [2]. With the Eye Phone, you can tweet and make phone calls. The latest development of which is the widespread use of touch screens. With this touch screen, it's much easier to interact with mobile phones. It can locate and map the user's eyes on the screen, using a wink to establish the association with any application. At no time does the user have to be touching the phone or the phone display to use the device. With constrained form factors such as smartphones, researchers and vendors constantly look for new ways to cut down on the amount of

---

[1] K. G. Kharade, Department of Computer Science, Shivaji University, Kolhapur, India
Email: kgk_csd@unishivaji.ac.in

effort required to access applications [3]. The user can use it as a phone, and it functions as a surveillance camera as well.

The current EyePhone system can follow a user's eye, and output images can be controlled using the phone's built-in camera [4]. The HPI can be used in a hands-free manner, as it is used in facial recognition.

Additionally, we believe Eyephone technology is a better alternative to voice activation systems in quiet locations. All you need is a front camera to use the phone is one of these. Today's smartphones have a front-facing camera; expect to see even more on future models [18]. Human-computer Interaction (HCI) is a multidisciplinary field of study focusing on the design of computer technology and, in particular, the interaction between humans (the users) and computers [5]. While initially concerned with computers, HCI has expanded to cover almost all forms of information technology design. An HCI in which users can control their phones with their eyes or faces while exercising and developing a proof-of-concept prototype is done using an Android phone fitted with an EEG headset and an eye tracker [19]. Each functional module can be individually launched and switched; it is possible to achieve a "no hands" environment. To show that minds and intentions on mobile devices can be accurately controlled, improve our society's lives and health. Although this paper demonstrates several laboratory problems, it lays the groundwork for further studies on HCI in the outside world [1].

2. Literature Review

(Sannidhan, 2018) identified that there are numerous academic and non-profit uses for computer vision research. Multiple technologies are currently available on the market that keeps track of drivers' minds and helps prevent similar incidents from occurring in daily life. Increasing road traffic has brought about a rising number of road accidents every day [20]. (Saba, 2017) found, how to ease human-phone interaction (HPI). From touch-sensitive screens to voice activation systems to EyePhones that incorporate eye-tracking software and algorithms, Biometric Goggles that utilizes eye tracking and biometric algorithms will be remarkable and significant to numerous customers. In noisy environments, the overall performance of voice control systems is lessened [4]. Very few researches have been carried out on this topic. Implementation of this technology is costly and can not be affordable by regular users.

3. Proposed Method

During this research, we have gone through various papers in this field. EyePhone technology has enormous advantages and can be integrated with many areas. Such areas cover fields like engineering and the medical domain. We can incorporate EyePhone technology to provide value addition to the existing applications.

4. Hardware Challenges

As opposed to HCI applications, any HPI implementation should not rely on any external hardware. Asking people to carry or wear additional hardware to use their phones might reduce the technology's penetration. Moreover, state-of-the-art HCI hardware such as glass-mounted cameras or dedicated helmets is not yet small enough to be conformably worn for long periods by people [6]. Any HPI application should

rely as much as possible on just the phone's onboard sensors. Over the last decade, non-invasive electroencephalogram (EEG) BCIs have greatly improved. But here are some challenges that remain, like the non-stationary and low signal-to-noise ratio of the EEG. It is imperative to consider optimizing both the individual parts of BCIs and their interrelationship to achieve the greatest possible bandwidth. In other words, using concepts from human-computer interaction (HCI) methods in HCI and neuroscience should be combined [19]. One of the primary stumbling blocks in broader adoption is the lack of friendly interfaces and GUIs for patients and medical staff. Heterogeneous sensors can also mean multiple sensor types, including wearable, smartphones, and IoT sensors that allow monitoring to be quick and patient-friendly [17].

5. Eye Phone Design

The eye phone algorithmic design breaks down into the following pipeline phases:
1. Eye detection phase
2. Open eye template creation phase
3. Eye-tracking phase
4. Blink detection phase
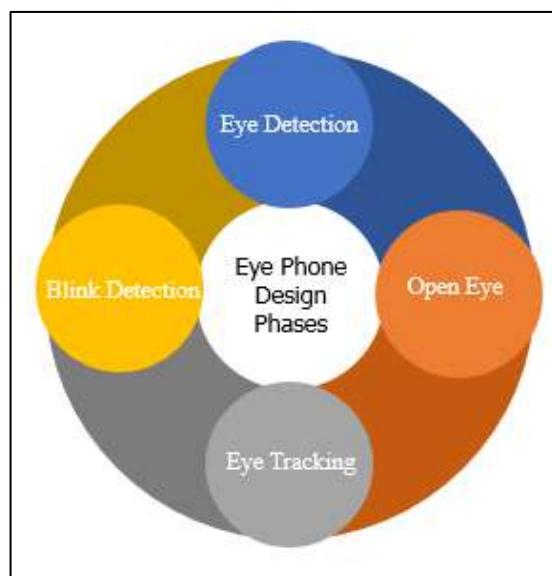
Fig. 1: Eye Phone Design Phases

a) **Eye Detection Phase:** By applying a motion analysis technique that operates on consecutive frames, this phase consists of finding the contour(outline) of the eyes. The left and right eye contours identify the eye pair. While the original algorithm identifies the eye pair with almost no error when running on a desktop computer with a fixed camera [7].

b) **Open Eye Template Creation:** While the author is adopting an online open eye template creation by extracting the template every time the eye pair is lost, the eye phone does not rely on the same strategy. Compared to a desktop

machine, the reduced computation speed and the restricted battery requirements imposed by the N810 dictate a different approach. Instead, an eye phone creates a user's open eye template once at the beginning when a person uses a system for the first time using the eye detection algorithm described above [8].

c) **Eye Tracking:** The eye-tracking algorithm is based on template matching. The template matching function calculates a correlations core between the open eye template created the first time the application uses and a search window. The search window is limited to twice the box's size, enclosing the eye to reduce the time template matching function's computation time and save resources[9].

d) **Blink Detection:** To detect blinks, we apply the thresholding (beginning) technique for the normalized correlation coefficient returned by the template matching function as suggested.

6. Applications of Eye Phone

EyePhone technology can be integrated with many technologies. Some of the applications of this technology are;

a) It is possible to use one's eyes to detect if one is sleepy and steer clear of the latter.

b) The EyePhone may be able to gauge drowsiness and inattention levels in drivers. While the industry is creating vehicles that monitor drivers' sleep and alertness, the EyePhone can apply to nearly any vehicle type.

c) It is a paradoxical notion to consider virtual reality as part of the game in that it emulates reality while distancing people from it [10].

d) Consumers will buy a higher quality product when they think it is essential.

e) Brands, marketing, offering product placement as a service

f) Due to physical infirmity, especially partial blindness, however, the phone is most often used to guide the eye. can be used while you are walking

g) If the problem lies outside the vision, do an MRI or CT scan, allowing inside eye examination for various medical issues, dental, and surgical issues [11].

7. Eye-Phone Apps in Healthcare

Providing healthcare in most developing countries remains a significant problem for healthcare providers in remote areas. Here are some apps where this technology is available;

a) **Peek:** Portable Eye Examination Kit (Peek) is a community health care worker-focused mobile app designed to enable them to do their work anywhere. The application can scan the eye for various ailments, including diseases, visual acuity, visual field, and color vision [16]. Indicating Peek may detect different types of blindness, such as glaucoma, macular degeneration, macular and nerve tumors, diabetic retinopathy, and brain hemorrhage. With only basic training, the app can conduct a comprehensive analysis. Hence, it is inexpensive, fragile, and expensive and does not require a constant power supply. In addition, it is often portable and effortless to use [12].

b) **D-EYE**: It's a window into the human mind's world. It's an innovative concept in eye examinations that will make them available to the general public, team up with telemedicine, and increase healthcare access.

c) **Wearable EEG headset:** It is a design with a graphical window that receives continuous EEG data from the headset and motion tracking. It is also possible to implement a cursor to simulate movement on the screen and face recognition [13].

d) **Mindo:** Mindo monitors every change in the driver's mental activity as it happens. With Java as a mobile application for online analysis, the system processes the brain wave recordings to derive the level of drivers of vigilance. A neuro-backpack system incorporating electroencephalography (EEG) and transcranial electrical stimulation (TES) is now available. Independent component analysis (ICA), neuro-system architecture incorporating the FFT and support vector machine (SVM) [14].

e) **Tele rehabilitation Device:** a novel device intended for the requirements of patients with rheumatoid arthritis. It has several sensors that can be used either in a clinic for functional hand evaluation or by the patient at home for therapy, with a PC integration. In the second case, the device assesses the patient's performance as he follows the session guidance, using a GSM/GPRS-based connection to collect statistics for future analysis [18].

f) **Mobicare Cardio Monitoring System:** It is a predictive virtual instrument that consists of a cellular phone with real-time Q-T onset detection algorithms, Q on-time, T out-time (mobile ECG), a Web-enabled ECG sensor, a patient's database, and a web-rich server was recently introduced. Under this arrangement, MobiECGPRS+ sends the abnormally detected signal to hospitals to not burden the doctor [6].

g) **WBSN based e-healthcare:** The Wireless Body Sensor Network-based e-healthcare Systems is wearable, wireless communication, multi-sensor data fusion is utilized. The BSN continuously monitors important network parameters, such as physiological data security [15].

8. Conclusion

We presented the eye-phone technology and focused on using it in practice, and assessed its usefulness. For a mobile interface, the eyephone relies on eye-tracking and blinks. Our preliminary findings show that using the eyephone to develop mobile applications in a hands-free manner is exciting. The growth in e-health, driven by wireless and sensor technologies, which have accelerated e-healthcare. As e-health architecture has developed, it has become more lightweight, compact, and cheap, with smartphone's help. In this paper, we have focused on eye-phone technology, its design, and its application. It also focuses on how it is helpful in e-health care. Finally, it also focuses on its challenges before such technologies.

References

[1]   English, E., Hung, A., Kesten, E., Latulipe, D., & Jin, Z. (2013). EyePhone: A Mobile EOG-based Human-Computer Interface for Assistive Healthcare. 6th Annual International IEEE EMBS Conference on Neural Engineering (pp. 105-108). San Diego, California: IEEE.

[2] Miluzzo, E., Wang, T., & Campbell, A. T. (2010). EyePhone: Activating mobile phones with your eyes. Proceedings of the Second ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile Handhelds - MobiHeld '10, 15. https://doi.org/10.1145/1851322.1851328

[3] Schleicher, R., Westermann, T., & Reichmuth, R. (2014). Mobile Human–Computer Interaction. In S. Möller & A. Raake (Eds.), Quality of Experience (pp. 339–349). Springer International Publishing. https://doi.org/10.1007/978-3-319-02681-7_23

[4] SSM College of Engineering & Technology, Kashmir, India, & Paul, Z. M. (2017). EYE PHONE-HAND FREE MOBILE CONTROL: A REVIEW. International Journal of Advanced Research in Computer Science, 8(7), 52–54. https://doi.org/10.26483/ijarcs.v8i7.4226

[5] Understanding Mobile Human-Computer Interaction. (2005). Elsevier. https://doi.org/10.1016/B978-0-7506-6352-6.X5000-2

[6] Kher, R. K. (2016). Mobile and E-Healthcare: Recent Trends and Future Directions. Journal of Health & Medical Economics , 2-10.

[7] Dongale, T. D., Kharade, K. G., Naik, G. M., & Kamat, R. K. (2017). Artificial Neural Network Modeling of NixMnxOx based Thermistor for Predicative Synthesis and Characterization. Journal of Nano And Electronic Physics, 03042(1)-03042(4).

[8] Katkar, S. V., Kharade, K. G., & Kamat, R. K. (2017). Role of WhatsApp in Farmer's Welfare. National Conference on Rural Livelihoods Models for Sustainable Development organized by Yashvantrao Chavan School of Rural Development, Shivaji University, Kolhapur.

[9] Katkar, S. V., Kharade, K. G., Kharade, S. K., & Kamat, R. K. (2020). An Intelligent Way of Modeling and Simulation of WO3 for Supercapacitor. In N. Thapa, Recent Studies in Mathematics and Computer Science (pp. 109-117). Hooghly, West Bengal, India: Book Publisher International.

[10] Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.

[11] Kharade, K. G., Kamat, R. K., & Kharade, S. K. (2017). Utilization of Copy Logger Software Package to Protect Sensitive Data from Malicious Individuals. National conference on "Innovative Trends in Electronics and Allied Technology (ITEAT-2017)" organized by Yashvantrao Chavan Institute of Science, Satara, (p. 19).

[12] Kharade, K. G., Kamat, R. K., Kharade, S. K., & Katkar, S. V. (2019). Automation of Paper Setting Process to Improve Effectiveness of The Examination System of The University. Journal of Emerging Technologies and Innovative Research, 490-493.

[13] Kharade, K. G., Kamat, R. K., Mudholkar, R. R., & Kharade, S. K. (2018). Removable Drive Blocker Application for Virus Detection. International Journal of Research Culture Society, 233-234.

[14] Kharade, K. G., Kharade, S. K., & Kumbhar, V. S. (2018). Impact of Digital India on Various Sectors. Indian Journal of Innovation in Management and Excellence In Research (IJIMER), 37-40.

[15] Kharade, K. G., Kharade, S. K., & Oza, K. S. (2018). Security Threats In E-commerce. Indian Journal of Innovation in Management And Excellence In Research (IJIMER), 14-17.

[16] Patil, B. P., Kharade, K. G., &Kamat, R. K. (2020). Investigation on Data Security Threats & Solutions. International Journal of Innovative Science and Research Technology, 5(1), 79–83.

[17] Rao, G., Kumari, K., Shankar, D., & Kharade, K. G. (2021). A comparative study of augmented reality-based head-worn display devices. Materials Today: Proceedings., 1-9.

[18] Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.

[19] Scherer, R., Josef Faller, D. B., Friedrich, E. V., Pröll, M., Allison, B., & Müller-Putz, G. (2012). Brain–computer interfacing: more than the sum of its parts. Soft Computing, 317–331..

[20] M S, Sannidhan & Aithal, Sunil Kumar & Bhandary, Abhir. (2019). A Comprehensive Review On Various State Of Art Techniques For Eye Blink Detection.

Systematic Study of Video Mining with Its Applications

Mallappa G. Mendagudli[a] [1], K. G. Kharade[b], T. Nadana Ravishankar [c], K. Vengatesan[d]

[a]BLDEA's A.S.Patil College of Commerce, Vijayapur, Karnataka
[b]Department of Computer Science, Shivaji University, Kolhapur, Maharashtra
[c]Department of Computer Science & Engineering, Vel Tech High Tech, Avadi, Chennai
[d]Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon

Abstract. Effective methods for video indexing will be more valuable as digital video data continues to grow. It has been years since we've seen this level of new multimedia research. The content analysis aims to create high-level descriptions and annotations by treating language and facts as data. Data mining is a technique that seeks out previously unknown facts and patterns in large datasets. A video can include several different kinds of data, such as images, visuals, audio, text, and additional metadata. Thanks to its broad application in various disciplines, like security, education, medicine, research, sports, and entertainment, it is often used differently. Data mining aims to discover and articulate exciting patterns that are hidden in a lot of video footage. While video mining is still in its infancy, data mining is more mature. A considerable amount of research must be done to turn the mined video into usable content

Keywords. Event Detection, Frequent Pattern, Video Classification, Video Mining, Video Processing

1. Introduction

The massive growth in multimedia databases has been fueled by the ongoing improvements in multimedia acquisition and storage technology. As computing power, sensors, and storage improve, multimedia data development increases. Multimedia data mining is the process of discovering patterns from a large set of previously unknown data. Implicit knowledge extraction deals with extracting non-explicit patterns hidden in the multimedia data or the relationships between the multimedia data [14]. In the non-structural nature of multimedia data, multimedia data management is among the essential tasks. An immediate threat is handling all the multimedia data, such as multimedia text, images, audio, and video, with an unnecessarily complicated structure such as multimedia text, images, audio, and video. Currently, video is widely accepted as one of the most prominent multimedia regarding information and entertainment. Video mining enables efficient browsing and harvesting of video information while also making video retrieval possible in a gigantic data store or the cloud. Video data comes in many different varieties [2].

---

[1] K. G. Kharade, Department of Computer Science, Shivaji University, Kolhapur
Email: kgk_csd@unishivaji.ac.in

There are two main types of videos: scripted and non-scripted. Produced videos follow a set script or plan, and are edited, compiled, and disseminated to viewers. Fiction movies, news stories are perfect demonstrations of pre-produced videos. The footage in non-scripted videos does not occur through an act of will, nor do they always happen in a predictable setting, such as surveillance footage, sports footage, and business meetings [4]. Non-scripted videos deal with occurrences of interest and highlights. While video mining does not extract everything about the video, it can also discover various patterns related to video structure, events, activities, and much more. Many different video mining techniques have been explored to extract meaningful information from video data [18]. Finding information in a highly compressed feature space is still a difficult task. Association mining, pattern detection, and clustering are video classifications referred to the following should be kept in mind when mining video;

a) It should be possible to solve the problem by applying relatively simple mathematical techniques

b) The implementation should be unsupervised

c) There should be some room for flexibility

d) It should display essential and significant events
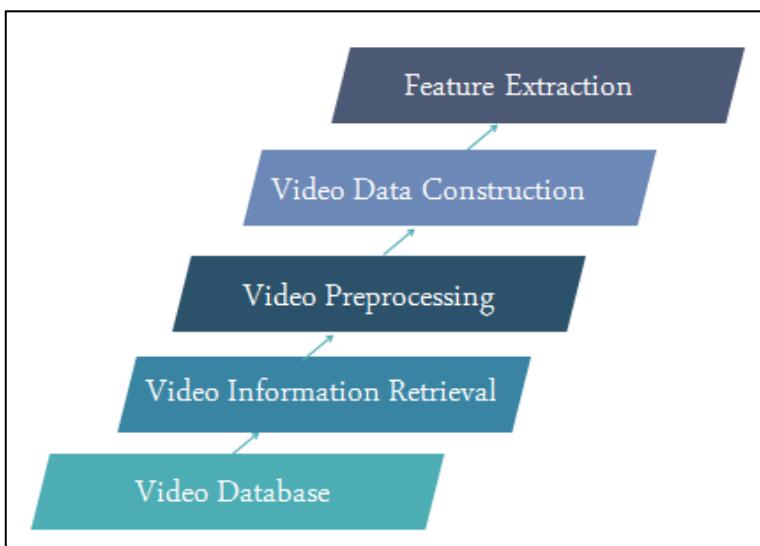
Fig. 1: Common Steps in Video Mining

2.   Framework

Video data mining applies the principles of implicit knowledge discovery when seeking to discover information that has been encoded into the data sequence. It's a difficult task to visualize patterns in a video frame sequence, because video contains both complex structured and unstructured data. When patterns are identifiable in the database, the discovery process incorporates discovering patterns [21]. Video is made up of individual images strung together sequentially, and images are referred to as frames. A single-camera has only one shooting sequence, called a shot. They utilize a shot-based strategy, while the object-based approach does not. This transaction-level analysis is done at the item level, and around each image is examined as well. Object-based representation makes content and indexing easier when it comes to search [5].

When doing object-based tracking, region-tracking and spatiotemporal segmentation are commonly employed. In the regular course of day-to-day life, we're interested in watching two kinds of videos: those containing structure and those that don't. When it comes to movies, videos, and news, these are all vehicles for carrying video content[17]. There is no video in the sports surveillance footage. Associated rule mining requires three steps: preliminary processing, association rule development, and database construction. To ensure that the video data is processed first, it should be first converted from its un-structural raw data format into a structured layout. Once the video data has been transformed, features are extracted to obtain the main features and stored in structured databases. By employing data mining methods, mining can be done with this generated output [15].

3. Video Pre-Processing

Most of the time, unstructured video data is a data source. No matter how hard you try, this knowledge cannot be duplicated. The video is broken up into individual video frames to convert video data files into a format in which it can be easily structured. The shot boundary is found in the first phase before any processing has taken place [6]. Video data is broken down into the shot, which is the basic unit. The primary components of video are text, audio, and video. Object segmentation techniques and color analysis are used to capture visual features. Both frequency-domain and time-domain audio features are included, and text features that use video text processing techniques. In the early stages of processing, all the various components are applied to process the raw video sequences into video clips that can be organized and sorted into time steps [7]. Video frames can be used as an input once the key frames have been extracted with the shot partitioning algorithm. Using multimodal features, time-lapse shots, the key frames are discovered, and in the next step, the video data is generated [16].

4. Video Data Construction

When it comes to creating a video sequence, it is not uncommon for every symbolic stream (visual, audio, text, and objects with frame window) to be assigned a symbol. Each frame in the video is treated as a time unit [13]. According to the Lookup Table (which maps every feature to a symbolic value), each time unit's extracted features are translated into extended streams to produce transformed elements at the end of the video. The temporal video sequence, which consists of multiple streams into a single stream, is made using the original video data. A transformed structured video sequence is utilized in the mine associations in the video database [1].

5. Feature Extraction

Visual features are extracted from the key frame images using image processing algorithms. Audio feature extraction systems use signal classification techniques to collect several audio segments from the input audio signal. It is possible to glean useful information from the text present in the video by conducting mining on the associations [8]. Videotext is divided into two categories: featured video and associated video. When the text is superimposed on video sequences, it helps the viewer know what is being presented. We apply a symbol the first time we find text within the data. Every step, procedure, and significant action is shown in step-by-step detail in a video [20].

6. Applications Of Video Mining

Many different applications, including defense, exploration, education, entertainment, and sports, use video mining technology. Video data mining is the search for knowledge to extract data from visual sources, establishing patterns and dynamics. The video mining options on this page might be of interest to you.

i)Vehicles monitoring procedures: Video is now an essential part of nearly every security system and can help you decide which cameras and monitoring setups you'll need to meet your needs. We provide a wide range of video recording, image management, as well as cutting-edge technology. Video data mining can be used to regulate various types of public transportation. Digital video processing is helpful for vehicle monitoring. Since it is a video-based system, the traffic department can track vehicles [9]. Traffic flow monitoring is done on major thoroughfares to keep transport speeds and tolls on schedule.

ii)Traffic control management: One of the new uses thermal imaging is discovering is on traffic video. Extractive analytics or mining for video data can lead to better decisions regarding traffic management Viewers will have access to the live feeds of traffic cameras installed on the traffic posts [12]. This real-time input is brought into the system to give it more data. According to the available video, the traffic management systems the procedures will be controlled by the controls. Due to this solution, such as a suspicious vehicle activity, the analytics dashboard is immediately notified [6].

iii)Health Monitoring: It's used as the method for gathering the observation data. Many experimental modal analyses have shown that there is effective when natural frequencies are used as measurements. A CCTV stream can give healthcare workers a window into the well-being of the patient's health. Eulerian video magnification and optical flow stream camera are excellent for tracking the infant's respiratory difficulties. Eliminated research will allow for better treatment options in the future [10].

iv)Automated Transcription Video Data: Due to transcribing, small businesses can take advantage of audio/video data mining; audio/video data mining offers excellent opportunities for gaining unstructured data. It results in accurate texts being transformed into an increase in business productivity and efficiency. Simultaneously, it runs various routines, including feedback analysis, demographics, and legal compliance [11].

v) Enhanced Security with Live Video Streaming: It is excellent for transcribing individual-level large small businesses because the audio/video/video mining of accurate texts improves productivity and effectiveness. It performs various analyses, demographics, including feedback, while it also operates in compliance with a legal regulation [19].

7. Conclusion

The application of data mining is known as data mining and describes processes that search for hidden patterns and knowledge in large volumes of data. Until relatively recently, much of data mining research has focused on alpha-numeric databases. It remains the case, as multimedia data mining has not received the same attention. This paper discusses the current state of video data mining and the challenges in the field. It would be best if you concentrated on addressing these issues to gather extensive amounts of valuable data from an abundance of video data. In the data mining realm, a tremendous amount of research and development has been done over the years. Low-

levels out of spatial, inferring concepts from low-level videos and dealing with complexity in unstructured video unabstracted data.

References

[1] Chen, X., & Zhang, C. (2006). An Interactive Semantic Video Mining and Retrieval Platform— Application in Transportation Surveillance Video for Incident Detection. Sixth International Conference on Data Mining (ICDM'06), 129–138. https://doi.org/10.1109/ICDM.2006.20

[2] Kexue Dai, Jun Zhang, &Guohui Li. (2006). Video Mining: Concepts, Approaches, and Applications. 2006 12th International Multi-Media Modelling Conference, 477–480. https://doi.org/10.1109/MMMC.2006.1651376

[3] Kharade, K. G., Kharade, S. K., & Kumbhar, V. S. (2017). A Comparative Study of Traditional Server and Azure Server. Journal of Advances in Science and Technology, 13(1), 329–331.

[4] Potharaju, S. P., & Sreedevi, M. (2017). A Novel Clustering Based Candidate Feature Selection Framework Using Correlation Coefficient for Improving Classification Performance. Journal of Engineering Science & Technology Review, 10(6).

[5] Katkar, S. V., Kharade, K. G., Kharade, S. K., & Kamat, R. K. (2020). An Intelligent Way of Modeling and Simulation of WO3 for Supercapacitor. In N. Thapa, Recent Studies in Mathematics and Computer Science (pp. 109-117). Hooghly, West Bengal, India: Book Publisher International.

[6] Kharade, K. G., Kamat, R. K., & Kharade, S. K. (2017). Review On Microsoft Windows Azure Technology. National conference on "Innovative Trends in Electronics and Allied Technology (ITEAT-2017), (p. 33).

[7] Kharade, K. G., Kamat, R. K., & Kharade, S. K. (2017). Utilization of Copy Logger Software Package to Protect Sensitive Data from Malicious Individuals. National conference on "Innovative Trends in Electronics and Allied Technology (ITEAT-2017)" organized by Yashvantrao Chavan Institute of Science, Satara, (p. 19).

[8] Kharade, K. G., Kamat, R. K., Kharade, S. K., &Katkar, S. V. (2019). Automation of Paper Setting Process to Improve Effectiveness of The Examination System of The University. Journal of Emerging Technologies and Innovative Research,490-493.

[9] Kharade, K. G., Kamat, R. K., Mudholkar, R. R., & Kharade, S. K. (2018). Removable Drive Blocker Application for Virus Detection. International Journal of Research Culture Society, 233-234.

[10] Kharade, K. G., Kharade, S. K., &Kumbhar, V. S. (2018). Impact of Digital India on Various Sectors. Indian Journal of Innovation in Management and Excellence In Research (IJIMER), 37-40.

[11] Kharade, K. G., Kharade, S. K., &Oza, K. S. (2018). Security Threats In E-commerce. Indian Journal of Innovation In Management And Excellence In Research (IJIMER), 14-17.

[12] Patil, B. P., Kharade, K. G., &Kamat, R. K. (2020). Investigation on Data Security Threats & Solutions. International Journal of Innovative Science and Research Technology, 5(1), 79–83.

[13] Rao, G., Kumari, K., Shankar, D., & Kharade, K. G. (2021). A comparative study of augmented reality-based head-worn display devices. Materials Today: Proceedings. 1-9.

[14] Li, X., Shi, M., & Wang, X. (Shane). (2019). Video mining: Measuring visual information using automatic methods. International Journal of Research in Marketing, 36(2), 216–231. https://doi.org/10.1016/j.ijresmar.2019.02.004

[15] Matsuo, Y., Amano, M., & Uehara, K. (2002). Mining video editing rules in video streams. Proceedings of the Tenth ACM International Conference on Multimedia - MULTIMEDIA '02, 255. https://doi.org/10.1145/641007.641058

[16] Othman, E. H., Abderrahim, G., & Jaber, E. B. (2017). Mining Moocs Videos Metadata Using Classification Techniques. Proceedings of the 2nd International Conference on Big Data, Cloud and Applications, 1–6. https://doi.org/10.1145/3090354.3090450

[17] Porreca, A., Scozzari, F., & Di Nicola, M. (2020). Using text mining and sentiment analysis to analyze YouTube Italian videos concerning vaccination. BMC Public Health, 20(1), 259. https://doi.org/10.1186/s12889-020-8342-4

[18] Rosenfeld, A., Doermann, D., &DeMenthon, D. (Eds.). (2003). Video Mining. Springer US. https://doi.org/10.1007/978-1-4757-6928-9

[19] Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.

[20] Vijayakumar, V., &Nedunchezhian, R. (2012). A study on video data mining. International Journal of Multimedia Information Retrieval, 1(3), 153–172. https://doi.org/10.1007/s13735-012-0016-2

[21] Santhosha, S. & Kiran, N.. (2018). A Systematic Review on Existing Data Mining Approaches Envisioned for Knowledge Discovery from Multimedia. International Journal of Electrical and Computer Engineering. 81. 908-916. 10.11591/ijece.v8i2.pp908-916.

Node Failure Management to Improve the Performance of Wireless Sensor Networks

Pallavi Katkar [a 1], Ashwini Pawar [b], Sheetal Zalte [c], Suhas Katkar [d]

[a] *Shivaji University, Kolhapur, India*
[b] *Modern College, Pune, India*
[c] *Shivaji University, Kolhapur, India*
[d] *Varroc Engineering, Pune, India*

Abstract. A sensor network can be defined an assembly of sensor nodes which associated by all together to complete particular detailed task. These sensor nodes are mostly in huge amounts also compactly installed moreover in the network area or very near to it. Sensor networks can be worked for several sectors such that: environmental monitoring, home, health care, Industries, military, and habitat. Failure of network is unavoidable in wireless sensor networks because of unfriendly location and non-reachable placement. Hence, it is needed that network faults are discovered in time and proper methods are engaged to bear network task. So, it is important to deliver fault forbearing systems for spread sensor applications. Numerous new work in this field yield severely different methodologies to talking the fault tolerance concern in routing. In this propose review and equate present fault tolerant practices to provision for sensor applications.

Keywords. Wireless sensor networks, node failure reasons, fault detection approaches, fault management, fault detection and recovery.

1. Introduction

A dynamic field of concern for researchers and manufacturing group is Wireless sensor networks (WSNs). A WSN is usually a network of scattered sensor nodes assembling data from the atmosphere to observe the situations of that atmosphere. There are two types of nodes are used for WSN application which are completely depend on the application requirement, static nodes and movable nodes are the categories of the nodes. Static nodes stay fixed and complete to recognizing the events, routing of packets, else perform as destination nodes, although movable nodes travel everywhere in the network to identify network situation and different jobs such as changing dead nodes. Node distribution is depend on application may be physical or random. In physical distribution, the sensors are physically dropped and data is routed with selected paths [1]. The presentation of WSNs completely rest on the assumption that the nodes are linked till the destination node., few nodes those are situated away from the sink node and because of this its require several steps to communicate with neighboring nodes and transfer the date to destination node[2, 3]. Restricted resources are the one of the design issues in sensor node.

---

[1] Pallavi Katkar, Shivaji University, Department of Comp.Sci., India
Email: pal.katkar@gmail.com

Several sensor nodes might be dying or congested because of shortage of energy resource, have physical loss or atmosphere interfering. The fault of sensor nodes would not disturb the complete performance of the sensor network [4, 5]. Faulty node or multiple nodes may disturb the connectivity and might be the reason for dividing of the network.

A resolution to this issue is to install renewed nodes changing the failed ones which are an expensive and mostly not feasible method. Instead of this routing algorithm it needed to be self management characteristics are self organizing, healing and fault tolerant. The approach to handling the fault is to plan a scheme that is self healing tolerant to initiate with. Self healing is the capability to sustain sensor network's activity lacking any intermission because of sensor nodes failure. So that routing algorithm can manage node failure management and improve the network performance.

2. Literature Review

In Maintaining connectivity while keeping complete handling undamaged by with restricted amount of movable nodes are a tough job, particularly while a particular otherwise numerous nodules collapse. The situation is problematic of extreme significance those are taken considered in the literature survey in recent ages. Several explanations require to be suggested concerning in the topic. [6] This work proposed a novel fault finding scheme to resolve fault related issues. This work suggested four methods. With each fault all methods are checked and after that an inquiry message are taken consider to minimize the wrong results. This find failed node and try to replace those with new one. But it not worked link failure and management. [7] Self-Sensor deployment geographically with automatic manner. Algorithm to define re-location of sensors. Algorithm failed to solve to manager node failure management. Rather they replace the nodes with new node. But every time it is not possible to replace nodes. It is most costly and time consuming. This scheme [8] proposed software oriented less overhead fault discovery technique to identify faults in numerous hardware modules. The failure recognition arrangement has remained performed in the SOS kernel of the sensor nodes. Self healing systems support hardware built sensors those are used in electronic devices. But, its failed serve mobile WSN. [9] An approach to quotation such material from a gathering of described facts are stated. Lastly, subsequent the planned method, a programmed fuzzy logic constructed method for failure recognized in LTE networks is scheduled. Its only work in LTE networks. It is not consider mobility as parameter while considering link failure. [10] An innovative method is projected to decrease battery operated intake those are grounded on the self-motivated grouping with the use of neural network. Using Boltzmann concept in neural network, former data distribution sensor nodes generate cluster dynamically. Loads remain correct conferring for the situations and the situation as well improve the competence of the active grouping. This works considered static nodes not mobile nodes.

3. Faulty Nodes in WSN

When it comes to faulty nodes in a sensor network, it is necessary to know about faulty node, node failure reasons, faulty node detection, and diagnosis of fault. In the network

nodes are get fail with the number of reasons. Some of node failure reasons are Power drain, Out of transmission range, Harsh Environment conditions, malicious nodes, Node movement Speed and direction.

3.1 Faulty Node

Sensor node remains assumed to be faulty node if the data sent to the destination or the intermediate (neighboring) nodes are present in the network is improper. Due to faulty nodules network faces many issues like network get splits, network performance get degrades and sometimes its cause end of the network.

3.2 Node Failure Reasons

So many reasons are there to have faulty nodes in a sensor network but they cannot be predicted exactly because the sensor nodes are deployed in a huge amount. Still some of the node failures reasons which are occur frequently are;

- Power Drain - Wireless sensor network having limited energy resources.
- Out of transmission range - Ehen nodes are mobile nodes are frequently move out of transmission range causes packet loss issues.
- Malicious Node - malicious nodes in network which behaves so, other nodes become unreachable or dead.
- Congestion - Find the malicious nodes in network which behaves so, other nodes become unreachable or dead.
- Node movement speed and direction – If nodes are mobile then mobility of nodes their speed and direction creates issue for network.

3.3 Fault Detection Approaches

Fault discovery mostly scheduled as per the category of the application and also which sort of failure, certain existing failure discovery patterns are discoursed. Here classified the present failure recognition methods are: Centralized, Distributed and Cluster

Table1. Fault Detection Approaches

Type	Description	Advantage	Disadvantage
Centralized	A geologically or reasonably integrated sensor node having central organizer or administrator which is concern for observing and locating faulty or disobedience nodes in the network.	1. Capable to perform extended variety of failure managing preservation. 2.Active detection model 3.Help for route selection	1. Unlimited resource required. 2. More storage required. 3. Increased network overhead. 4. If center node fails whole system fails.
Distributed	Distributed method inspires the conception of local management, which uniformly allocates fault managing into the system.	1. Node level decision 2. Less network traffic 3. Co-operation from all network nodes 4. Not depend of single node 5. Less Storage required	1. Difficult to decide when to start fault detection 2. Malicious nodes effects decisions

Cluster	The cluster based architecture implemented fault detection mechanism in a distributed manner through intra cluster communication and reports the failed nodes to the upper layer of communication hierarchy.	1. Combination of center and distributed types 2. Managing fault node is easy 3. No single node centric 4. Allow node level decision	1. Cluster head selection 2. Hierarchical topology required 3. Different approach needed to select fault in node and cluster head.

4. Proposed Work

Proposed work supposed that each and every sensor nodes are arbitrarily spread in the network range as well as each sensor node is expected to identify its individual place and also its neighboring node and sink node too. Also all sensor nodes have the equal quantity of initial energy and equal maximum communication range. In this proposed scheme, every time due limited energy resources sensor nodes get fail to keep communicating with neighboring node and also link with destination node get break. This causes end of the network. In proposed scheme using self healing concept proposed routing protocol predict the fault which may cause due to node energy drain or node moving out of transmission range. Self healing system is set of algorithms to identify, and analyze the root cause of the faults (Network Fault Detection System) and then correct them using Fault Management System. Critical part of Self-healing is able to detect failure of a node and once detect those nodes then how re-organize network and recover from node failures without losing any information and controllable delay. Proposed work performance comparison done with existing Ad hoc On Demand Distance Vector (AODV) routing protocol. Proposed system increases throughput, decrease delay which verified and presented in performance analysis.

5. Performance Analysis of Routing Protocol

5.1 Simulation Methodology and Parameters

The planned method considered network simulator Ns3 for simulation. Simulation helps to recognize the actions and presentation of the network and its protocol. The proposed scheme considered node 0 as sink node and others are source nodes.

Table2. Simulation Parameters

Parameters	Values
Network Simulator	Ns3.20
Connection Type	UDP
Simulation Time	100ms
Number of Nodes	120
Simulation Area	500m*500m
Node Pause	0s
Traffic Flow	CBR

6. Result Analysis

6.1 Throughput
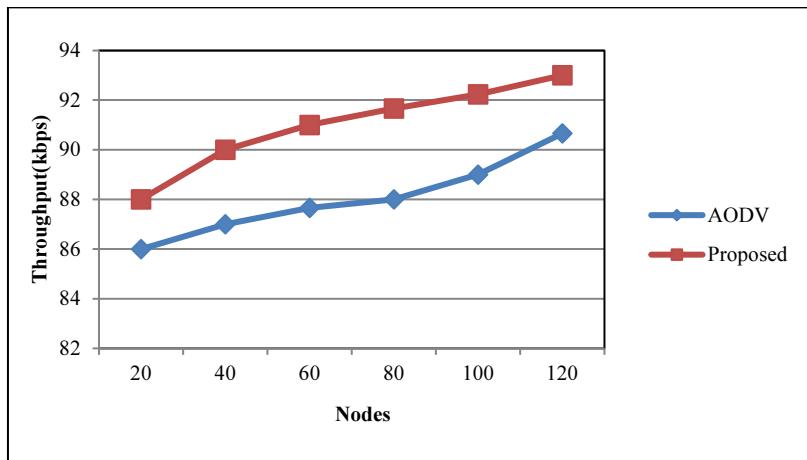
Figure1. Throughput Graph

The amount of successful message delivery from source node to destination node in the network is called throughput. Proposed protocol throughput is more than AODV. Fault management scheme handle the fault and reduces repeated path discovery.

6.2 Delay
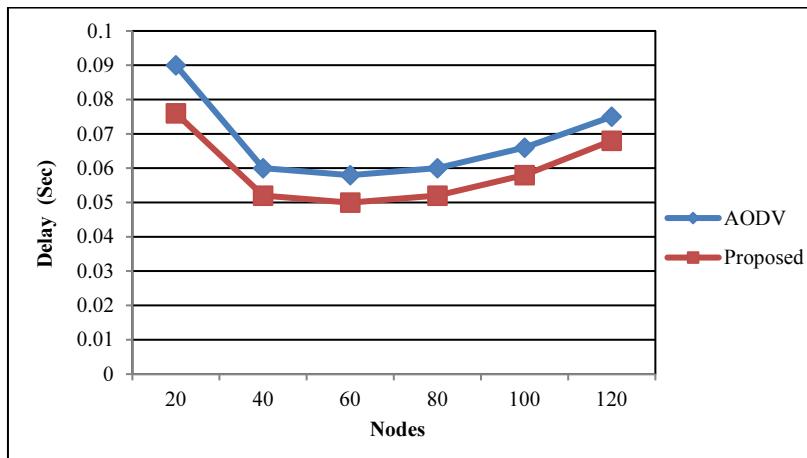
Figure2. Delay Graph

Delay means specifies the amount of time data the packets takes to reach the destination. Proposed scheme reduces delay than AODV routing protocol.

7. Conclusion

Wireless sensor network is linked of numerous wireless detecting nodes. These tiny sensor devices have limitation of resource and randomly installed in harsh location. Hence, it is very common for wireless networks to occur faults, node failure, routing path or network fault etc. Proposed scheme, presented self healing fault managing scheme for wireless sensor network to detect mistakes and execute suitable methods to improve sensor network from failure. The proposed fault managing scheme is help to increase the performance of wireless sensor network. Comparative simulation result shows that proposed work improves the throughput. Main reason behind increase in throughput less link failure reduces packet drop and self healing avoids repeated path discovery leads to reduce in delay. Reduced delay and increase in packet delivery, resulted in better throughput.

References

[1] Pallavi Katkar, V. R. Ghorpade, A Survey on energy Efficient Routing Protocol for Wireless Sensor

[2] Networks, International Journal of Computer Science and Information Technology, Vol.6, 2015, 81-83

[3] V. Ranga, M. Dave, A. K. Verma, Node stability aware energy efficient single node failure recovery

[4] approach for WSANs, Malaysian Journal of Computer Science, Vol. 29, 2016, 106–123

[5] C. Zhu, C. Zheng, L. Shu, and G. Han, A survey on coverage and connectivity issues in wireless sensor networks, Journal of Network and Computer Applications, Vol. 35, 2012, 619–632

[6] I.F.Akyildiz, W.Su, Y.Shankarasubramaniam, E.Cayirci, Wirless sensor networks: a survey, Computer Network, Vol. 38, 2002, 393-422

[7] Jamal N. AL-Karaki, Ahmed E. Kamal, ROUTING TECHNIQUES IN WIRELESS SENSOR NETWORKS: A SURVEY, IEEE Wireless Communications, 2004

[8] Shahram Babaie, Ahmad Khadem-zadeh and Kambiz Badie, Distributed Fault Detection Method and Diagnosis of Fault Type in Clustered Wireless Sensor Networks, Life Science Journal, 2012

[9] Ting-Yu Lin, Hendro Agus Santoso, and Kun-Ru Wu, Global Sensor Deployment and Local Coverage-Aware Recovery Schemes for Smart Environments, IEEE Transactions on Mobile, 2014, 14(7)

[10] Hsung-Pin Chang and Tsung-Yu Yeh, A Software-based Fault Detection Scheme for Wireless Sensor Networks, IEEE International Conference on Computer and Information Science, 2016

[11] Gómez-Andrades, P. Muñoz, E. J. Khatib, I. de-la-Bandera, I. Serrano, and R. Barco, Methodology for the Design and Evaluation of Self-Healing LTE Networks, IEEE Transaction on Vehicular Technology, 2015, Vol. 65

[12] Pardeep Kumar Tanisha Saini, TO PROPOSE A TECHNIQUE FOR FAULT TOLERANCE IN WIRELESS SENSOR NETWORKS, International Journal of Computer Engineering and Technology, 2016, Vol. 7

[13] Rab Nawaz Jadoon, Adnan Anwar Awan, Muhammad Amir Khan, WuYang Zhou and Aamir Shahzad, An Efficient Nodes Failure Recovery Management Algorithm for Mobile Sensor Networks, Mathematical Problems in Engineering Hindawi, 2020

[14] Chiatanya Vijaykumar Mahamuni, Sensor Node Failure Affecting Coverage of WSNs: An Overview, International Journal of Emerging Technology in Computer Science and Electronics, 2015, Vol.14

[15] Rukhsar Begum Shaikh, Abid H Sayed, SENSOR NODE FAILURE DETECTION IN WIRELESS SENSOR NETWORK: A SURVEY, International Journal of Emerging Technology in Computer Science and Electronics, 2015, Vol. 14

[16] Elham Moridi, Majid Haghparast, Mehdi Hosseinzadeh, Somaye Jafarali Jassbi, Fault management frameworks in wireless sensor networks: A survey, Computer Communications ELSEVIER, 2020, Vol.155

COVID-19:Face Mask Detector with Open CV and CNN Algorithm

Tambe Pragati [a] [1], Tanpure Akshada [a], Wakchaure Asmita [a], Zaware Prachi [a], Prof. S. A. Bhosale [a]

[a] *Dept of Computer Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra, India*

Abstract. Now a day, covering our faces with a mask has become a new normal habit in this pandemic, as face masks are effective in preventing the virus outbreak. Masks reduce risk from an infected person whether they have symptoms or not. In this paper, we propose a system that restrict the growth of COVID-19 by finding out peoples with mask and without mask. Where all the public places are monitored with CCTV cameras. A deep learning architecture is trained on a dataset which consists of images of people with wearing mask and without wearing is masks collected from various sources.

By using image processing analysis and machine learning method we can find out face mask wearied or not. Face mask detection can be done using various methods. Mainly convolutional neural network and OpenCV method is used. The accuracy and decision making of CNN algorithm is higher than other algorithms.

Keywords: COVID-19: Corona Virus Disease of 2019, CNN: Convolutional Neural Network, SVM: Support Vector Machine, KNN: K Nearest Neighbor

1. Introduction

India has a high-test positivity rate for COVID-19. In many countries' lockdowns are released when their COVID-19 numbers started reducing. Covering our faces with a mask has become a new normal habit in this pandemic, as face masks are effective in preventing the virus spreading. In so many developed and underdeveloped countries over worldwide have made it compulsory for people to wear masks while leaving home or visiting public places. On the other hand, it will be challenging to recognize faces with masks in monitoring systems. In order to prevent the spread of COVID-19, it is necessary to wear a mask. Therefore, it is very necessary to improve the face recognition performance of the existing technology on masked and unmasked faces.

Face mask detection is a challenging task. It has been getting more attention in the era of covid 19 due to the spreading of disease in large quantity. In the case of medical field, mask reduces the chances of exposure risk from an infected person whether they have symptoms of covid-19 or not. Face mask detection system is used in Airports, Hospitals, Offices and Educational Departments, Industries etc.

Tambe Pragati, *Dept of Computer Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra, India*

Email: tambe@sanjivani.org.in.

1) Face Recognition

2) Feature Extraction

This paper aims at designing a system to find out whether a person is using a mask or not and displaying the message if person is not wearing the mask. Firstly, CCTV cameras are used to capture real-time image of different public places or at the entrance of office. From that image, facial images are extracted and these images are used to identify the mask on the face. Convolutional Neural Network (CNN) is used for feature extraction from the images then these features are learned by multiple hidden layers.

2. Related Work

There are many techniques are used for face mask detection. Some of them are explained below.

1. Coronavirus disease 2019 has affected the world seriously. One major protection method for people is to wear masks in public areas. Furthermore, many public service providers require customers to use the service only if they wear masks correctly. However, there are only a few research studies about face mask detection based on image analysis. In this paper, we propose Retina Face Mask, which is a high-accuracy and efficient face mask detector.[1].

2. A modern image recognition that has millions of parameters and requires a lot of training data as well as high computing power that is hungry for energy consumption so it becomes inefficient in everyday use. Machine Learning has changed the computing paradigm, from complex calculations that require high computational power to environmentally friendly technologies that can efficiently meet daily needs. We then demonstrate the effectiveness of MobileNets across a wide range of applications and use cases including object detection, face attributes and large scale geo-localization [2].

3. In order to effectively prevent the spread of COVID19 virus, almost everyone wears a mask during coronavirus epidemic. This almost makes conventional facial recognition technology ineffective in many cases, such as community access control, face access control, facial attendance, facial security checks at train stations, etc. Therefore, it is very urgent to improve the recognition performance of the existing face recognition technology on the masked faces. Most current advanced face recognition approaches are designed based on deep learning, which depend on a large number of face samples [3].

4. Deep convolutional neural networks have been successfully applied to face detection recently. Despite making remarkable progress, most of the existing detection methods only localize each face using a bounding box, which cannot segment each face from the background image simultaneously. To overcome this drawback, we present a face detection and segmentation method based on improved Mask R-CNN, named G-Mask, which incorporates face detection and segmentation into one framework aiming to obtain more _ne-grained information of face [4].

5. Faces in the dataset have various orientations and occlusion degrees, while at least one part of each face is occluded by mask. Based on this dataset, we further propose LLE-CNNs for masked face detection, which consist of three major modules. The Proposal module first combines two pre-trained CNNs to extract candidate facial regions from the input image and represent them with high dimensional descriptors. After that, the Embedding module is incorporated to turn such descriptors into a similarity-based descriptor by using locally linear embedding (LLE) algorithm [5].

6. Face Detection has evolved as a very popular problem in Image processing and Computer Vision. Many new algorithms are being devised using convolutional architectures to make the algorithm as accurate as possible. These convolutional architectures have made it possible to extract even the pixel details. We aim to design a binary face classifier which can detect any face present in the frame irrespective of its alignment.

7. Coronavirus disease 2019 has affected the world seriously. One major protection method for people is to wear masks in public areas. However, there are only a few research studies about face mask detection based on image analysis. The proposed RetinaFaceMask is a one-stage detector, which consists of a feature pyramid network to fuse high-level semantic information with multiple feature maps, and a novel context attention module to focus on detecting face masks. [7].

8. According to data obtained by the World Health Organization, the global pandemic of COVID-19 has severely impacted the world and has now infected more than eight million people worldwide. Wearing face masks and following safe social distancing are two of the enhanced safety protocols need to be followed in public places in order to prevent the spread of the virus. Thus, the proposed system favors the society by saving time and helps in lowering the spread of corona virus. It can be implemented effectively in current situation when lockdown is eased to inspect persons in public gatherings, shopping malls, etc. Automated inspection reduces man- power to inspect the public and also can be used in any place [8].

9. Head pose classification is widely used for the preprocessing before face recognition and multi-angle problems, because algorithms such as face recognition often require the input image to be a front face. But affected by the COVID-19 pandemic, people wear face masks to protect themselves safe, which makes cover most areas of the face. The proposed HGL method combines the H- channel of the HSV color space with the face portrait and grayscale image, and train the CNN to extract features for classification[9].

3. Methodology

We proposed an automated smart framework for screening persons who are not using a face mask in this paper. Office or any public place are monitored by CCTV cameras. The cameras are used to capture images from public places; then these images are feed into a system that identifies if any person without face mask appears in the image. If any person without a face mask is detected then this information is sent to the proper authority to take necessary actions. The block diagram of the developed framework is depicted in Fig. 1.
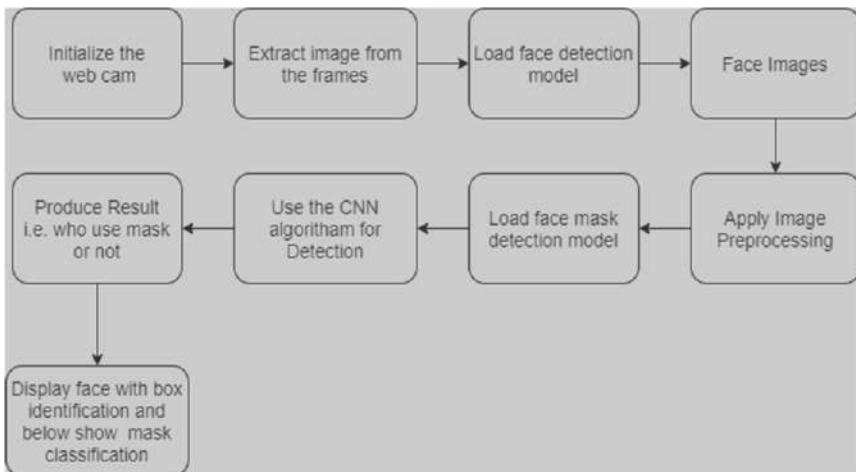
Figure 1. System Architecture

A. Image Preprocessing

The images captured by the CCTV cameras required preprocessing before going to the next step. In the preprocessing step, the image is transformed into a grayscale image because the RGB color image contains so much redundant information that is not necessary for face mask detection. RGB color image stored 24 bit for each pixel of the image. On the other hand, the grayscale image stored 8 bit for each pixel and it contained sufficient information for classification. Then, we reshaped the images into (64×64) shape to maintain uniformity of the input images to the architecture. Then, the images are normalized and after normalization, the value of a pixel resides in the range from 0 to 1

B. Deep Learning Architecture

The deep learning architecture learns various important nonlinear features from the given samples. Then, this learned architecture is used to predict previously unseen samples. To train our deep learning architecture, we collected images from different sources. The architecture of the learning technique highly depends on CNN. All the aspects of deep learning architecture are described below.

i)Dataset Collection

Data from two different sources are collected for training and testing the model. We collected a total of 1400 images of people with masks and 1400 images of people without a mask. For training purposes, 80% images of each class are used and the rest of the images are utilized for testing purposes. Fig. 2 shows some of the images of two different classes.

Figure 2 Faces Without Mask

Figure 2. Faces With Mask

ii) Architecture Development

The learning model is based on CNN which is very useful for pattern recognition from images. The network comprises an input layer, several hidden layers and an output layer. The hidden layers consist of multiple convolution layers that learn suitable filters for important feature extraction from the given samples. The features extracted by CNN are used by multiple dense neural networks for classification purposes.

iii) Screening and displaying Message

The main goal of our proposed system is screening persons who are not following guidelines of using a facial mask. The learning architecture identifies whether any input image contains persons without a face mask. If such a person is detected, then system will display message to wear the mask.

4. Algorithm
CNN:

A **Convolutional Neural Network (ConvNet/CNN)** is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imagery.
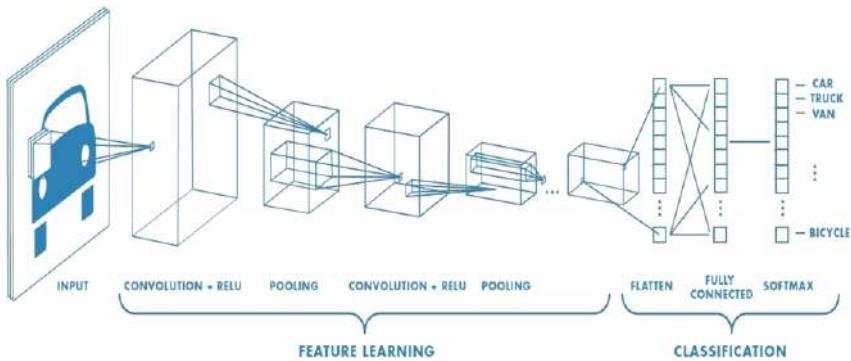
Figure 3. CNN Algorithm Blocks
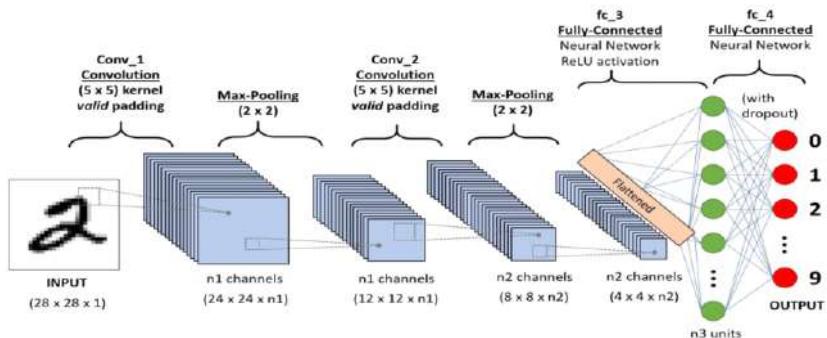
Figure 4.CNN actual Matrix Representation

5. Conclusion

This paper presents a system for a public place e.g., office to reduce the spread of coronavirus by informing the authority about the person who is not wearing a facial mask that is a precautionary measure of COVID-19. The system captures the images of employees and these images are classified using different AI algorithms and a combination of them, then their performance was evaluated to detect the image who wore the mask. These algorithms include a convolutional neural network (CNN), Softmax, support vector machine (SVM), Random Forest, and K nearest neighbor (KNN). The system contains a face mask detection architecture where a deep learning algorithm is used to detect the mask on the face. To train the model, labelled image data are used where the images were facial images with masks and without a mask.

6. Limitations and Future Work

The developed system faces difficulties in classifying faces covered by hands since it almost looks like the person wearing a mask. For a very densely populated area, distinguishing the face of each person is very difficult. For this type of scenario, identifying people without face mask would be very difficult for our proposed system.

In order to get the best, result out of this system, office entrance should have two or more CCTV cameras to monitor the entrance of office as well as dedicated manpower to enforce proper laws on the violators.

References

[1]. Barlian Khasoggi, Ermatita, "Efficient mobilenet architecture as image recognition on mobile and embedded devices." Samsuryadi Master of Informatics Engineering, Sriwijaya University, Indonesia.

[2]. Andrew G. Howard Menglong Zhu Bo Chen Dmitry Kalenichenko Weijun Wang Tobias Weyand Marco Andreetto Hartwig Adam " MobileNets: E_cient Convolutional Neural Networks for Mobile Vision Applications."

[3]. Mingjie Jiang, "RETINAFACEMASK: A FACE MASK DETECTOR" Department of Electrical Engineering City University of Hong Kong Hong Kong xinqifan2c@my.cityu.edu.hk Hong Yan Department of Electrical Engineering, City University of Hong Kong Hong Kong h.yan@cityu.edu.hk June 9, 2020

[4]. Kaihan Lin , Huimin Zhao , Jujian Lv , Canyao Li, Xiaoyong Liu, Rongjun Chen,"Face Detection and Segmentation Based on Improved Mask R-CNN " ,School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China Correspondence should be addressed to Huimin Zhao; zhaohuimin@gpnu.edu.cn and Jujian Lv; jujianlv@gpnu.edu.cn, MAY 2020

[5]. Shiming Ge1, Jia Li2, Qiting Ye1,3, Zhao Luo1,3,"Detecting Masked Faces in the Wild with LLE-CNNs", Beijing Key Laboratory of IOT information Security, Institute of Information Engineering, Chinese Academy of Sciences 2State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University 3School of Cyber Security, University of Chinese Academy of Sciences .

[6]. Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore

[7]. Mingjie Jiang, Xinqi Fan* Hong Yan,"RETINAFACEMASK: A FACE MASK DE- TECTOR" Department of Electrical Engineering City University of Hong Kong Hong Kong minjiang5-c@my.cityu.edu.hk, june 2020.

[8]. J. Shashi Yadav, \ Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID- 19 Safety Guidelines Adherence ," IInternational Journal for Research in Applied Science Engineering Technology (IJRASET), ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429 , ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429.

[9]. Lisette Garc__a-Moya, Henry Anaya-S_anchez, and Rafael Berlanga-Llavori, 2013, Multi-angle Head Pose Classi_cation when Wearing the Mask for Face Recognition under the COVID- 19 Coronavirus Epidemic," 1 Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China.,International Conference on High Performance Big Data and Intelligent Systems (HPBDIS),2020.

Mineral Rock Classification Using Convolutional Neural Network

Shanmuk Srinivas Amiripalli [a 1], Grandhi Nageshwara Rao [a], Jahnavi Behara [a],K Sanjay Krishna[a], Mathurthi pavan venkat durga ram[a]

[a] *Dept of CSE, GIT, GITAM University, Visakhapatnam, Andhrapradesh ,India*

Abstract. The main aim of the research is to build a model that can effectively predict the type of mineral rocks. Rocks can be predicted by observing it is colour, shape and chemical composition. On-site technicians need to apply different techniques on rock sample in order to predict rock type. Technicians need to apply different techniques on rock samples, so it is a time-consuming process, and sometimes the predictions may be accurate, and sometimes predictions may be false. When predictions are false, it might show a negative impact in several ways for workers and organization as well. We considered an image dataset of rock types, namely Biotite, Bornite, Chrysocolla, Malachite, Muscovite, Pyrite, and Quartz. We applied CNN (Convolutional Neural Network) Algorithm to get a better prediction of different mineral rocks. Nowadays, CNN is mainly used for image classification and image recognition tasks.

Keywords. Deep learning, convolutional neural networks (CNN), image pre-processing, convolution layer, pooling layer, flatten layer, fully connected layer

1. Introduction

The traditional method for rock classification is manual work with many problems like time-consuming and low accuracy. We can reduce the amount of time required and increase the accuracy by using the domains present in developing technology. Already some scholars or students researched image classification. Deep learning is a sub-domain of Machine learning that is used to know how a human brain works on data by creating patterns for each object in data and producing results based on those patterns. It has the capability of learning unsupervised data (un labelled data) using its hierarchical networks. Deep learning has gained much interest from researchers because it can be used for processing unsupervised data. Convolution Neural Network (CNN) is a class of deep learning which is commonly used for image classification, medical problems (examples detection of cancer, brain tumor, etc.), video recognition (for example identifying students in zoom call for proper attendance, etc.). All these tasks can be completed using CNN with better accuracy when compared to other algorithms. CNN is one of the best techniques used for extracting features from an image. It can automatically extract features from images with high accuracy,

---

[1] Grandhi Nageshwara Rao, Dept of CSE, GIT, GITAM University, Visakhapatnam, Email: nageshwarao1127@gmail.com

Whereas its predecessors cannot learn or identify features without any human supervision. It has a wide range of applications such as plant classification basing on images, face recognition, handwritten recognition, analyzing documents. So, we used CNN to create a model that would help users predict the type of mineral rock. This paper focuses on the creation of a rock classifier using CNN's concepts. Our dataset contains various images of seven kinds of rock minerals, namely biotite, bornite, chrysocolla, malachite, muscovite, pyrite and quartz, and all of them are col oured images. Binary classification is a type of classification in which an image is related to only two classes. For example, it can be a "dog(or)cat" prediction. Here, if a user provides an image to the model as input, then the model will identify features and then try to predict the result to which class the image relates. Multi-class classification is a type of classification in which an image can be related to n-number of classes, where 'n' depends on the training dataset, for example, "flower" prediction. Here 'n' depends on how many types of flowers the user provided for training. If the user-provided only sunflower, jasmine, and rose, then the 'n' value will be 3. Value of 'n' should be greater than two(n>2) then only it would be considered as multi-class classification.

Image classification is a process of recognizing the data to which class does it relates to. We can predict the features in an image by looking at it, whereas computers cannot identify images directly. So, we use different algorithms to develop a model that can identify the class of an image. Here class can be any real-time object. For example, it can be an animal, thing, humans, etc., so we provide a set of data to the Algorithm, which is helpful in the model's training process to identify the class to which the given image relates. Images are divided into a training set and a testing set. The testing set will not be included in the training set, and these images will be new images to the model user build. Binary classification is a type of classification in which an image is related to only two classes. For example, it can be a "dog(or)cat" prediction. Here, if a user provides an image to the model as input, then the model will identify features and then try to predict the result to which class the image relates. Multi-class classification is a type of classification in which an image can be related to n-number of classes, where 'n' depends on the training dataset, for example, "flower" prediction. Here 'n' depends on how many types of flowers the user provided for training. If the user-provided only sunflower, jasmine, and rose, then the 'n' value will be 3. Value of 'n' should be greater than two(n>2) then only it would be considered multi-class classification. The human brain has three main parts called Cerebrum, Cerebellum, Brainstem. When humans try to view anything using their eyes, the information is passed to the region called Visual Cortex, present in Cerebrum. The visual cortex contains many layers, and these layers can also be called filters. Each layer has its job to find edges, identify multiple objects in data, vertical edge detection, horizontal edge detection, detect human faces, detect moving objects, etc. Convolution neural network is well-known image recognition. If we give some binary image, some processing will be done, and it looks for feature orientation and gives output. In CNN, we generally use the 3-dimensional image in which every pixel will have three values, separate channel for red, green end blue colour value. The mathematical function of CNN has two functions (f and g) resulting from the third function(f*g) that identifies how the shape of one is changed by the other. This can be achieved by taking a 3*3 stride or window and placing it on the image, and then it is multiplied with a filter so that we can get weights to the features in images. After completing the first matrix multiplication, we need to move our stride to the next column and perform matrix multiplication again. If

stride reaches the end of a column, then we move stride to the next row. We continue this process till stride reaches to last of the image.

2. Literature survey

In [1], a visual technique that gives insight into the function of intermediate feature layers and the operation of the classifier was introduced to know why the CNN model works so well and how it can be improved. In [2], three different networks Alex Net, GoogLeNet, and ResNet50, were applied on the 3 different datasets ImageNet, CIFAR10, and CIFAR100, better in capabilities and limitations. The analysis stated that GoogLeNet and ResNet50 were able to recognize objects with better precision. In [3] Artificial neural networks and other methods of image classification 2007, Techniques Used is Artificial Neural Networks (ANN), Support Vector Machines (SVM), Fuzzy measures, Genetic Algorithms (G.A.), Fuzzy Support Vector Machines (F.S.V.M.). Relevance is This page demonstrates different accuracy values for traditional classification methods. Here images are pre-processed using Artificial Neural Network (ANN) to compare the accuracy with other machine learning algorithms. The future scope was to Improve the automatic techniques and apply CNN for a higher accuracy result. This paper also fails to consider over fitting between models. In [4], they are Pruning Convolution Neural Networks for Image Instance Retrieval 2010. The technique used is Convolution Neural Network (CNN), node pruning. Relevance is in this paper, and the author tried to optimize the default CNN algorithm by adding pruning techniques to them, which results in the loss of accuracy in the neural network and increases the speed by 3% for evaluating and classifying datasets over 35000 entries. The future scope is that we can try to reduce the loss by using any other optimized techniques. In [5-6], they improved Optimization of Convolution Neural Networks through parameter fine-tuning 2019.

Techniques used are Convolution Neural Network (CNN), hyperparameter tuning. In this paper, the author compares the original CNN build with another same network that is optimized by using parameter tuning and feature reduction. The future scope is this is a comparative study not much is to be improved. In [7], A method was proposed to predict lung cancer from C.T. scan images by using Convolutional Neural Networks. The experimental study was carried on a real-time dataset collected from Iraq-Oncology Teaching Hospital/National Centre for Cancer Diseases (IQ-OTH/NCCD). The model achieved 95% accuracy. In [8], A new methodology was proposed to detect real-time gender classification based on facial images. In the newly proposed methodology, the number of processing layer was reduced. Unlike in conventional CNN, convolution operation was replaced with cross-relation, which reduced computational loads. It was performed in two datasets i.e., SUMS and AT&T to which the obtained accuracy were 98.75% and 99.38%, respectively. In [9], Very Deep Convolutional Neural Network-based image classification using small training samples. Techniques used are Convolutional Neural Network (CNN), batch normalization, regularisation. The author uses regularisation and batch normalization on CNN to fit small datasets with proper and straightforward modifications and do not need to redesign specific small networks. The future scope is we can think of optimizing the overfitting problem by removing some of the nodes in the network to make it slightly weighted. In [10], Optimization of Convolution Neural Network (CNN) parameters for image classification 2017. A technique used is Convolution Neural Network (CNN).

Relevance is an approach is specified in order to boost up the accuracy of CNN by two steps. One is to add the number of layers, and another is to decrease the window size of the image. The future scope is this is a comparative study not much is to be improved. Optimization can also be achieved by TGO model [11-17].

3. Proposed Model

Initially, collect the dataset by downloading images individually or downloading the entire dataset from online websites. The dataset is divided into training and test datasets [18]. Then image pre-processing is applied to the training dataset, which helps the model train well by making all the images into the same size. Users can add extra features like rotating images up to a certain angle and zooming in images up to a specific rate. These features help the model to train well. In training and test datasets, images are placed in their respective sub-folders so that the model would give labels to each sub-folder, and the model returns these labels or numbers as output after the prediction [19].

Proposed Algorithm to build CNN Model
Step 1: Start.
Step 2: Import all required packages, libraries, and modules.
Step 3: Divide dataset into training and test sets.
Step 4: Apply pre-processing image tool to the dataset.
Step 5: Create a basic model using the Sequential method.
Step 6: Add a Convolution layer, Flatten layer, Pooling layer, Hidden layers, and Output layer to the model.
Step 7: Train the model and save it using ('.h5'-extension).
Step 8: Create a python file (Flask application) and load the model into the file.
Step 9: Create HTML pages for the application.
Step 10: Run the flask application for the website to load.
Step 11: Provide the image input on the website and get the result back.
Step 12: Stop.

To extract all the features from training images, we need to build a CNN model containing a group of layers. Now add all the layers (specified in the above steps in the same order). With the help of the training dataset, we train the model and save the model. Typically, users need to run entire programs to predict the images, but users can save the model and load that model into a web application. So users do not need to run the program every time for predicting images. Now users can load the model into a web application using respective packages available in python and host their web application. Now users can provide an image as input to application, and then the model will then find features in an image and return the matched class label as output [20].

4. Results and Discussion

We have seven types of mineral rocks in our dataset: biotite, bornite, chrysocolla, malachite, muscovite, pyrite and quartz, and have 951 images (859 images for training and 92 images for testing). We have 951 images for seven classes, so we took 10% for testing and 90% for training. We used CNN Algorithm to implement our research because CNN produces better results for image classification. Our model produced an accuracy of ~85%. We applied image pre-processing to images which helped our model to train well [21].
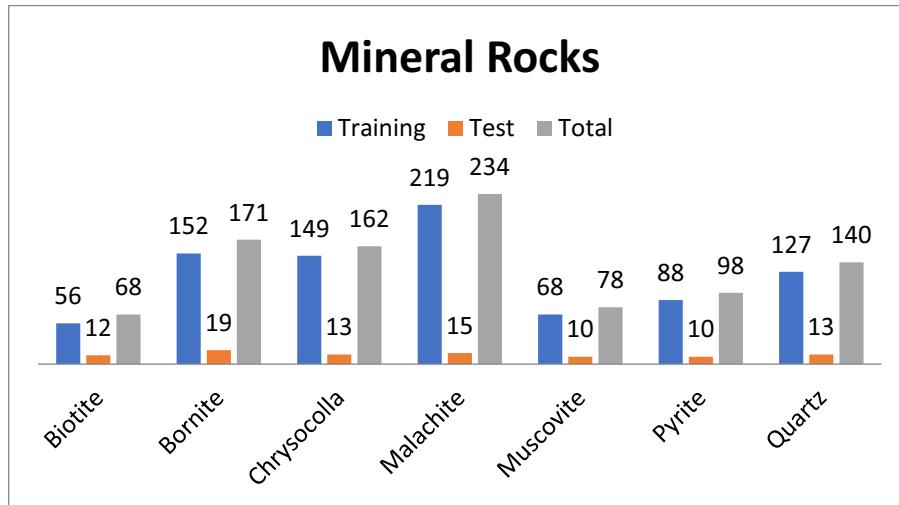
Figure 1. Comparison of mineral rocks dataset with training, test and total.
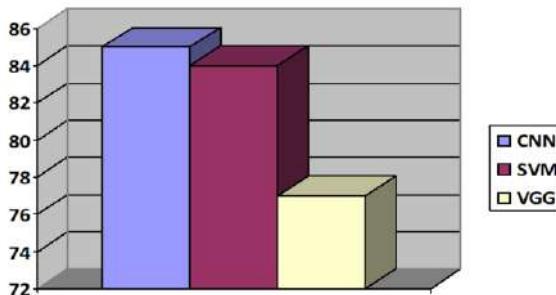
Figure 2. Comparison of accuracy

5. Conclusion

This research aims to build a good model that predicts the mineral rock using the images. So, we used CNN for better identification of features in an image. We even applied SVM and V.G.G. algorithms to our dataset. SVM algorithm scores an accuracy of 84% and the V.G.G. algorithm scores an accuracy of 77%. V.G.G. algorithm

accuracy can be increased by increasing the number of epochs but the time taken to train the model would be more. We state that our model (CNN algorithm) scored an accuracy of 85%, like the SVM algorithm's accuracy.

References

[1]. Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham.

[2]. Sharma, N., Jain, V., & Mishra, A. (2018). An analysis of convolutional neural networks for image classification. Procedia computer science, 132, 377-384.

[3]. Seetha M., Muralikrishna, I. V., Deekshatulu, B. L., Malleswari, B. L., & Hegde, P. (2008). Artificial neural networks and other methods of image classification. Journal of Theoretical & Applied Information Technology, 4(11).

[4]. Manek, G., Lin, J., Chandrasekhar, V., Duan, L., Giduthuri, S., Li, X., &Poggio, T. (2017). Pruning Convolutional Neural Networks for Image Instance Retrieval. arXiv preprint arXiv:1707.05455.

[5]. Becherer, N., Pecarina, J., Nykl, S., & Hopkinson, K. (2019). Improving optimization of convolutional neural networks through parameter fine-tuning. Neural Computing and Applications, 31(8), 3469-3479.

[6]. Hossain, T., Shishir, F. S., Ashraf, M., Al Nasim, M. A., & Shah, F. M. (2019, May). Brain Tumor Detection Using Convolutional Neural Network. In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (I.C.A.S.E.R.T.) (pp. 1-6). IEEE.

[7]. Potharaju, S. P., & Sreedevi, M. (2018). A novel subset feature selection framework for increasing the classification performance of SONAR targets. Procedia Computer Science, 125, 902-909.

[8]. Al-Yasriy, H. F., AL-Husieny, M. S., Mohsen, F. Y., Khalil, E. A., & Hassan, Z. S. (2020, November). Diagnosis of Lung Cancer Based on C.T. Scans Using CNN. In I.O.P. Conference Series: Materials Science and Engineering (Vol. 928, No. 2, p. 022035). I.O.P. Publishing.

[9]. Liew, S. S., Hani, M. K., Radzi, S. A., &Bakhteri, R. (2016). Gender classification: a convolutional neural network approach. Turkish Journal of Electrical Engineering & Computer Sciences, 24(3), 1248-1264.

[10]. Liu, S., & Deng, W. (2015, November). Very deep convolutional neural network-based image classification using small training sample size. In 2015 3rd I.A.P.R. Asian conference on pattern recognition (ACPR) (pp. 730-734). IEEE.

[11]. Sinha, T., Verma, B., & Haidar, A. (2017). Optimization of convolutional neural network parameters for image classification. In 2017 IEEE Symposium Series on Computational Intelligence (S.S.C.I.) (pp. 1-7). IEEE.

[12]. Potharaju, S. P. (2018). An Unsupervised Approach For Selection of Candidate Feature Set Using Filter Based Techniques. Gazi University Journal of Science, 31(3), 789-799.

[13]. Amiripalli, S. S., Bobba, V., & Potharaju, S. P.: A novel trimet graph optimization (T.G.O.) topology for wireless networks, (2019) doi:10.1007/978-981-13-0617-4_8.

[14]. Amiripalli, S. S., Kumar, A. K., & Tulasi, B.: Introduction to T.R.I.M.E.T. along with its properties and scope. In A.I.P. Conference Proceedings (Vol. 1705, No. 1, p. 020032). A.I.P. Publishing L.L.C., (2016).

[15]. Amiripalli, S. S., & Bobba, V.: Research on network design and analysis of T.G.O. topology. International Journal of Networking and Virtual Organisations, 19(1), 72-86, (2018).

[16]. Amiripalli, S. S., & Bobba, V.:Trimet graph optimization (T.G.O.) based methodology for scalability and survivability in wireless networks. International Journal of Advanced Trends in Computer Science and Engineering, 8(6), 3454-3460, (2019). doi:10.30534/ijatcse/2019/121862019.

[17]. Amiripalli, S. S., & Bobba, V. (2019). An Optimal T.G.O. Topology Method for a Scalable and Survivable Network in I.O.T. Communication Technology. Wireless Personal Communications, 107(2), 1019-1040.z

[18]. Amiripalli, S. S., & Bobba, V.;Impact of trimet graph optimization topology on scalable networks. Journal of Intelligent & Fuzzy Systems, 36(3), 2431-2442 (2019).

[19]. Amiripalli, S. S, V. Bobba, "A Fibonacci based T.G.O. methodology for survivability in ZigBee topologies". INTERNATIONAL JOURNAL OF SCIENTIFIC &TECHNOLOGY RESEARCH, 9(2), pp. 878-881. (2020).

[20]. Ramiah Chowdary, P., Challa, Y., Jitendra, M.S.N.V.: "Identification of MITM Attack by Utilizing Artificial Intelligence Mechanism in Cloud Environments" Journal of Physics: Conference Series, 1228 (1),012044, (2019).

[21]. Thota, J.R., Kothuru, M., Shanmuk Srinivas, A., Jitendra M, S.N.V.: Monitoring diabetes occurrence probability using classification technique with a U.I., International Journal of Scientific and Technology Research, 9 (4), pp. 38-41, (2020).

[22]. Amiripalli, S.S., Venkatarao, R., Jitendra, M.S.N.V., Mycherla, N.M.J.:Detecting emotions of student and assessing the performance by using deep learning International Journal of Advanced Trends in Computer Science and Engineering, 9 (2), pp. 1641-1645, (2020).

[23]. Jitendra, M.S.N.V., Radhika, Y.: A review: Music feature extraction from an audio signal, International Journal of Advanced Trends in Computer Science and Engineering, 9 (2), pp. 973-980, (2020).

A Survey on Electric Power Demand Forecasting

Sandip Ashok Shivarkar[a][1], Dr. Sandeep Malik [a]

[a] *Department of Computer Science and Engineering, Oriental University, Indore, M.P. India*

Abstract: Recently there has been tremendous change in use of the forecasting techniques due to the increase in availability of the power generation systems and the consumption of the electricity by different utilities. In the field of power generation and consumption it is important to have the accurate forecasting model to avoid the different losses. With the current development in the era of smart grids, it integrates electric power generation, demand and the storage, which requires more accurate and precise demand and generation forecasting techniques. This paper relates the most relevant studies on electric power demand forecasting, and presents the different models. This paper proposes a novel approach using machine learning for electric power demand forecasting.

Keywords: Electric demand forecasting, short-term load forecasting, machine learning.

1. Introduction

Electric load forecasting is essential for smooth energy exchange in various electrical domains such as energy generation, energy transmission, energy distribution. The accurate load forecasting helps in the operation and scheduling of a utility company. Load forecasting based on the predicting time span, can be broadly divided into mainly four types: Very Short Term Load Forecasting (VSTLF) forecast demand minutes to hours; Short Term Load Forecast (STLF) forecast demand ranges in days, while Medium Term Load Forecast (MTLF) and Long Term Load Forecast (LTLF) forecasts are longer than a year. Load forecasting methods are categorized into three major groups: traditional forecasting techniques, soft computing techniques, lastly modified traditional techniques.

Due to emerge of new electric appliances and penetration of renewables, the electricity industry has seen ground breaking changes in the energy demand and supply over the last few years in India. This results to over or under estimation of electricity load. Over estimation results in unused power, whereas under estimation causes power deficiency, which results in power outage in rural areas. To tackle this problem, proper methodology with almost accurate solution is required.

---

[1] Sandip Ashok Shivarkar,Department of Computer Science and Engineering, Oriental University, Indore, M.P
Email: shivarkarsandipcomp@sanjivani.org.in

Existing demand forecasting methods doesn't consider local weather, economic growth, class of customer, social parameters, environmental and climate change. This is supposed to be the biggest shortcoming of traditional methods.

Hence a need of novel method is required which would incorporate climatic as well as environmental changes to get more accurate predictions. For improvement in the accuracy of the predicting model, yearly load growth with population growth can be predictable and employed as load affecting factors. The load demand may also increase because of the socio-economic factors, like massive occasion, sporting competition, religious occasions like Diwali, Christmas, Ramzan Eid, Carnivals etc. The load demand varies during weekdays and weekends. The climate decides the typical weather conditions above the defined period in particular fixed area, which also have impact on demand. The weather decides the state of the metrological features like temperature, humidity, wind, rainfall etc. which further decides the load demand.

2.   Literature Survey

Rahma et al. [1] presented the combination of machine learning and big data approach for electricity generation prediction, to predict amount of power required for United States. Here authors proposed data analytic to process power management data which was collected for past 20 years. The back propagation neural network model is used for prediction. The designed framework has three stages, First, processing of raw data and future extraction, second, standardize the data in the organized format using Hadoop then third, training the back propagation ANN for the forecasting. The MAPE is calculated for both, entire generation prediction and separate state prediction. It was 4.13% on behalf of total power generation, then 4-9% for separate state prediction.

Braun et al. [2] proposed multiple regression analysis towards forecast power and gas intake of super market for UK. The authors considered consumption data and weather data for year 2012.The future energy was predicted for the next year. The multiple regression analysis established relationship between dependent variable, energy consumption and independent variables, predictors. The model was assessed by means of normalized mean biased error (NMBE) and coefficient of variation of the root mean square error (CVRMSE).

Almeshaiei et al. [3] presented electric power load forecasting for Kuwaiti electric network. The time series segmentation and decomposition along with statistical analysis was used for decision making using probability plots.

Haque et al. [4] presented performance evaluation for different optimization algorithm of daily forecast. The authors proposed hybrid model on behalf of short-term load forecasting founded on Wavelet transform, Fuzzy ARTMAP plus FireFly. The authors used wavelet transform in conjunction with Fuzzy ARTMAP, whose output was enhanced using FireFly optimization algorithm.

Che et al. [5] presented hybrid model for short term load forecasting. For this authors considered seasonal cycle, where load was changed as per the season. The multiple linear regression was used to analyze time series data through seasonal cycle. To improve effectiveness of multiple linear regression, support vector regression (SVR) with optimal training subset (OTS) was also used. The proposed model was evaluated for California electricity market data with MAPE value 3.5%.

Li et al. [6] presented IoT centered self-learning home management system aimed towards very short-term load forecasting. The authors proposed a long short-term memory recurrent neural network (LSTN RNN) model aimed at electricity price prediction. For clustering the price data k-means clustering was used.

Javed et al. [7] presented the machine learning approach with IoT to improve energy consumption of a commercial building. The system was developed using random neural network (RNN). The random neural network was trained with particle swarm optimization (PSO) also sequential quadratic programming (SQP) optimization algorithms. Training dataset was generated using Fangers equation for PMV. The result shows that energy consumption of the building was reduced by 19.8%.

Li et al. [8] presented short-term load prediction, aimed at electrical vehicle charging station. The authors proposed forecasting model based on convolutional neural network (CNN) as well as lion algorithm (LA). The effectiveness of lion algorithm was enhanced through niche immunity algorithm. Niche immunity lion algorithm (NILA) searches for optimal weights with thresholds of CNN. This system was trained using CNN model.

Daneshi et al. [9] presented long term load prediction for electricity marketplace, region New England. The authors proposed multi-layer perceptron (MLP) for forecasting, to adapt over training in the direction of complex relationships besides fuzzy logic to simulate uncertainties of real data.

Guan et al. [10] presented very short-term load forecasting for region New England. The authors proposed wavelet neural network (WNN) to perform moving forecasting for every five minutes.

Zhang et al. [11] presented index with classification approach aimed at load pattern analysis for large electricity customers in Northern China. The authors proposed k-means plus self-organizing map (SOM) for cluster load curves of customers.

Agrawal et al. [12] presented long term load prediction by recurrent neural network, which consist of long short term memory network. The authors used ISO New England Electricity dataset for years 2004 to 2015. This has total 105120 records by fields like load demand for entire region, dry bulb temperature, dew point temperature then day ahead locational marginal price. The data of years 2004 to 2009 was used for training and years 2010 and 2011 was used for testing. The RNN with LSTM network keeps both long-term and short-term states, so it was resulted in better accuracy. This model has three LSTM layers, each consists of 15 neurons. The output of each layer was given as input to the next layer. At every layer corrected linear unit was used as activation function.

Singh et al. [13] presented short-term load forecasting by ANN through different profiles aimed at weekdays also weekends. The ANN using 20 neurons were used to predicting load of NEPOOL region of ISO New England. The model was trained with hourly data for years 2004 to 2007 and was tested through the data of the year 2008. This proposed ANN prototype has three layers, which are: first input layer, second hidden layer and third output layer. The sigmoid activation function was for hidden layer and linear activation function used for the output layer.

Niu et al. [14] proposed support vector machine and ANN to short-term load prediction. The authors used data of December 2004 of Baoding Hebei Province, which consist of 720 data points. These data tuples were divided into three types of high load, medium load then low load. The ANN was used to forecast load type and SVM was used for load forecasting. The data tuples concerning 1st December 2004 to 25th

December 2004 were used to train ANN-SVM and data tuples concerning 26th December 2004 to 30th December 2004 were used for testing.

Lekshmi et al. [15] proposed time series autoregression integrated moving average (ARIMA) model for short-term load forecasting of 400 kv substation. In this work authors considers one day and one week data samples of the time interval of one minute to train and test the deployed prototype. It was detected that if the number of samples considered are increased, the error is decreased, also error is decreased if the ambient temperature is considered.

Kong et al. [16] proposed a long short term memory centered deep learning framework for short term load prediction based on resident behaviour. In this work authors used AMPds dataset, which consist of minutely current reading of a Canadian house with nineteen appliances intended for a complete year. This LSTM model was trained and tested using almanc of minutely power (AMPds) dataset. Also feed forward neural network (FFNN) with K-nearest neighbour (KNN) implemented to compare results. In this work authors addressed the volatile problem of the resident's activity.

Xie et al. [17] presented long-term retail energy prediction using regression analysis with survival analysis. The survival analysis was used to forecast customers, those will remain with the electricity company. In this work authors developed two models, first one to forecast load per customer and second one customer attrition and forecast tenured customers. To forecast load per customer, the Hong's load forecasting technique was used.

Niu et al. [18] presented hidden markov model aimed at mid long term load prediction. The periodic peak load from year 2001 to 2007 of local grid was used as dataset to train and test hidden markov model. This trained HMM was used for searching variable of interested dataset pattern. The prediction was done through interpolating the adjacent values of these dataset.

Sun et al. [19] presented support vector machine based on mass recruitment with group recruitment continuous ant colony optimization (MG-CACO) model for mid-term and long-term power demand forecasting. The SVM regression was used for finding a nonlinear map after input space just before output space with map the data for a higher dimensional feature space over the map. To select parameters of SVM ant colony algorithm was used.

Imani et al. [20] presented long term and short-term memory network in conjunction with support vector regression aimed at short term load prediction. The authors performs experiments in this work using almanc of minutely power dataset (AMPds). In this work LSTM was intended for feature extraction, whereas support vector regression model was intended for short term load prediction. The authors considers three cases; First, two separate LSTM networks were intended to load and temperature feature extraction. Then SVR was used for forecasting. Second, in addition to first case, load with temperature time series of previous 24 hrs. were specified as input to SVR for forecasting. Third, this LSTM was intended to feature extraction and load prediction.

Hamadi et al. [21] presents linear fuzzy regression model for long-term power demand prediction. The linear regression model was a linear combination of load affecting factors, such as load growth, annual population growth, economic growth, total industrial output etc. of previous year. The authors performs experiments using load data of Liaoning Province between year 1989 to 2007.

Ma et al. [22] proposed short term load forecasting using isolation forest (iForest) and long short-term memory (LSTM) recurrent neural network. Isolation forest

algorithm was intended towards prepossesses the historical data. It clears the anomalies present in the dataset. The historical data was selected from city Power Company for the period 1st Jan 2016 to 7th Jan 2017. To forecast the short-term load, LSTM model was developed. It uses the sigmoid activation function in forgetting gate for control the retention of the state information in the cell.

3.   Limitation of Existing System

The clustering will be more time-consuming task, when the dataset size is very large, so multi-level clustering methods should be developed [11]. The weather parameters and customer class can be taken into account, to improve forecasting results [13]. The large number of load samples are required for more accuracy, using ARIMA load forecasting model [15].

4.   Proposed System

Figure 1 shows the flowchart of the proposed system. It consists of five steps as discussed below.

A.   *Step I- Dataset formation:*
As per the forecasting model the dataset is prepared. e.g. if data is monthly it should be divided as day wise, hour wise.

B.   *Step II- Dataset cleaning amd pre-processing:*
Data is prepared for analysis by removing inaccurate, incomplete, irrelevant, duplicated, and improperly configured data. Then dataset is divided in two parts as training dataset and testing dataset.

C.   *Step III- Model learning:*
The machine learning algorithms are selected and trained using training dataset. For the better accuracy different models are combined together.
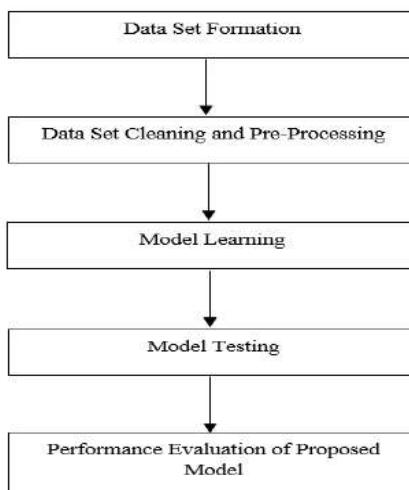
Figure 1: Flow Chart of Proposed System

D.   *Step IV- Model testing:*
     Once the learning is over, model is tested using testing dataset.

E.   *Step V- Performance evaluation of the proposed model:*
     After the testing, the proposed model assessed and evaluated for higher accuracy using different parameters like MAE, MAPE and RME parameters.

5.   Conclusion

This paper relates the most relevant studies on electric power demand forecasting. It emphasizes on the different machine learning models and the data sets used in the era of electric power demand forecasting.

References

[1] M. N. Rahman, A. Esmailpour and Junhui Zhao," Machine learning with big data an efficient electricity generation forecasting system", Elsevier, Big Data Research,Vol.no.5, pp. 9-15, Sept.2016.

[2] M.R.Braun, H.Altan and S.B.M.Beck, "Using regression analysis to predict the future energy consumption of a super market in the UK", Elsevier, Applied Energy, Vol.no.130, pp. 305-313, Oct.2014.

[3] Eisa Almeshaiei and H.Soltan, " A methodology for electric power load forecasting", Science Direct, Alexandria Engineering Journal, Vol.no.50, Issue 2, pp. 137-144, June 2011.

[4] A.U.Haquea, P. Mandalb, J. Menga and R. L. Pinedab, "Performance Evaluation of Different Optimization Algorithms for Power Demand Fore-casting Applications in a Smart Grid Environment", Science Direct, Procardia Computer Science, Vol.no.12, pp. 320-325, Sept 2009.

[5] Jin Wing Che, "A novel hybrid model for bi-objective short term electric load forecasting", Elsevier, Electrical Power and Energy Systems, Vol.no.61, pp. 259-266, March 2014.

[6] Weixian Li, T. Logenthiran, V.T. Phan and W.L. Woo, "Implemented IoT based self-learning home managemt system (SHMS) for Singapore", In proc. IEEE, Internet of Things Journal, Vol.no.5, Issue 3, June 2018.

[7] A.Javed, H. Larijani and A.Wixted, "Improving energy consumption of a commercial building with IOT and machine learning", IEEE, Journals and Magazines, IT Professionals, Vol.no. 20, Issue 5, pp. 30-38, Sept. 2018.

[8] Yunyan Li, Yuansheng Huang and Meimei Zhang, "Short term load forecasting for electric vehicle charging station based on niche immunity lion algorithm and convolutional neural network", MDPI, Energies, Special Issue on Short-Term Load Forecasting by Artificial Intelligent Technologies, Vol.no.11, Issue 5, pp. 1-18, May 2018.

[9] Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.

[10] C.Guan, P.B.Luh, L.D.Michel, Y. Wang and P.B.Friedland, "Very short term load forecasting: Wavelet neural network with data pre filtering", IEEE Trans. Power Syst. Vol.no.28, issue no.1, 2013.

[11] T.Zhang, G.Zhang, J.Lu, X.Feng and W.Yang,"A new index and classification approach for load pattern analysis of large electricity customers", IEEE Trans. Power Syst.,Vol.no.27, Issue no.2, pp. 153-160, 2012.

[12] Rahul Kumar Agrawal, Frankle Muchahary and Madan Mohan Tripathi, "Long term load forecasting with hourly predictions based on long-short-term-memory networks", In proc. IEEE Conference Texas Power and Energy Conference (TPEC), 12 March 2018.

[13] Saurabh Singh, Shoeb Hussain and Mohammad Abid Bazaz, "Short Term Load Forecasting Using Artificial Neural Network", In proc. IEEE Conference on Image Information Processing, International, pp. 159-163, 12 March 2018.

[14] Dong-Xiao Niu, Qiang Wanq and Jin-Chao Li, "Short term load forecasting model using support vector machine based on artificial neural network", in proc. IEEE Conference on Machine Learning and Cybernetics, (ICMLC) International, pp. 4260-4265, 07 November 2005.

[15] Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.

[16] Weicong Kong, Zhao Yang Dong, David J Hill Fengji Luo and Yan Xu, " Short term load forecasting based in resident behaviour learning", IEEE Transactions On Power Systems, Vol. 33, No . 1, pp. 1087-1088, January 2018.

[17] Jingrui Xie, Tao Hong and Joshua Stroud, "Long term retail energy forecasting with consideration of residential customer attrition", IEEE Transactions On Smart Grid, Vol. 6, No. 5, pp. 2245-2252, September 2015.

[18] Dong Xiao Niu, Bing En Kou and Yun Yun Zhang, " Mid long term forecasting using hidden markov model", In proc. IEEE International Symposium on Intelligent Information Technology Application(IITA), pp.481-483, 31 December 2009.

[19] Wei Sun and Wei Zhao, " Mid long term power load forecasting based on MG-CACO and SVM method", In proc. IEEE, International Conference on Future Computer and Communication,Vol.no.1, pp. 118-121, 28 June 2010.

[20] Maryam Imani, "Long short term memory network and support vector regression for electrical load forecasting". In proc. IEEE International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET), 28 October 2019.

[21] H.M. Al Hamadi, " Long term electric power load forecasting using fuzzy linear regression technique", in proc. IEEE Conference on Power Engineering and Automation, pp. 96-99, 19 January 2012.

[22] Yuan Ma, Qian Zhang, Jinjin Ding, Qiongjing Wang and Jinhui Ma, " Short term load forecasting based on iForest LSTM", In proc. 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), pp. 2278-2282, 16 September 2019.

A Wi-Fi Based Smart Irrigation Monitoring for an Agricultural Environment

Nikhil Patankar [a,1], Manoli Charmal [a], Nikhil Bhaskar [a], Swati Janrao [a], Aniket Kamble [a]

[a] *Dept. Of Information Technology Sanjivani College of Engineering, Kopargaon-423601 (MH), India*
.

Abstract. A producer who controls irrigation using a smartphone. A firebase that uses weather data to predict when to water crops. Sensors that read how much water is in the soil. From water availability and unpredictable weather patterns to regulations from outside entities, it can be a challenge to irrigate crops. Nowadays IoT has different solutions to overcome with this kind of problem. The particular research targeted successfully by development effected system using NodeMCU, Sensors, firebase, and assertive application. These tools which could conserve a good amount of water, it will become vital to success as the population groves and water availability dries up. The tool works by helping farmers with limited irrigation capacity determine the best time to water their crops. The main ai m of automating the system is to provide a certain amount of water required by crops by monitoring the moisture of soil and surrounding temperature. This obtains with the help of sensors and NodeMCU for interfacing their values. The values are displayed on a mobile application in real-time using Google's firebase. Irrigation using IoT is a key component of precision agriculture. By changing manual irrigation with automatic valves and systems reduces the human error. Farmer can monitor his crop yield from anywhere at any time.

Index Terms—Iot, NodeMCU, Firebase, wireless sensors

1.   Introduction

India's agricultural sector is on the cusp of a breakthrough technological transformation. After decades of evolution, starting from mechanization and the Green Revolution, disruptive technologies have ushered in Precision Agriculture. Digital Farming is much required in the Indian context as by 2050, India will have an estimated 1.64 billion population1 and will need 333 million tonnes of annual food grain production. Climate change aggravates the situation and threatens to reduce yields of major crops by 25The Wi-Fi-based smart an irrigation system is an IoT based device that monitors the different aspects like soil moisture, Water level, Humidity at a threshold value and then it sends a message to the monitoring part of the proposed system. It provides water supply to crop at the right time, in the right quantity, and at the right place in crop field which plays a vital role in the plant's growth. Monitoring a water level is also a challenging task, especially the management becomes difficult during the shortage of water which may otherwise damage the crop. Using sensors like moisture, temperature, etc.

[1] Nikhil Patankar, Dept. Of Information Technology Sanjivani College of Engineering, Kopargaon
Email: patankarnikhilit@sanjivani.org.in.

water level can be managed easily by analyzing the condition of soil and climate. The sensors which are used in the proposed system like soil moisture sensors smartly monitor the soil moisture and depending on that data, the field irrigated automatically with fewer human mediate

2.   Literature Survey

This proposed system is a smart irrigation strategy using the IoT domain. In this system the sensors are located in the agricultural sector, measuring the soil wetness content, water level of the tank, and the mobile app interface to analyze data and act on results obtained to perform the desired action [1]. This particular paper represents smart irrigation using WSN and water pumps. The proposed system supports aggressive water management for the land. The database connected to the web server analyzes the data which is set at threshold values that are set by the farmer. If the values do not compare with the specified values aware the farmer of the water requirement [2]. The paper is aimed to provide valuable information related to automated smart irrigation. This proposed system reads the Temperature and Humidity, Water level, and soil moisture of soil at a specified value. The main aim is to maintain the nutrient balance of the soil for the overall development of a plant [3]. This paper presents the wireless nodes of an intelligent agricultural system to monitor and controlling crop production framework based on IoT. The sensor node is stored in an agricultural and battery environment with electrical power. The control node is located on the monitoring side. It can look at and offer different prices so that the farmer can easily determine the yield status of their crop as long as there is internet availability [4].

3. Components and Technology used

A.   NodeMCU (ESP8266)

The NodeMCU is an open-source firmware that is built on-chip called ESP8266. It contains the element like CPU, RAM, Wi-Fi and even modern operating system and SDK,NodeMCU refers to the firmware which is the set of instructions programmed on a hardware device. The hardware prototype used on NodeMCU consists of the dual-in-line package with an integrated USB controller onboard with an antenna and MCU.

Figure 1. NodeMCU

B. DHT11 Sensor

DHT11 is used to measure relative digital signal as an output of temperature and humidity as serial data. customized digital module technology, as well as heat and

humidity technology, are used to take care of products that have high reliability and good appropriate balance. The sensor incorporates the opposite sensor of the wet part and the NTC temperature measuring device and is connected by an efficient 8-bit microcontroller.

Figure 2. DHT11 Sensor

C. **Soil moisture sensor**

Soil wetness sensors compute the presence of water in the soil. It looks like a fork-shaped probe with two-conductor performing as a variable resistor. whose resistance changes concerning the water content present in the soil. the better conductivity means that the amount of water present in the soil is decent this result shows in lower resistance. similarly with the poor conductivity which means the volume of water present in the soil is less it will result in case of higher resistance. soil wetness sensor provides both analog and digital output. An output voltage produced by the sensor in terms of resistance, by measuring it helps to determine the wetness level of the soil.

Figure 3. Soil moisture sensor

D. **Relay Module**

A relay is an electronic appliance.it mainly uses to operate as a controlled system that is input circuit or output circuit. It is always working as a control unit for the automatic circuit. Simply, it is an actuator that controls a very high circuit concerning a low current signal. when the relay senses the fault current flowing through the system, it electrifies the electromagnetic field this magnetic field is responsible for opening and closing the connection. hence it plays a vital role in systems implementation for sustain the flow of current

Figure 4. Relay Module

E. Motor Pump

A DC motor is a continuous actuator that converts energy in the form of mechanical energy via electrical energy. The DC motor operated on direct current. A DC motor consists of two parts, A "Stator" is the stationary part and a "Rotor" is the rotating part. The Relay acts as a switching and controlling unit for the motor pump.

Figure 5. Motor pump

F. Breadboard

The breadboard is the basis for building electronic prototyping. Originally, the term referred to a piece of literal bread, a piece of polished wood used to cut bread. In the 1970s a solderless breadboard (a.k.a.boardboard, a terminal array board) was discovered and nowadays the term "breadboard" is often used to refer to these. This makes it easy to use in building temporary prototypes and attempting circuit construction.

Figure 6. Breadboard

G. Connecting Wires

A jump wire (also called jumper wire, or jumper) is an associate electrical wire, or cluster of them during a cable with an instrumentality or pin at every finish point that is generally accustomed to interconnect the parts of a breadboard or different model. These jump wires mainly established the connection between the NodeMCU and various sensors which are used in the system. The jump wire is used to connect point-to-point connections on the breadboard.

Figure 7. Connecting wires

H. Float Sensor

A floating sensor is a type of standard sensor, a tool used to determine the level of water inside a tank. We can say it as a level sensor as it is used to measure the liquid level of the tank. water level float detector or float switches, use the principle of fabric buoyancy (differential densities) to follow fluid levels.

Figure 8. water level float sensor

I. Google Firebase Cloud

The firebase real-time database continuously monitors the real-time data flow of the system. which is stored on a cloud-hosted database in JSON format. The firebase uses data synchronization whenever data changes. It integrates well native with ios , android as well as web. Firebase automatically scales when there's high traffic and scales down when not in use so it is cost-effective it provides security over the data using authentication, we authenticate firebase with firebase id and secret key. Firebase Realtime Database lets allow you to build an interactive application that securely gets access from the database at the user side.

Figure 9. Firebase

J. Arduino IDE

Arduino Integrated Development Environment (IDE) is an open-source software platform used for writing and compiling the code in c and c++ languages. The set of instructions(code) in the Hex file format is uploaded on a microcontroller. It is having a two-part Editor and compiler editor responsible for writing the code and compiler used to compile the uploaded code

Figure 10. Arduino IDE

4.Proposed System

As mentioned in the introduction the implemented system helps the farmer to sustain the quality of the crop by an automated irrigation system. The proposed system is sensing the temperature and humidity, soil wetness content, and tank water level from the farm without any human interaction. The Proposed system used to sense the physical world data using wireless sensors are used in the area to get real-time data, a master node get the data from the sensor and transmit the data to the cloud section, Each node includes sensors like temperature, humidity, and soil moisture, are control by a microcontroller and for water pump relay is used and it works as an actuator. The data received from the sensor of each node is transmitted to the cloud section via NodeMCU. The firebase database represents the sensed data as a result. as per the threshold values defined in the algorithm based upon that motor, the pump is operated and we used an android application to deliver the information associated with the field to the remotely located farmer.

The system consists of the following section.

A.Wireless sensor Node

Figure 11 shows the overall system architecture of the proposed system that includes wireless sensors and the cloud section. The temperature and humidity sensor, float sensor, and soil wetness sensor are sensed real-time data. Data received from the

sensors is processed by a micro-controller and transfer to the firebase database via NodeMCU for further Processing. The threshold values are set according to the crop that needs to be irrigated. The NodeMCU micro-controller works as firmware that is the programmable hardware device. The data obtained from sensors is further transferred to a cloud base database (firebase) via NodeMCU.
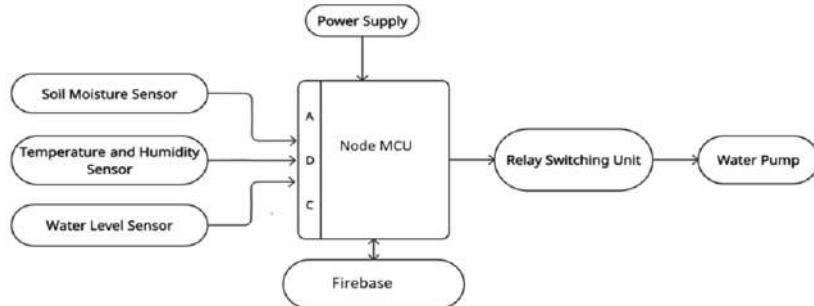
Figure 11. System Architecture

B.Cloud section

At cloud server which is nothing but the firebase used to monitor and store the real-time data. The farmer will Figure out the better decision for better crop yield based upon past data. Figure 12 shows the interaction between the microcontroller and the cloud server. The farmer will able to interact with the cloud server using the android and web application interface. The Decision of whether irrigation is on or off is based on comparisons between soil moisture values, air temperature values, the water level of the tank, and appropriate threshold values. The system controls irrigation according to a decision based on real-time parameters
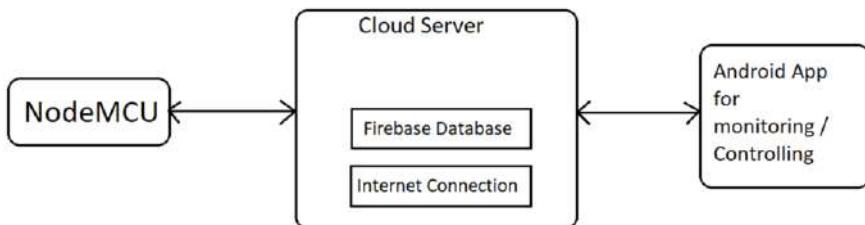
Figure 12. Cloud Server

5. System Implementation

An algorithm designed for the proposed system as mentioned in Figure 13. All information obtained from various sensors and parameters is provided by the NodeMCU microcontroller as an analog input. The pre-determined value of the soil wetness sensor, as well as water height in the tank, is set in the microcontroller. When it exceeds a certain amount of water the irrigated plants are watered and once the required amount of water has been filled, they stop. The microcontroller transmits that information online via the IoT framework in form of the Wi-Fi module attached to it. As the system works itself the irrigation via the motor pump can be switch-on or switch-off with the information provided by the controller. This proposed Irrigation system is used to detect real-time. This method is designed to advance the automatic irrigation process without power by balancing the various boundaries associated with the field and thus improve irrigation.
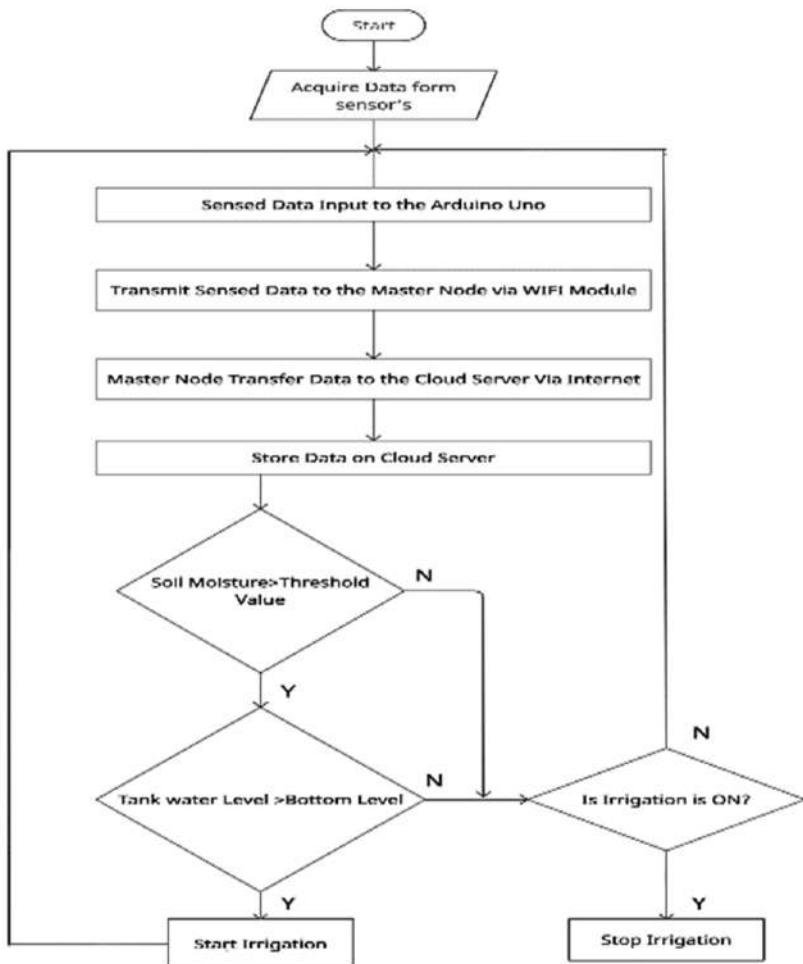
Figure 13. Algorithm of proposed system

6. Results

A proposed irrigation system was tested on a plant in the garden. The water requirement of the plant is 600-800mm per day and the soil temperature requirement is from 40oC - 600oC. In the Arduino code, the humidity and temperature range is set to 300-700 and 450-800 respectively (meaning the corresponding resistance value in digital format). In addition, the program appears to be cost-effective and has the potential to save water and reduce waste. it also indicates the level of water in the tank. In Figure 14 shows a mobile application developed using MIT app inventor which is used to monitor the environment of the area and control the water pump.
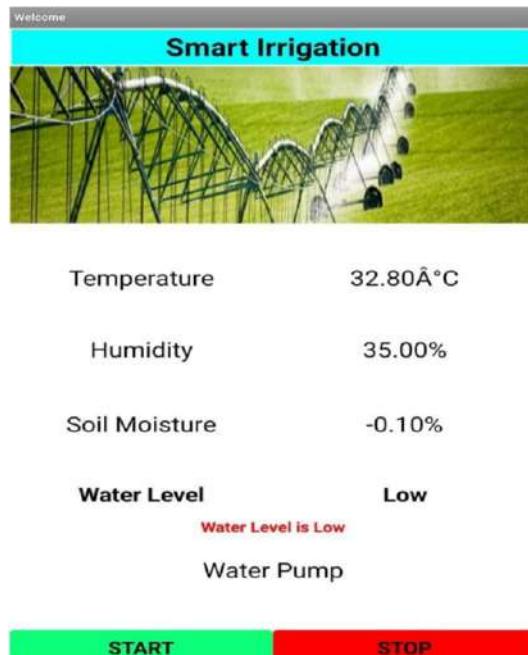
Figure 14. Android application interface

7. Conclusion

The implementation of the proposed system is successfully tested. As a result of this system, it can measure soil moisture, Humidity, Temperature. Using this technology, it is become more efficient to solve farming problems like water shortage problems, soil moisture, and temperature. Farmer can monitor the crop from any place. By building an agriculture IoT platform to comprehensively monitor the growth status of crop it will help to solve the issues more effectively. This project uses a modern database like Google's firebase for fetching data from sensors. Also further proposed

system can also monitor water level by water level sensors it would lead system to have efficient use of water to the crop. Not only this it will also control the consumption of water for irrigation of the field. The on-field research tells us that the results are executed for the better crop and the lessening of water. The system decreases the manpower necessary to field.

8. References

[1]   Shweta B. Saraf Dhanashri H. Gawali "IoT Based Smart Irrigation Monitoring And Controlling System"2017 2nd IEEE International Conference On Recent Trends in Electronics Information Communication Technology (RTEICT), May 19-20, 2017, IndiaPavithra D.M.S .Srinath "GSM based Automatic Irrigation Control System "for Efficient use of Resources and crop Planning by using an Android Mobile "IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE), e-ISSN: 2278-1684,p-ISSN: 2320-334X , Volume 11, Issue 4 Ver. I (Jul- Aug. 2014), PP 49-55.

[2]   Pavithra D.M.S .Srinath "GSM based Automatic Irrigation Control System "for Efficient use of Resources and crop Planning by using an Android Mobile "IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE), e-ISSN: 2278-1684,p-ISSN: 2320-334X , Volume 11, Issue 4 Ver. I (Jul-Aug. 2014), PP 49-55.

[3]   Sensor based Automated Irrigation System with IOT" International Journal of Computer Science and Information Technologies, ISSN: 0975-9646, Vol. 6 (6) 2015, 5331-5333.

[4]   G. H. E. L. de Lima, L. C. e Silva and P. F. R. Neto, "WSN as a Tool for Supporting Agriculture in the Precision Irrigation," 2010 Sixth International Conference on Networking and Services, Cancun, 2010, pp. 137-142.

[5]   Kumar, A., Vengatesan, K., Vincent, R., Rajesh, M., & Singhal, A. (2019). A novel Arp approach for cloud resource management. International Journal of Recent Technology and Engineering (IJRTE), 7(6).

[6]   A. Kumar, K. Kamal, M. O. Arshad, S. Mathavan and T. Vadamala, "Smart irrigation using low-cost moisture sensors and XBee-based communication," IEEE Global Humanitarian Technology Conference (GHTC 2014), San Jose, CA, 2014, pp. 333-337 Sukru Ozan, "A Case Study on Customer Segmentation by using Machine Learning Methods", IEEE,Year: 2018

Exploring Cardiac Disease Based on ML in IOT: A Comprehensive Study

M.Safa [a, 1], *Dr. A. Pandian* [a]

[a] *School of computing, College of Engineering and Technology*
SRM Institute of Science and Technology, Kattankulathur, Chennai
Tamil Nadu, India

ABSTRACT. Every human being in today's age has high importance in a stable human life. With daily new technology is launched in a world loaded with innovative business sectors to tackle more effectively the prevalent challenges of the world. We ought to figure out if as many people as possible can be spared. The death rate of multiple non-communicable heart disease forms is rising steadily each year. Many non-communicable disorders are long-term, gradual and too serious that a patient's situation is too critical to control. This leads to a sudden heart attack for most people or learning about their condition before it is too late. In this document we suggest a cardiovascular state prediction approach for IoT and Machine Learning, which will use IoT system (sensors) to capture the data needed from the human body and move it to the cloud where data is saved with user verification. The obtained information from the human body is then normalized to quantify and forecast the total condition before AI calculations are applied to them.

1. Introduction

In the course of recent years, the innovation business has blast higher than ever and has Different application spaces worked with (for example clinical diagnostics [1], [2] , risk factors selection [3], [4] , robotics [5], [6] and etc.). This has been even more contributed with the emergence of the internet of things [7]. In particular, an intelligent health tracking device developed primarily to check heart rate at regular intervals, taking correct steps depending on the findings. This method supersedes the conventional approach to diagnostics, in which the doctor is bounded to the patient and the appearance of the specialist is the best model for patients' prosperity. Moreover, the best benefit of this scheme is that people are not actually limited to hospitals and physically present. They are tested 24/7 by smartphones and other linked devices (sensor). Furthermore, people are spared from the bother of daily checks on their doctors, giving financial benefits. When you are alone and unable to do something yourself, the most commonly encountered condition in case of heart attacks is.

---

[1] M.Safa, School of computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, India

Email: safam@srmist.edu.in.

The evolved method therefore has the capacity to conduct proper operations, such as calling an ambulance, calling the doctor/family, etc., ensuring that the system can save vital human lives and prolong lifespan times. In addition, patients are better conscious of their illnesses across these programmes and may take an active role in managing the illness to prevent it sooner. According to the WHO, health benefits are the fundamental rights of any citizen and essentially their obligation to reach them, and the IoT position is one of the most significant developments in the field of healthcare.

2. Internet of Everythings

The Network or IoT can be the structure of integrated computing devices, automatic and electrical processes, objects, persons or entities with a similar identifier (UIDs) and the capacity to share statistics through organizations without the need to communicate with people or individuals. It can also be identified as an online object that can be a person with an embedment with a natural display, a grower with a biochip, carriers that have built-in sensors to caution the carrier when the tyre mass is moo, and any additional function or natural question that can be selected as an IP address that can share information over a network. Administrations in a variety of businesses continually use IoT to operate more productively, changing how consumers can maximize their consumer benefit, advance decision making, and raise market esteem. Innovative aspects of the internet of things include remote innovation, micro-electromechanical (MEM) structures [9], micro-services and the internet. The mixture broke the silos between working invention and invention of information to allow for the analyses of amorphous computer-generated data for little information to improve purpose. Although Ashton's was the main sign of the IoT, the concept of the associated appliances has already been under the network and the inevitable computing of this specific name since the 1970's [7]. In an IoT setting, web-enabled main devices are given, which use embedded processors, sensors and correspondence devices to relay and operate entirely on data from their circumstances. Sensor data is collected by IoT systems through communication with an IoT gateway or other edge gadgets, and afterward shipped off the cloud to be prepared or examined locally. These gadgets interact with one another from time to time. The below is a short description of the following types of machine learning problems: 1. Classification 2. Regression 3. Clustering

Classification Algorithm C4.5

For classification, the C4.5 calculation is utilized. Choice trees are constructed from the same amount of training data on each tree node using C4.5 using the knowledge entropy principle as ID3. Then C4.5 chooses one attribute of the data what separates its example set all the more adequately into subsets advanced in one gathering or the other. The parameters are the standard knowledge benefit (entropy difference) that results in selecting an attribute for data division. There are some simple cases in this algorithm.

1. Any example in the rundown is in a similar class. It simply makes an edge hub to the choice tree to choose the class.

2. No intelligence gain from any of the characteristics. In this case, C4.5, using the predicted class value, produces a choice hub higher up in the tree.

3. Occurrence of beforehand invisible class. Again, with the anticipated benefit, C4.5 generates a choice hub in the trees

Data Mining in Neural Networks

A parallelly distributed information processing structure composed of several processing components, i.e., a neural network. They are linked by a one-way signal channel called connections. A single output link is linked to several connections in each processing element. They are regulated learning and unaided learning. This is an essential model where an organization decides and afterward contrasts a response to each information and the objective worth. If the measured result varies from the target value, the network weights are modified in compliance with a learning guideline. e.g.: perceptron of one layer, perceptron multilayer. The Unaided learning systems can benefit from particular characteristics of problems, such as self-organizing maps. The neural network involves nonlinearity, learning capacities, input-yield planning, versatility, verification reaction, disappointment resistance, neurologic correlation, etc

Naïve Bayesian Classifier

The mathematical classificatory Naïve Baye does not presume reliance on attributes to increase the later likelihood in the class determination. Theoretically, although it will not always be the case, this classifier has a minimal error rate. Inaccuracies are due to expectations of class dependency and the lack of probability evidence available. On the basis of the formula listed above and based on such conditional probability data Bayesian classifier computes the contingent likelihood of a case having a place with each class and orders the occurrence as the class of the greatest restrictive likelihood.

Induction from the Decision Tree

The development of the choice tree is completed utilizing the C4.5 inductive calculation. This delivers a choice tree dependent on preparing information which limits the normal estimation of the quantity of information arrangement tests. Any internal tree node is a main component, while a single outgoing branch is a potential scope of that segment. The leaf hubs are the class that is allotted to an examination. The center factor in the C4.5 calculation is its ability to pick the necessary usefulness for every hub naturally. The function of each node is chosen to efficiently separate input samples. The measure of success is the knowledge received. By using a cutting tool, the scale and complexity of trees can be minimized. Pre-tapping and post-taking are the two techniques that are frequently used to carve.

3. Related Work

Neha Anilrao Sarnaik et al. have suggested a method for implanting heart-beat monitor and a system for monitoring heart attacks in which a sensor is attached to a heart rate reader that sends it through the Internet. The device uses two loops one at the end of the patient, the other at the end of the practitioner. A sensor device is used to read and view the data on the LCD screen at heart rate and microcontroller operation [11]. Ingole Abhilasha and at. Proposed the use of Raspberry Pi to track human body temperature and pulse and give the parameters on the B+ Raspberry Pi model to a variety of users. This machine works at -55 cm to +125 cm and the accuracy of DS18B20 is around 0.5 cm [12] In this device, Pic16F77A gathers data that include cardiac beat, temperature and saline levels by means of a sensor using IoT that tracks patients everywhere. Vishal Jagtap et al. This machine uses the Wireless Internet ESP8266 and Gecko IoT modules. A device Healthcare Package using IoT has been launched by Deepika Agarwal et al. This device uses the second-generation development kit of INTEL GALILEO board

4. Methodology

Parameters

The heart conditions of the patients are our biggest concern, because we need to gather data from patients with cardiac disease. In fact, there are 75 parameter numbers from the previous research on cardiac disease. The findings of 14 limitations have proved to be reliable enough to continue the study on 14 only limitations. The researchers have experimented on 14 limitations in Cleveland. In comparison, we have seen a similar outcome for 12 parameters. Thus, the above 12 criteria will be adequate to estimate the patient's heart condition correctly. The following parameters are related to the sensors available.

Parameters Through Manual Input

1) Age: This paper indicates that one out of two people with age 40 are at risk of heart disease, and one in three for women. If we take the age of 70 years into account, the numbers show that for men it is one out of three and one out of four for women [7].
2) Gender: Middle-aged people suffer heart disease more often, especially two to five times higher in men compared to women.
3) Type of chest pain: the most frequent signs of various coronary diseases are chest pain. This is why one of the restrictions is selected.
4) Sleeping pulse: From the paper of J.Gomez[8] and C. Li et al. [9], we get the numbers from which it gives the idea that the vast majority of the patients biting the dust from cardiovascular breakdown have been experiencing Overweight 2.6M - 4.4M, Blood Pressure 7.1M. Which drives us to pick RBP (Resting Blood Pressure) as one of the boundaries.
5) Fat: Raised One of the is cholesterol significant reason for death among individuals saying he has a heart disease by J. Gomez et al. in [6].

6) Fasting blood pressure: J.Gomez [5] and C. Blood pressure. Li et al. [10] indicates the number of individuals who have experienced Overweight of 2.6M-4.4M, Blood pressure 7.1M for most of the deaths of heart disease, which is why RBP is selected as one of the criteria. Which is why RBP is the most significant consideration.

7) Cholesterol: higher cholesterol, said in J, is one of their key causes of mortality in people with heart disease. Gómez et al. in [4].

8) Sugar levels in the blood acceleration: Speeding leads to weight loss and decreasing blood sugar levels, although it is considered normal for less than 100 mg/dL. For this cause, heart prediction has been chosen as a parameter

9) Blood sugar speeding: speeding results in loss of weight or reduced blood sugar levels but is considered normal at less than 100 mg/dL. The parameter was then chosen to make a guess for the patient's heart.

10) ECG rest: Electrocardiograms are ECG's norm. ECG is used for heart rhythm [8] detection. What is significant, since the cardiac condition of a patient must be respected and multiple abnormalities observed. That's ECG's supporting us. In addition, the BPM listening with the ECG can be measured.

11) Optimum heart rate: the most extreme pulse is around equivalent to the heart cadence so allows us to notice the heart work during the day.

12) Exercise involved angina: Angina is a type of chest pain torment that actuates because of the absence of blood supply to the heart. It is a side effect of coronary illness.

13) a previous high: exercise-related sT despair measured across the vertical gap of the iso-electric line to the footprint of the patient.

14) Slope: slope of the ST section of peak practice which is a coronary condition predictive value.

15) CA: a collection of large fluoroscope-coloured vessels that are a prescient worth in a 0 to 3 territory.
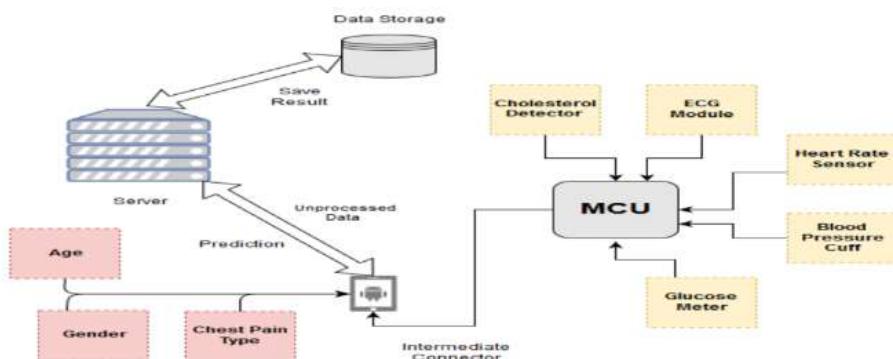
Figure 1. Combined prototype block diagram.
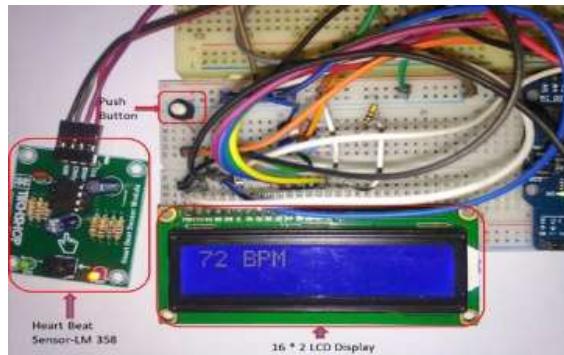
Figure 2. Implementation of a monitor for the heart rate module (LM358).

Affording to the report, those 15 limits are sufficient to quantify and forecast a patient's total heart condition, but we are focusing on the first 8 parameters because of certain This is a worked-on block graph (Figure 1) of the model wanted to be carried out. We suggested using a Microcontroller Module (MCU). The MCU is a microcontroller device is attached to sensors and components in electronics necessary. Using Bluetooth to sending the data to the smartphone after all of the hardware components have been obtained. The mobile also gathers the remaining 3 parameters. Then all 8 boundaries are shared and guided into the cloud running a calculation for ace discovering that gives the client the outcome. This is the real acknowledgment of MCU-associated Sensor Module for Heart Rate (Figure 2). The module tests the information for 10 seconds and delivers the Arduino's simple yield, which is then increased by 6 to acquire the patient's BPM. Constraints and left the exercise for future work, which involves angina, old max, pitch and CA.

The patient's cardiac disease can be demonstrated by repeated variations in heart rhythm or pulse regularity. The measured Beats per minute (BPM) for a patient in our applied module dictates whether the patient's body state is ordinary or whether the respiratory rate is extremely rapid. See the table for more detail for more information on our module. A single ECG module is staffed to calculate the electrocardiogram using 3 electrodes. First, the production pin was attached to an oscilloscope in which we had a signal detected. By analyzing a paper [1], we have correlated our signal received with the actual ECG signal. Then we also visualized the output of the serial plotter with a very promising output waveform. The most confounded piece of an ECG signal is the QRS (top worth), the Q section and the S fragment discovered in the output of our signal Nonetheless, we are attempting to develop and achieve a even better outcome with the ECG signal output. the sensor for Infrared (near-infrared) light is delivered to the human body in order to measure cholesterol without the need for blood sampling [11]. 30 seconds are taken each second, with 30 samples taken in each second. We follow up the detection protocol. We then take the average voltage decrease around the output terminal. We take cholesterol in a balanced person as a normal value of 5.2 mm/L.

Proposed Algorthim

1: Start
2: assign id to the sensor

3: Check if Sensor are connected
4: If connected then
5:      Proceed to step 7
6: else
7:      Establish connection between Sensor and Node Processor
8: for Every Sensor Ss, do
9:      for Every Parameter value do
10:     Read all the value from step 5·,
11:        Publish value check the timing interval
12: analyses the value
13: go to the step 3

5. Results and Discussion

The following are the categories of benefits of IoT: higher reaction times and the capacity to give timely information for decision-making by automation. Improved planning as a result of new insights gained from larger data sets Lower operating expenses. These findings suggest to researchers and practitioners that expected benefits are frequently related to the use of IoT data, such as operational reactions to data generated by IoT sensors, and that IoT, while obviously useful for action-reaction use cases, can provide organizations with far more benefits than purely operational benefits. The values from the nodes are collected and for analysis used various analytics models and addition to that add some intelligence for improved results
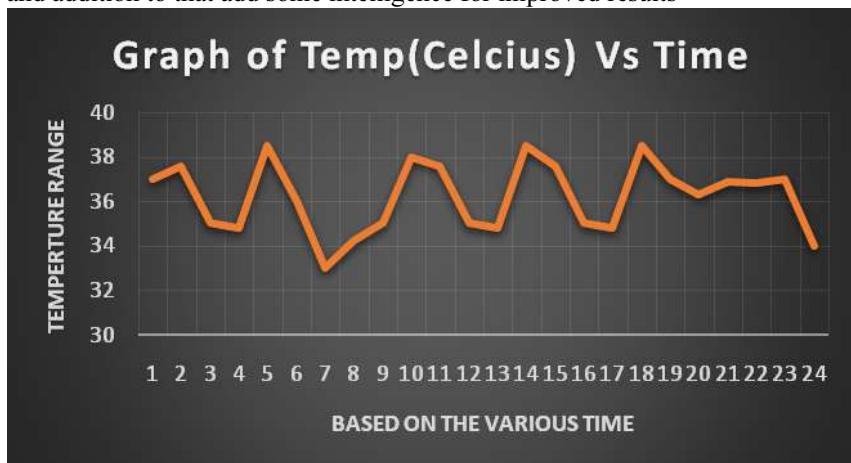
Figure 3. Graph Analysis from Temperature sensor

6. Conclusions

The advances in mobile and remote welfare programmes lead from the determination to the care of various disorders such as cardiovascular diseases in many areas of our lives. Android apps are now a part of the numerous options which are announced as side-by-side treatment initiatives as a variety of wellness monitoring applications. We encountered some challenges during the implementation process. Due to the lack of

supplies in our region, it was difficult to locate the actual IoT devices to enforce our requisite criteria. The feedback was tough to obtain from a photodiode during the implementation of the near-infrared sensor for cholesterol from the human body. After a lot of work, we get a photodiode electrical current and transform it into voltage. On some subjects, we tried our system and compared the true cholesterol values to the end results of a patient's cholesterol Collecting an electrocardiogram (ECG) signal from the human body introduced one more test. From the outset, the outcomes we acquired didn't appear to be in accordance with our prerequisites. Like a measure, the ECG output signal was distorted. substantially different from the previous one.

Acknowledgement

Reference

[1]. F. Akhtar, J. Li, M. Azeem, S. Chen, H. Pan, Q. Wang, and J.-J. Yang, "Effective large for gestational age prediction using machine learning techniques with monitoring biochemical indicators," *The Journal of Supercomputing*, pp. 1–19, 2019.

[2]. F. Akhtar, Y. Pei, M. Azeem, *et al.*, "A semi-supervised technique for lga prognosis," in *Proceedings of The International Workshop on Future Technology*, pp. 36–37, 2019.

[3]. F. Akhtar, J. Li, Y. Guan, A. Imran, and M. Azeem, "Monitoring biochemical indicators using machine learning techniques for an effective large for gestational age prediction model with reduced computational overhead," in *International Conference on Frontier Computing*, pp. 130– 137, Springer, 2018.

[4]. F. Akhtar, J. Li, Y. Pei, A. Imran, A. Rajput, M. Azeem, and Q. Wang, "Diagnosis and prediction of large-for-gestational-age fetus using the stacked generalization method," *Applied Sciences*, vol. 9, no. 20, p. 4317, 2019.

[5]. V. Verma, V. Chowdary, M. K. Gupta, and A. K. Mondal, "Iot and robotics in healthcare," in *Medical Big Data and Internet of Medical Things*, pp. 245– 269, CRC Press, 2018.

[6]. T Karthick, Snigdha Pandey, Gayatri Dutt, Tarunjot Singh Kohli, "Prediction of water quality and smart water quality monitoring system in IoT environment", International Journal of Pure and Applied Mathematics, 2018, vol 118, pp. 3969-3977

[7]. Vijayakumar, K.P., Pradeep Mohan Kumar, K., Kottilingam, K. *et al.* An adaptive neuro-fuzzy logic based jamming detection system in WSN. *Soft Comput* 23, 2655–2667 (2019). https://doi.org/10.1007/s00500-018-3636-5

[8]. M. A. Khan and K. Salah, "Iot security: Review, blockchain solutions, and open challenges," *Future Generation Computer Systems*, vol. 82, pp. 395– 411, 2018.

[9]. Jaree Thongkam, Guandong Xu, Yanchun Zhang and Fuchun Huang 'Breast Cancer Survivability competitive neural networks. Canadian Conference on Electrical and Computer Engineering (Vol. 1, pp.289–293).

[10]. Dilip Roy Chowdhury, Mridula Chatterjee& R. K. Samanta, An Artificial Neural Network Model for Neonatal Disease Diagnosis, International Journal of Artificial Intelligence and Expert Systems (IJAE), Volume (2): Issue (3), 2012.

[11]. M. Anandan, M. Manikandan, T. Karthick, "Advanced Indoor and Outdoor Navigation System for Blind People Using Raspberry-Pi," *Journal of Internet Technology*, vol. 21, no. 1 , pp. 183-195, Jan. 2020.

A Wide Survey on Data Mining Approach for Crop Diseases Detection and Prevention

Vikas N Nirgude [a,1], Dr.Sandeep Malik [a]

aDepartment of CSE, Oriental University, Indore, India

Abstract. India is agriculture land and major revenue manufacturing sector. However, because of amendment in temporal parameters and uncertainty in climate directly have an effect on quality and amount of the assembly and maintenance of crops. Also, quality even a lot of degrade once the crops area unit infected by any malady. The main focus of this analysis in agriculture is to increment the crop quality and potency at lower price and gain profit as result of in India the majority of the population depends on agriculture. Big selection of fruits is growing up in India such as apple, banana, guava, grape, mango, pomegranate, orange is the main one. Fruit production gives around 20% of the country's development. However, because of absence of maintenance, inappropriate development of fruits and manual investigation there has been scale back in generate the standard of fruits.So, Data Mining Approach used in the agriculture domain to resolve several agricultural issues of classification or prediction. During this paper complete survey of several data mining approach for crop disease management has been done. Detection of disease in early state will improve in quality of crop still as decrease the production cost. Also, we can improve the production of the particular crop. Several major parameters are used for the crop disease classification or prediction.

Keywords: Data Mining, Support Vector Machine, Classification, Naïve Bayes Machine

1. Introduction

India is agriculture country; whereas major population depends on it. Indian farmers have a choice to decide their crops. Our farmers work hard to feed large population. But due to drastic change in climate condition crop failure will occur which leads more farmers are committed suicides. So, by using different data mining approaches, can predict the different diseases which occur on early stage so that we can increase the crop quality and quantity. In agriculture fields Vegetables and Fruits are most significant products. To increment the crop quality and potency at lower price and gain more profit as result of in India the majority of the population depends on agriculture due to this main focus is given on this analysis. Wide range of fruits is growing up in India such as apple, banana, guava, grape, mango, pomegranate, orange is the main one. Fruit production gives around 20% of the country's development. However, because of absence of maintenance, inappropriate development of fruits and manual investigation there has been scale back in generate the standard of fruits [1].

---

[1] Vikas N Nirgude, Ph.D Scholor, Department of CSE, Oriental University, Indore, India
Email: vikas.nirgude2006@gmail.com.

In Agriculture quality of seed and soil is important for producing good quality of product. The quality of agriculture product is reduced if crop is infected by some malady. The operation like crosspollination transpiration, fertilization, and germination etc., will affects the plant which may cause the disease. The pathogen like bacteria, fungi, viruses and also because of unfriendly conditions will produce the diseases on the crops [2]. In plants different parts are like fruits, leaves, stem, flowers on which disorder can be found. So, without knowing the proper disease unnecessary pesticides are sprays and used which cause huge amount of air and soil pollution, that also effect on human health. So, it is necessary to predict the right disease depend on the atmospheric condition. Those natural details could be particular climate condition like humidity, temperature, leaf wetness and rainfall. Sometimes experience is insufficient. For specific prediction of potentially fruit plant disease, and right time for protection the fruit plant using data mining. To extract meaning information from huge database the Data mining is the technique used. Data mining is rising exploration field in agriculture plant protection too. Compare the available data with historical datasets as well as previous results of pathogen disease has been recorded [3].

Some factors of soil and environmental parameters like temperature, humidity, moisture etc. are affecting the quality and quantity of fruits. Just by observing the all-environmental parameters manually unnecessary pesticides are used by farmers sometimes which are actually not needed which will increase the production cost and degrade the quality of the fruit. So, there is need of new approach which detect the actual disease of fruit, which lead to improvement in quality and quantity.

2. Literature Survey on Crop Disease Management

Till dates several research works carried by different researchers such as, Dnyanesh Nawale et al [4] proposed machine learning framework with the help of IOT to early detect pomegranate disease, the framework uses Hidden Markov model to uncover disease and alert farmer initial phase. The framework also uses sensors to keep track of temporal details. The above framework required more time for testing and giving the result.
B. Balaji Bhanu et al [5] the temporal details have been taken trough different sensors for improving crop production wireless sensor networks system was proposed. S. R. Rajeswari et al [6] uses random forest method with the help of temporal details crop yield prediction and cost of product will be able to decide. Basically, this approach work for Maize and Rice type crop. Proposed system work has shown on static data set and basically it works on small area.
Geraldin B. Dela Cruz et al [7] uses principal component analysis (PCA) and genetic algorithm (GA) for crop classification. Basically, proposed framework uses crop images of maize. The proposed framework shows the improvement in classification further we can improve the classifying result by applying a novel data mining method.
X.E. Pantazi et al [8] proposed ANN model for detecting wheat yield prediction within field variation. This system uses soil parameters for checking the crop growth, also the factor which affecting the yields has been discussed. The temporal parameters which affect the quality of crop well as yield production is not considered.
Sunil More et al [9] suggested a framework for farmers about pesticides and nutrients that can be used, also it uses wireless sensor network for measuring the different

temporal details. Proposed system shows the better result as compared to traditional methods.

Manisha Bhange et al [10] proposed a framework by using web-based tool to detect fruit disease The proposed system having already processing the images of different diseases based on that image it gives the results. The proposed framework takes the online image as input further it extract the features and compare it with available images. The proposed framework used SVM classifier to classify the disease and according to that it will give the result. As the proposed system will work based on available database images it will not give accurate result for newly introduced disease.

Yuguang Huang et al [11] proposed a Naive Bayes Classification method which work on small data set. Proposed method is most effective in text classification. The proposed framework work well for large training sample set. For large training sample set put heavy work for classification and more storage is required.

Dhiman Mondal et al [12] proposed a framework by combining Image processing and Naive Bayesian classifier.

The proposed method is useful for detect as well as classify the present disease in Okra leaf of YVMV disease. The proposed system only focuses on okra leaf and limited disease prediction and classification.

G. Prem Rishi Kranth et al [13] proposed different machine learning framework for analysis of plant disease prediction. By using the feature of crop such as size, wilting, dryness, shape the disease of the crop is predicted. The above framework further extended to large area and large image dataset. There are several factors are there which will leads to different disease. Also, there is chance in improvement for shadow images because above framework will not give good result for shadow images.

Khumukcham Robindro et al [14] proposed Naive Bayes machine learning framework for detecting the diseases of rice plant. The proposed framework is a future for the detecting common diseases in rice plant may occur during entire life of rice plant. The framework will give the result by doing the assumptions of symptoms and precondition of diseases. Sometimes leads to prediction of wrong disease will affect the quality and also increase the cost productivity.

Chithra, P. et al [15] proposed K-means framework for detecting the defected apple, the proposed framework uses the images for analysis and extract the different features like colour.This system does not take any temporal details which will affect the quality of fruits.

Varughese et al (2016) [16] proposed K-mean framework to detect the apple disease. This framework used the segmentation approach to detect the affected area. The result can be further improved by using Artificial neural network.

Dubey, S et al [17] design a K-mean system for detection and classifying the disease of apple fruit. In this approach mainly focus given on scab, blotch and rot types of diseases. This framework work on images but disease may occur due to temporal details. So, for achieving better accuracy temporal details will be considered.

Mitunkumar Balar et al [18] proposed neural network framework for crop prediction. In this framework rain water and soil parameters such as PH, nitrogen, soil temperatures are considered. In this research focus given on cotton crop. It works on only static data sets and temporal details are not considered while predicting the production

Siddique et al [19] designed a framework by using multiple linear regression and k-nearest neighbour regression. This framework proposed for detecting which crop is

cost effective as well as give maximum yield for this soil parameters are considered. The result is calculated for small area and limited data sets.

3. Challenges in Recent Technology

CR-I. Which method of data mining is more suitable for crop disease management?
CR-II. Which are the major challenges that are faced during crop disease prediction as well as detection?

CR-I: Crop Disease Management by using different Data Mining approaches

Here we discuss various Data Mining approaches based on SVM, Naïve Byes that are used for crop disease
a. Identification
b. Classification
c. Prediction.

   1. SVM approach for crop disease prediction:
   Support vector machines are supervised learning models that analyse data for regression analysis and classification. They come with associated learning algorithms. A Support Vector Machine classifies data by determining the hyperplane that maximises the difference between two classes. It is used in machine learning applications.
   A support vector machine used for classification and give better result if the problem is two class problem but in crop having more number of diseases are there then it not able to classify that problem properly.

   2. Naïve Bayes approach for crop disease detection:
Naive Bayes assumes that all features which we will going to considered are independent, but in real life or for disease prediction or detection most of the disease features are dependent on each other.so again there is challenge if factors are more then it will not give accurate result of disease.
Naïve Bayes will give equal importance to each feature but in case of disease detection or prediction each factor having different values.

CR2-II: Crop disease prediction and identification are difficult tasks.

- Diseases are appearing on crops due to a variety of aspects like soil, temporal details, rainfall, seeds, and weeds because of this take decision of automatic real-time disease management is difficult. Crop diseases and quality are influenced by these factors, also prediction of correct disease is difficult.
- The majority of researchers have only looked at single parameters such as temperature, water, soil factors, Here the agricultural decision-making process takes into account all of them. As a result, by taking into account all of these considerations, crop disease accuracy can be enhanced.
- Due to the dynamic crop production method, real-time field data collection based on these parameters is critical.

- Most of the researchers considered only detected or predicted one or two diseases. But if more than two diseases are there then how to perform correct classification of these diseases.
- Very less research has been carried out on Indian soil therefore in India farmers are not get expected quality and quantity of the crop.
- In India because of the less research on the disease so farmers are not able to produce good quality of the crop even though they are using more pesticide unnecessary cost will get increase as not understanding the proper disease.

4. Analysis of Different Research Paper

Literature review was conducted for different paper which is shown in following Table 1. Based on a review of these research papers and data mining techniques, it was discovered that very little research has been carried out on real-time environmental temporal parameters and crop disease management.

Table 1. Analysis performance of Data Mining technique on Different Dataset

Sr. No.	Paper	Author	Explanation Features/ Algorithm	Application/ Advantage	Gap Identify
[1]	"Comparative Study of Knowledge in Crop Diseases Using Machine Learning Techniques"	P. Revathi et al. (2011)	The techniques of data mining as, • C4.5 • SVM machine learning • ABC algorithm • Artificial neural networks were presented	Application of data mining technique. Accuracy and Performance obtained	A Multidisciplinary approach to combining computational with agriculture can aid in the accurate forecasting and management of agricultural crops.
[2]	Disease Prediction in Data Mining Technique – A Survey	Laharika et al.	1.The training dataset Is used for reference. 2.Different algorithm used such as C4.5, Decision Tree, Naïve Byes, KNN	Here, some statistical tests are performed to find out possible output prediction with the help of some static data set and compared with this technique with data sets	1.Accuracy is different for each sample dataset 2.Can perform the Analysis on image dataset
[3]	"Early Detection of Pomegranate Disease Using Machine Learning and IoT"	Dnyanesh Nawale et al.	In this system Hidden Markov Model is used for crop disease Detection also sensor network, is used.	It predicts disease in early stage	It required more time in testing.
[4]	"Naive Bayes Classication Algorithm Based on Small Sample Set"	Yuguang Huang et al (2011)	For classification purpose Naive Bayes algorithm is used	For Text classification	Get Accurate Result only in case of large dataset sample
[5]	"Application of Data Mining Techniques for Medical Data Classification: A Review"	Saima Anwar et al	Data Mining algorithm KNN, SVM, Naïve Bayes	It predicts medical disease	Accuracy is depending on cleanness of dataset
[6]	"Heart disease Prediction Using fuzzy logic System"	Long,et al	Rough set based on • Attribute reduction • Fuzzy logic system used	Heart Disease diagnostic system	--
[7]	"Decision Support System for Rice Plant Disease Diagnosis using Naive Bayes' Algorithm"	Khumu et al (2017)	Naïve Bayes Algorithm	The designed system is intended for the diagnosis of common diseases in rice plant occurred during the life span	

5. Conclusion

A broad survey of data mining approaches used in the agriculture domain for crop disease management was conducted in this paper. Using the most up-to-date classification, prediction methods in the context of data mining and applying them to age-old agricultural practises would be a novel approach to solving some of the farmer's most pressing issues. According to the survey, the majority of researchers have focused on crop disease identification and classification, but there has been less work on crop disease prediction using real-world data. So by using Hybrid data mining approach we can predict correct disease on real filed dataset. We can increase the productivity and quality of the fruit.

References

[1]. Dr. Kamaljit Kaur, Manpreet Kaur," Prediction of Plant Disease from Weather Forecasting using DataMining" IJFRCSCE | April 2018, Available @ http://www.ijfrcsce.org

[2]. Wilfried Baudoin, Remi Nono-Womdim, NeBambi Lutaladio, Alison Hodder , "Good Agricultural Practices for greenhouse vegetable crops", Food and Agriculture Organization of the United Nations Rome, 2013

[3]. Hui Wang, Jorge Antonio Sanchez-Molina,Ming Li and Francisco Rodríguez Díaz ,"Improving the Performance of Vegetable Leaf Wetness Duration Models in Greenhouses Using Decision Tree Learning" Water 2019, 11, 158; doi:10.3390/w11010158

[4]. Dnyanesh Nawale, Kunal Patil, Rakesh Mahajan, Sona Pawara, "Early Detection of Pomegranate Disease Using Machine Learning and IoT." International Journal of Innovative Research in Science, Engineering and Technology 2018. Issue 6, June 2018.

[5]. B. B. Bhanu, K. R. Rao, J. V. N. Ramesh and M. A. Hussain, "Agriculture field monitoring and analysis using wireless sensor networks for improving crop production," 2014 Eleventh International Conference on Wireless and Optical Communications Networks (WOCN), Vijayawada, 2014, pp. 1-7, doi: 10.1109/WOCN.2014.6923043.

[6]. S.R.Rajeswari , Parth Khunteta, Subham Kumar,Amrit Raj Singh,Vaibhav Pandey,"Smart Farming Prediction Using Machine Learning" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-7 May, 2019.

[7]. Geraldin B. Dela Cruz, Member, IACSIT, Bobby D. Gerardo, and Bartolome T. Tanguilig III,"Agricultural Crops Classification Models Based on PCA-GA Implementation in Data Mining",International Journal of Modeling and Optimization, Vol. 4, No. 5, October 2014

[8]. Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.

[9]. S. More and M. Nighot, "An agro advisory for pomegranate field using wireless sensor network," 2016,International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, 2016, pp. 557-562, doi: 10.1109/ICACDOT.2016.7877647.

[10]. Manisha Bhange, H.A.Hingoliwala,"Smart Farming: Pomegranate Disease Detection Using Image Processing." Second International Symposium on Computer Vision and the Internet IEEE 2015.

[11]. Y. Huang and L. Li, "Naive Bayes classification algorithm based on small sample set," IEEE International Conference on Cloud Computing and Intelligence Systems, Beijing, 2011, pp. 34-39, doi:10.1109/CCIS.2011.6045027.

[12]. D. Mondal, A. Chakraborty, D. K. Kole and D. D. Majumder, "Detection and classification technique of Yellow Vein Mosaic Virus disease in okra leaf images using leaf vein extraction and Naive Bayesian classifier," 2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI),Faridabad, 2015, pp. 166-171.

[13]. Potharaju, S. P. (2018). An Unsupervised Approach For Selection of Candidate Feature Set Using Filter Based Techniques. Gazi University Journal of Science, 31(3), 789-799.

[14]. G. Prem Rishi Kranth, M. Hema Lalitha, Laharika Basava ,Anjali Mathur,"Plant Disease Prediction using Machine Learning Algorithms." International Journal of Computer Applications (0975 – 8887) Volume182 – No. 25, November- 2018

[15]. Chithra, P. L., & Henila, M. (2017)," Defect identification in the fruit apple using k-means color image segmentation algorithm", International Journal of Advanced Research in   Computer Science, 8(8).

[16]. Suraj Khade., et al(2016),". Fruit quality evaluation using k-Means clustering Approach", International Journal of Advances in Science Engineering and Technology,4(2),2321-9009.

[17]. Arlimatti, S. R. (2012),"Window based method for automatic classification of apple fruit", International Journal of Engineering Research and Applications, 2(4), 1010-1013.

[18]. Mitunkumar Balar1 Ronakkumar Patel2," Crop Production Estimator", IJSRD - International Journal for Scientific Research & Development| Vol. 6, Issue 11, 2019 | ISSN (online): 2321-0613

[19]. T. Siddique, D. Barua, Z. Ferdous and A. Chakrabarty, "Automated farming prediction," 20 Intelligent Systems Conference (IntelliSys), London, 2017, pp. 757-763

[20]. P. Revathi , R. Revathi and Dr.M.Hemalatha ,"Comparative Study of Knowledge in Crop Diseases Using Machine Learning Techniques",(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (5) , 2011, 2180-2182.

Enhancing Agricultural Product Supply Chain Management Using Blockchain Technology: Concept

M A Jawale [a,1], A B Pawar [a]

[a] *Department of Information Technology, Sanjivani College of Engineering, Kopargaon, India*

Abstract. In traditional agricultural supply chain management, due to involvement of many stakeholders in the entire procedure ranging from farmers, retailers to final vendors, it is merely the producer of the agricultural products i.e., farmers get its direct benefit. The middle stakeholders are always more beneficial than farmers and hence, the conditions of farmers are always the same though they sometimes get good earning and it is becoming a very serious concern in India. The major reason behind it, the transparency and traceability of the entire supply chain of this agricultural products journey from farm to vendor shop. In addition, consumers are becoming more conscious of where their food and food products come from. Block chains have distributed ledger technology (DLT) which has potential to provide transparency and trust for agricultural product supply chains at its different stages and even useful for improving its efficiency. This can boost confidence of all stakeholders who are involved in this farming supply chain. This research paper proposes the same concept in its subsequent sections.

Keywords - *Agricultural Supply Chain Management, Blockchain Technology, Distributed Ledger Technology, Transparency, Traceabiltiy*

1. Introduction

In today's unpredictable world, everyone is now becoming more health conscious and always curious to know about his or her food details. Mainly, with agricultural food products; consumers want to know where their food comes from, when it is processed, how it is best for their health, freshness of it, and other parameters too. Therefore, many agricultural related businesses are in search of good supply chain management technology-based solutions to enhance food quality, food safety and its traces throughout the entire supply chain of agricultural products. In the agriculture field, there are so many existing technologies in place namely precision farming, crop management systems, IoT based solutions, transport systems which are improving this sector related businesses to get better outcomes related to food production and farming related supply chain. Increase in food demand brings new problems with it like fake or forged agriculture products, which are hovering this supply chain at its different levels. Lack of a properly secured mechanism to ensure transparency and poor efficiency lead towards disadvantages at farmer as well as at consumer's side.

---

[1] M A Jawale,Department of Information Technology, Sanjivani College of Engineering, Kopargaon
Email: jawale.madhu@gmail.com.

Eventually, block chains have distributed ledger technology (DLT) which has Potential to provide transparency and trust for agricultural product supply chains at its different stages and even useful for improving its efficiency. This can boost confidence of all stakeholders who are involved in this farming supply chain.The proposed research work is intended to take advantage of block chain technology and want to bring transparency in farming supply chain management, which will improve trust and will give its benefits from farmers, market players to consumers.

2.   Origin of the Research Problem

The farming supply chains are complicated as well as risky as compared to any other existing business supply chains as its product production relies on natural factors like climatic conditions, pesticides and various diseases -which are difficult to trace, predict and control. This uncertainty brings lack of transparency in this supply chain, Even it requires a lot of manual work to maintain it which slows down financial transactions. Additionally, fake operations may cause at each stage of this supply chain and result in destructive damages to all of its stakeholders including farmers, government, market players and consumers too. As per reported in [1], at each stage of agricultural product supply chain management, there are very serious counterfeit occurs. At manufacturing level, it occurs usually at seed supply, chemical supply and brings huge losses to manufacturers. For farmers, it causes a decrease in crop production, health issues due to fake chemicals in pests, financial losses due to untested fertilizers. Even, for consumers it results in huge losses like decrease in revenue, public health issues, price variations, and unemployment.

This leads towards the proposed research problem as to how to ensure transparency, efficiency in agricultural products supply chain management from seeds to grocery shops and provide a chance to use the potential of block chain technology to encounter it.

3.   Interdisciplinary Relevance

The proposed system has a large number of applications ranging from agricultural field to data science, machine learning applications which can use this use case for extending its usefulness.

4.   Review of Research and Development in the Subject

In today's digital world and in continuous population growth, consumers demand services on digital platforms. The consumers are now more interested to know about their foods [2]. Therefore, farmers and agricultural related businesses should use this technology to bring transparency, financial performance improvements and at the same time, it should fulfill demand for the food with high quality. Therefore, the agricultural sector needs to adopt technological changes as early as possible to meet requirements of all stakeholders who depend on the agriculture field and in its supply chain. The major challenge in the agricultural supply chain is its food transportation. With the help of techniques like barcodes, cryptography movement of each product transport and shipping can be observed. The physical movement of this product needs an immutable product and its supply chain process link. Following are existing technologies, which are currently in use to trace the physical movements of the product [3].

Use of QR codes on the products and on their packaging: This technology is in place and in use for agricultural products in the supply chain. It needs more regulations to monitor and authenticate from seed bags to a ready grocery product.

Radio Frequency Identifier (RFID) based supply chain management system.

Contactless technology such as NFC agriculture technology IBM crypto anchor for product authenticity [useful but still in its initial development stage]

Blockchain technology is able to trace information at each stage of the farming supply chain and is useful to bring transparency in the system. It provides a secure way to manage and handle data in a distributed, immutable manner which ensures trust on the system and increases system efficiency too. Due to this promising potential of the Blockchain of information traceability, data driven innovativeness, safety; many organizations have been introduced to solutions for food security. The details of few prominent solutions are described in subsequent sections in connection with the proposed research work.

IBM Food Trust: Big IT giant, IBM introduces it as a smart food supply chain which claims support food safety, food quality and traceability. It has multi-faceted features to bring a food ecosystem on a smart platform. [4]

The other proposed solutions and their reviews with current status of these solutions are available in [5]. Most of these solutions are still in the preliminary stage of implementation which includes agricultural insurance, smart agriculture, food supply chain, e-commerce of agricultural products. Also, researchers explore the merits and demerits of Blockchain technology for its effective usage in future research work.

India has its own Blockchain native platform Eleven01 and it has partnership with The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and KHETINEXT, a mobile based agricultural solution provider for BlockChain Technology use.

The joint venture of these partners is enabling BlockChain Technology to enhance crop productivity and to raise income for small farmers in India.

This solution is in its initial stage of development [6].

5.   Significance of the study

The major advantage of this study involves understanding the farming supply chain and its impact on Indian agricultural products traceability from seeds to grocery ready products and how it can be ensured with the potential of the Blockchain Technology with its usage beyond crypto currency to useful technology to enhance lives of farmers in India.

6.   Objective and Research Methodology

The proposed research work mainly focuses on following objectives during this research work tenure.

- To perform a detailed literature survey which will leverage the current state of the agricultural product supply chain system, its pitfalls and impact on all stakeholders from farmers, manufacturers to consumers. In addition, it will give in depth knowledge about Blockchain Technology for implementing the proposed research work.

- To design a solution based on Blockchain Technology for identifying counterfeit in the agri. product supply chain through the following four steps
  - Data generation based on IoT at Farm or collection of required crop related details from farmers.
  - Supply of grown crops to the food processing industries.
  - Supply of processed agricultural products to wholesalers and retailers.
  - Consumers can trace the product details and can trace back the supply chain to get information about food products regarding its location details, processing details and quality details.
- To evaluate the performance of a developed system on suitable evaluation parameters.

7. Research Methodology

As stated in the Objective section, the following methodology will be applied to bring the proposed solution in real implementation based on stated steps. i.e.
  - Data generation based on IoT at Farm or collection of required crop related details from farmers.
  - Supply of grown crops to the food processing industries.
  - Supply of processed agricultural products to wholesalers and retailers.
  - Consumers can trace the product details and can trace back the supply chain to get information about food products regarding its location details, processing details and quality details.

The following steps depict the proposed research methodology for the selected research problem solution:
1. Crop details will be collected in Smart Contract of the Public Blockchain either IOT enabled network or through mobile phone application.
2. These entered details will be processed using machine learning algorithms to predict the quality of crop.
3. Details of the quality of crop will be available for Agricultural industries for bidding purposes through its own smart contract.
4. Once a bid is accepted, the payment details will be entered into a smart contract and the concerned crop load will be picked up and will be brought into the agricultural industry for its further processing. Even, auto payment transfer will be available to farmers.
5. The processed agricultural or food product details will be maintained in its own smart contract and traceable to wholesalers for bid.
6. Once a processed agricultural product bid is passed, its payment details will be entered into a smart contract to do auto payment transfer to agricultural industries and shipment will be processed to the concerned wholesaler.
7. At last, once consumers get the agricultural or food products from a wholesaler or retailer; he or she can scan barcode details and can trace contracts for exact location and other details instantly.

8. Similarly, government agencies or authorities can trace out each and every transaction, costing and other information through these implemented contracts of BlockChain Technology.

8.   Conclusion

The proposed supply chain traceability system based on BlockChain Technologies can transform the food supply chain from farmers to consumers in the following manner and can provide transparency, trustworthiness and efficiency for agricultural product supply chain systems.

- Simplification and reformation of all steps in the agricultural product supply chain management system.
- Possible tracking and exact food product information along with its complete path information from farmland location to grocery store shelf.
- Elimination of counterfeits in the supply chain and enhance food safety.
- Direct Access to agricultural related financial services to farmers and related businesses.
- Getting smarter agricultural data, for better decision making
- Certification for legal authorities involved in agricultural businesses.

References

[1]. The report on "Counterfeiting in African Agriculture Inputs – Challenges & Solutions", Bill and Melinda Gates Foundation in collaboration with Monitor Deloitte, and was used to facilitate a workshop in Nairobi, Kenya on February 14, 2014.

[2]. Shaosheng Jin Yan Zhang, Yining Xu, "Amount of information and the willingness of consumers to pay for food traceability in China", 2017.

[3]. https://www.intellias.com/crop-management-software-for-sustainable-farming/ Accessed Date: 25/12/2020

[4]. https://www.ibm.com/in-en/blockchain/solutions/food-trust, Accessed Date : 25/12/2020

[5]. Hang Xiong, Tobias Dalhaus, Puqing Wang and Jiajin Huang, "Blockchain Technology for Agriculture: Applications and Rationale", In Front. Blockchain, 21 February 2020 https://doi.org/10.3389/fbloc.2020.00007

[6]. https://www.downtoearth.org.in/news/agriculture/high-tech-on-the-ground-blockchain-te chnology-to-boost-small-farm-incomes-, Published Date: 28 May 2019

A Review on BCI Emotions Classification for EEG Signals Using Deep Learning

Puja A. Chavan [a,1] and Dr. Sharmishta Desai [b]

ªSchool of Computer Engineering & Technology, MIT World Peace University, Pune,Maharashtra, India
ᵇAssociate Professor, School of CET, MIT World Peace University

Abstract. Emotion awareness is one of the most important subjects in the field of affective computing. Using nonverbal behavioral methods such as recognition of facial expression, verbal behavioral method, recognition of speech emotion, or physiological signals-based methods such as recognition of emotions based on electroencephalogram (EEG) can predict human emotion. However, it is notable that data obtained from either nonverbal or verbal behaviors are indirect emotional signals suggesting brain activity. Unlike the nonverbal or verbal actions, EEG signals are reported directly from the human brain cortex and thus may be more effective in representing the inner emotional states of the brain. Consequently, when used to measure human emotion, the use of EEG data can be more accurate than data on behavior. For this reason, the identification of human emotion from EEG signals has become a very important research subject in current emotional brain-computer interfaces (BCIs) aimed at inferring human emotional states based on the EEG signals recorded. In this paper, a hybrid deep learning approach has proposed using CNN and a long short-term memory (LSTM) algorithm is investigated for the purpose of automatic classification of epileptic disease from EEG signals. The signals have been processed by CNN for feature extraction from runtime environment while LSTM has used for classification of entire data. Finally, system demonstrates each EEG data file as normal or epileptic disease. In this research to describes a state of art for effective epileptic disease detection prediction and classification using hybrid deep learning algorithms. This research demonstrates a collaboration of CNN and LSTM for entire classification of EEG signals in numerous existing systems.

Keywords. Deep learning, machine learning, DCNN, signal processing, social data analytics

1. Introduction

Nowadays the subfield of Brain Computer Interaction (BCI) encompasses everything we do with our Human Computer Interaction (HCI). It provides the ability to connect to electronic devices such as a computer and a cell phone with the human brain. BCI has played a significant role in helping disabled people. An EEG-based BCI framework involves a user who interacts with the hardware that employs other technologies as well. In the previous expandable BCI, several processing steps were taken to find the reason for brain signals and transform them into something that the user wanted to do. BCI techniques collect signals from the brain tissue, learn about these signals, and use this information to identify the subject's intentions.

[1] Puja A. Chavan, School of Computer Engineering & Technology, MIT World Peace University, Pune, Email:apujachavan02@gmail.com.

Similar to the imaging studies, electrodes may be used for non-medical purposes such as for sports, educating, tracking, and for entertainment purposes as well. Feelings are extremely important in human cognition, especially in rational decision-making and interpretation, in addition to helping people communicate and help them gain understanding. Recent affects-related computing appears to have caught up to HCI by linking emotions and systems with HCI, with this as well as other fields of computing Measuring emotional states using human-computer interactions refers to ascertains consumer engagement with the technology People may have an emotional awareness of other than a reaction to stimuli in their bodies, particularly for their emotional memories. psychology, neuroscience, and computer science continue to collaborate to improve analysis of emotion recognition [link to brain and thought processing working together] Curious to see how the algorithms of modern artificial intelligence could be applied to any form of modern society, modern militaries, manufacturing, and academic have always demonstrated and fantasy on computer screens. Alternative strategies include those that use emotional behavioral patterns such as well as well as facial and voice cues, including movement and tone of voice, but there are also the ones that employ additional factors, though less definite, clues like gut feelings and vague feelings as identifiers. Noninvasive sensors may typically record biological activity as well as well as electrochemical and/electrical signals. These models include conductivity, electrocardiogram, and electrocardiogram measurements.

1.1 Overview of Deep learning:

It is the sub-branch of machine learning that makes the machines smart enough so the machines are capable to learn from past experience and understand the real-world concepts. When a computer gathers more real-world intelligence, it becomes better at making decisions on what to do base on new experiences [4].

In Deep Learning, "deep" indicates the number of hidden layers in the Neural Networks. Deep Learning models are trained by using any large set of labeled data. Deep learning techniques are used for image sentiment analysis and providing optimum results. With increasing investment in deep learning, a lot of such approaches such as Convolutional Neural Network, Deep Belief Network, Deep Belief Network, and Deep Learning play an important roles in providing algorithms that utilize various features including word embedding, word embedding, word co-oriented belief, and connectionist analysis, to deliver optimal results, state of the art results It is also expressed as image and text here: When people have differing emotional conflicts, undesirable, or incommoded feelings, the main issue we see is in our studies is that these feelings are expressed in both image and text [5].

2. Review of Literature

A wide range of signal processing and machine learning experiments have been done on seizure recognition, many of which have included studies of the convolving features. the authors isolated the frequencies in [1] using the wavelet transform (WT), which provides an understanding of the frequency content of the signal, and its standard deviation, and the Lyapun's characteristics, as two sets, and then measured the signal's three main distinguishing features: variance and coherence. Depending on the method, several different methods have been found to be more suitable for seizures:

wavelet analysis and the wavelet-based network learning system. in the article titled "Classifying Learning with Fourier Expansions, a Lifts, in [3], features were extracted with the FFT, and classification trees were used to analyse these features PCA review found that these two methods to be applicable to various models: Principal Component Analysis (PCA) and Genetic Algorithm (GA) were usable for all models of different types, from linear to non-linear. Computation of the initial principal variable using a non-linear algorithm came up with better results than a genetic algorithm. according to the paper, which showed a comparison of wavelet-based techniques for seizure detection, a 2004 saw in [4]. Expanding Spectral Epileptosis and Detection of Epileptiform Pulses and Training and Testing Sets" (also known as "expanding spectral epilepsy extract and test feature set extraction"), authors listed training and testing sets for recognizing and classifying seizure segments by extracting additional spectral features." Classification SVMs are helpful for purposes of identifying strong support vectors, for instance.

Using a strategy of the empirical mode decomposition of signals from the EEG to do the study It was used to pull out the real-valued components in the same way as the Hilbert–Huang transform (HHT). These functional capabilities were applied as features for the differentiation of EEG signals. using distinct brain regions, as well as groups of ECG cycles, [5] as opposed to individuals, according to the study [6] using distinctive regions of the brain, expanded the research in [8] used the non-non-linear dynamics of the Lyapunov exponent, and largest Lyapunov wavelet-based method to measure the EEG signals, both calculated from correlation coefficients, for study. This model used logistic trees to derive optimal seizure detection algorithms that draw on statistical features and incorporated a statistical detection technique that utilises the logistic feature space. Some statistical features were used as input to a logistic model (LMT) for the epileptic seizure recognition. This algorithm was used to validate the accuracy of the presented approach using an EEG dataset. Some of the papers previously discussed illustrate various classification objectives and feature extraction strategies. These models have been created for the systems they are considered to be essential features of the systems. Since the researchers in the paper [9] employed the discrete wavelet transform (DWT), the brain wave signals were subdivided into frequency bands and thus made into a collection of discrete statistics in [9] were obtained In order to reduce the amount of data, the PCA, independent components analysis (ICA) and linear discriminant analysis (LDA) were utilised. Once the extracted features had been added, SVM was used to distinguish non-epileptic seizures, the model got better at identifying epileptic seizures [10].

Instead of drawing an arbitrary pattern, interictal EEG data points were used in [in 11] to form a data Figure [a universal universe Figure] in [as a random.]. using feature extraction techniques and an implementation of Universum SVM, [Feature Vector Machine (SVM) Classifier], they were able to Structure out of important features in [12] from the EEG data was developed and made apparent, using NCA (neighbourhood component analysis). Also, they tried out AdaBoost (adaptive boosting), SVM, and K-NN classifiers to determine the device efficiency of the system. According to the article [13 14], a deep learning neural network was used to examine the EEG data. Using CNN, the features were extracted and they were assigned to each of the three types of seizure classifications, before, during, and after the seizure.

In the detection of real-time seizure activity from intracranial EEG signals was applied with a new method called machine learning, which used the prior seizure observations to arrive at an optimal or expected seizure detection threshold. We had to get

spectra and time stamps for each photo and extract temporal and spectral features to train the pattern recognition. Utilizing linear discriminant analysis and wavelet wavelet transform-based features, [15] as well as triadic wavelet transform discriminants, K-NN classifiers were used to identify the seizure signals the main goal of this study is to discover if seizures are present due to epilepsy. These systems provide means of intercession for those who were never affected by the seizure to be monitored as well as systems that surgery or treatment has zero effects. Because of their architectures, above-constructed systems also require feature extraction and classification steps, the word we're looking for can be found using a method known as a multi-layered query The methods used in this research, for seizure identification on the scalp, were used in CNN as well to locate the seizure originating in the body [16] [17]. CNN works with unsupervised learning that automatically looks for representation layers within a collection of feature databases [that have been created by previous layers of processing] and learning features naturally within the CNN, as it learns to identify them an empirical study of artificial intelligence, known as the deep learning, has confirmed its ability and shown itself to outclass human cognition on the difficult problems of image and audio recognition [18,19]. A related to machine learning has been applied in a number of diverse ways, such as diagnosing Alzheimer earlier, calculating the degree of illness, and determining the severity of diseases in the chest, all of Algorithms, and strength assessment.

According to MansiAgrawal et al. [21] proposed a system of hand gesture identification using deep CNN, this module they proposed a collect the streaming data from a live webcam and converted it into various frames. Then using preprocessing and gesture sign recognition, the final signal has been indeed using soft computing techniques. According to Swati Avad and DrSharmishta Desai [22] proposed a system of Indian sign language recognition using machine learning techniques; this system is the collaboration of image processing techniques and machine learning approaches. The initial image processing techniques, life segmentation, histogram generation, feature extraction and selection and finally, classification process has done for final detection.

According to Meet Patel, Sharmishta Desai [23] proposed a system music recommendation for paralyzed peoples; this system author has collected a brain EEG signal in the runtime environment and classify the emotional using deep learning-based CNN algorithms. This system has evaluated synthetic and teal time signals data that produces high accuracy over the traditional machine learning algorithms.

According to DrSharmishta Desai [24], proposed student behaviour analysis using dynamic induction module, The decision tree algorithm visualizes data, which is beneficial for social media data analysis. Since they store all data in memory for constructing a model, neural networks like C4.5 or CART have a memory designed to regulate. As a result, these algorithms are ineffective when dealing with large amounts of data. When the amount of data is small, these algorithms perform well, but the algorithms perform poorly as the amount of data grows. We used the Hoeffding tree for a large volume of data in this paper, and the results show that the Hoeffding tree outperforms other Data Mining algorithms.

DrSharmishta Desai [25] proposed a framework for analyzing consumer data from social media sites that compared the performance of various supervised learning models to market data. The Clustering method with one-level logistic regression is found to work the best against trade data. They used AdaBoost to enhance the productivity of decision trees, and the results were impressive. The various stages of large-scale data extraction are also thoroughly clarified.

3. Proposed works

In proposed research work evaluate the entire system with supervised learning algorithms, initially collect the data from the brain as EEG signals. To extract the various features from inputs Set and generate the trained model accordingly from CNN and LSTM algorithm. The objective of the system to identify the epileptic disease based on given ECG signals. In the testing system classify each input signal with respective labels and show the system efficiency.
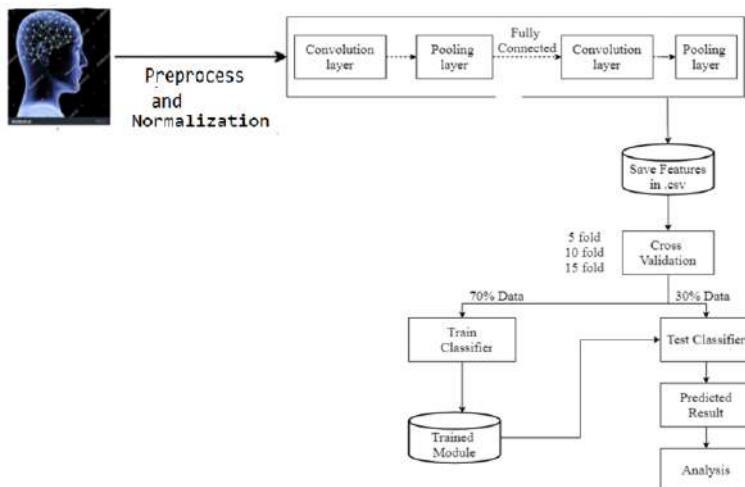
Figure 1. Proposed system overview

Features from the training data have been examined and training models have been created, which has the various specific capabilities as a result. Another image feature extraction technique is implemented on the testing dataset, which extracts each feature only if it'svalue exceeds a certain threshold. The feature weight identification method involves determining the similarities between testing and training conditions. When comparing two features sets, we go deeper into the data in order to discover what they have in common. The weight factor takes values between which corresponds to a person's preference for objects and attaches them to those values. Initial weight is 0, while the threshold can be system defined.

CNN-LSTM classifier has used for classification of epileptic disease in proposed architecture. It is a supervised learning model, used mainly for study on classification and regression. The classification frequently used for challenges regarding object recognition, performs particularly well in aspect-based applications recognition, and classification based on colour. Encourage vector computers aren't better than others machine learning approaches; however, they do at the state-of-the-art, and with a lot of current both conceptual and empirical. CNN-LSTM is known as by many researchers as a better algorithm for performing rankings. The EEG Input signals the values are entered in the classifier and the epileptic disease are classified by hyper plane using drawn with the epileptic disease identified.

4. Observations

- From the above literature review, several systems deal with deep convolutional neural networks, and various boosting approaches have been used to improve computational complexity.
- To achieve a better time and space complexity, a convolutional neural network and various types of boosting methods were used.
- To classify the sentiment of stratified image datasets, researchers used region-based convolutional neural networks and other computer program algorithms such as PNN, RNN, and DCNN.
- The features were extracted using the ImageNet library, and the transport model was built accordingly. It can be difficult to achieve better precision than CNN using PNN, RNN, and other techniques.
- It is proposed that a combination of fast recurrent neural networks and CNN yield the highest accuracy with the least amount of time complexity.
- When a multi convolution layers is formed, it takes a long time for CNN to produce, and it also takes a lot of data when the system would be dealing with heterogeneous datasets. Irrelevant features are removed by function extraction techniques, resulting in high dimensionality problems.
- Several current studies on CNN have found that the average accuracy for text categorization on the flicker picture dataset is about 96.50 percent.

5. Results and Discussions

Getting equipped to efficiently perceive emotions would be Important for decision support structures such as Mental Health conditions which is a method for controlling risk and protection used by Practitioners in mental health and people for evaluation. Emotional evaluations, independent of self-reporting would help calibrate and guide the advice further delicate encounters. Brainwave analysis applied it is too invasive in this paper but would be useful to have a benchmark for discovering other approaches which are more acceptable. The implementation has done in python base open-source platform. Below table 1 shows the classification accuracy with two different algorithms respectively.

Table 1. Accuracy of system using KNN, SVM and CNN

Algorithm	Accuracy	Dataset
KNN	56.25%	Synthetic
	56.64%	EEG-IoT
SVM	65.25%	Synthetic
	64.45%	EEG-IoT
CNN	92.50%	Synthetic
	93.60%	EEG-IoT
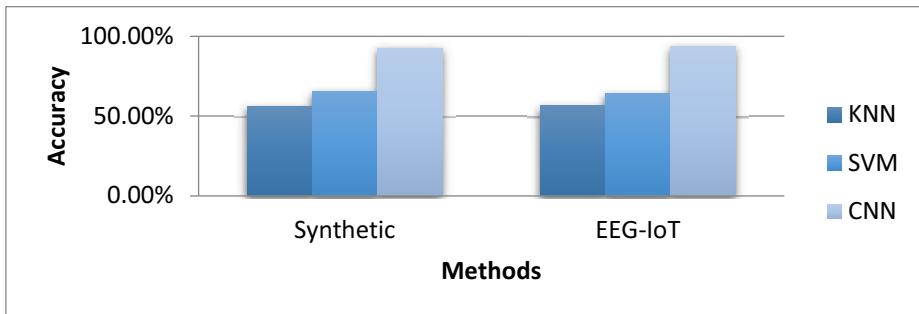
Figure 2. Accuracy of proposed system using KNN and SVM algorithm

The proposed system used KNN, SVM and CNN three different machine learning and deep learning classifiers which are evaluated with different cross-validation of real-time EEG datasets. Various feature extraction techniques have been carried out during the model training and a similar strategy has been applied for testing. The above Figure 2 demonstrates synthetic and real time IoT dataset.

6. Conclusion

Our research looks at the use of deep learning techniques in BCI Emotions classification systems. Images can be used in a variety of ways, including programmed marking with sentimental parts in order, automatically classifying video scenes into thrillers, comedy, romance, and other genres, and automatically classifying video classes with emotions categories. According to the results of the survey, CNN-LSTM can be created with different confusion matrix parameters when dealing with different datasets. With more than 96 percent average classification accuracy, DCNN using the RESNET library achieves the highest accuracy.

Images from one specific event have been effective with allusions to particular news organizations such as CNN has already been demonstrated in a lot of experiments, particularly the context of metaphor construction. Since the CNN-LSTM architectures were used to classify the EEG data, we take these into consideration as well. In order to use the unprocessed input data as training and development data, we first expanded it into a two-dimensional representation. The researchers developed an effective and cutting-edge method of using only visual, auditory, and textual features to accurately and easily identify human emotions. The best features of the Full Feature Equivalent from a long feature table are chosen and used to expand the network to achieve high-quality results faster expansion. Here, we use wavelet-based transforms to translate one-dimensional signals into two-transformed time-frequency representations. Features from the four models are combined to reduce the function vector's scale by including negative information in the proposal calls for the reduction of the number of channels via differential entropy. A model will develop that is several steps ahead of previous models in its ability to obtain functionality. The classifier's efficiency has improved greatly over the preceding versions. Expansion also makes it possible for feature detection and real-time recognition of emotions by using an EEG. According to this work, it appears that neural networks and deep learning can be combined in order to develop a complex model. Further study is still remaining to be done to investigate the emotional analysis, to be performed by the combined use of many neural networks.

References

[1]. Ghosh-Dastidar, S.; Adeli, H.; Dadmehr, N. Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection. IEEE Trans. Biomed. Eng. 2007, 54, 1545–1551.

[2]. Polat, K.; Güne¸s, S. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. Appl. Math. Comput. 2007, 187, 1017–1026.

[3]. Liang, S.F.; Wang, H.C.; Chang, W.L. Combination of EEG complexity and spectral analysis for epilepsy diagnosis and seizure detection. EURASIP J. Adv. Signal Process.2010, 1, 853434.

[4]. Faust, O.; Acharya, U.R.; Adeli, H.; Adeli, A. Wavelet-based EEG processing for computer aided seizure detection and epilepsy diagnosis. Seizure 2015, 26, 56–64.

[5]. Acharya, U.R.; Sree, S.V.; Chattopadhyay, S.; Yu, W.; Ang, P.C.A. Application of recurrence quantification analysis for the automated identification of epileptic EEG signals. Int. J. Neural Syst. 2011, 21, 199–211.

[6]. Bajaj, V.; Pachori, R.B. Classification of seizure and nonseizure EEG signals using empirical mode decomposition. IEEE Trans. Inf. Technol. Biomed. 2012, 16, 1135–1142.

[7]. Andrzejak, R.G.; Lehnertz, K.; Rieke, C.; Mormann, F.; David, P.; Elger, C.E. Indications of Nonlinear Deterministic and Finite Dimensional Structures in Time Series of Brain Electrical Activity: Dependence on Recording Region and Brain State. Phys. Rev. E 2001, 64, 061907.

[8]. Adeli, H.; Ghosh-Dastidar, S.; Dadmehr, N. A wavelet-chaos methodology for analysis of EEGs and EEG sub-bands to detect seizure and epilepsy. IEEE Trans. Biomed. Eng. 2007, 54, 205–211.

[9]. Kabir, E.S.; Zhang, Y. Epileptic seizure detection from EEG signals using logistic model trees. Brain Inform. 2016, 3, 93–100.

[10]. Subasi, A.; Gursoy, M.I. EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Syst. Appl. 2010, 37, 8659–8666.

[11]. Richhariya, B.; Tanveer, M. EEG signal classification using Universum support vector machine. Expert Syst. Appl. 2018, 106, 169–182.

[12]. Raghu, S.; Sriraam, N. Classification of focal and non-focal EEG signals using neighbourhood component analysis and machine learning algorithms. Expert Syst. Appl. 2018, 15, 18–32.

[13]. Acharya, U.R.; Oh, S.L.; Hagiwara, Y.; Tan, J.H.; Adeli, H. Deep convolutional neural network for the automated detection and diagnosis of a seizure using EEG signals. Comput. Biol. Med. 2018, 100, 270–278.

[14]. Truong, N.D.; Nguyen, A.D.; Kuhlmann, L.; Bonyadi, M.R.; Ippolito, J.Y.S.; Kavehei, O. Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. Neural Netw. 2018, 105, 104–111.

[15]. Shoeb, A.H. Application of Machine Learning to Epileptic Seizure onset Detection and Treatment. Ph.D. Thesis, Massachusetts Institute Technology, Massachusetts, UK, 2009.

[16]. Kharbouch, A.; Shoeb, A.; Guttag, J.; Cash, S.S. An algorithm for seizure onset detection using intracranial EEG.Epilepsy Behav. 2011, 22, 529–535.

[17]. Chandel, G.; Upadhyaya, P.; Farooq, O.; Khan, Y.U. Detection of Seizure Event and Its Onset/Offset Using Orthonormal TriadicWavelet Based Features. IRBM 2019, 40, 103–112.

[18]. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. Nature 2015, 521, 436–444.

[19]. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. ACM Commun. 2017, 60, 84–90.

[20]. Ortiz-Garcia, A.; Munilla, J.; Gorriz, J.M.; Ramirez J. Ensembles of deep learning architectures for the early diagnosis of Alzheimer's disease. Int. J. Neural Syst. 2016, 26, 1650025.

[21]. ManasiAgrawal, RutujaAinapure, ShrushtiAgrawal, SimranBhosale, Sharmishta Desai "Models for Hand Gesture Recognition using Deep Learning", 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA),589-594.

[22]. Swati Avad, Dr. Sharmishta Desai "A research review on Indian Sign Language Recognition Using Machine Learning", Journal of Computer Science Engineering and Software Testing (e-ISSN: 2581-6969) Volume 6 Issue 2 Year 2020.

[23]. Meet Patel, Sharmishta Desai "Music suggestion according to emotion for paralyzed and bed rest patient using EEG Signals", Journal of SeyBold report, July 2020,1533-9211.

[24]. Dr.Sharmishta Desai "Dynamic Induction Model for Student's Behavior Analysis",International Journal of System and Software Engineering,7(1), June 2019, 13-19.

[25]. Sharmishta Desai, Dr.S.T.Patil "Boosting Decision Trees for Prediction of Market Trends,Journal of Engineering and Applied Sciences",552-556,April 2018,ISSN:1818-7803.

Aspect Based Emotion Detection and Topic Modeling on Social Media Reviews

Mr. Ganesh N. Jorvekar [a,1] and Dr Mohit Gangwar [b]

aPh.D. Research Scholar, Department of Computer Science and Engineering, Bhabha University, Bhopal, Madhya Pradesh, India
bDean-Engineering, Bhabha University, Bhopal, Madhya Pradesh, India

Abstract. In recent years, the number of user comments and text materials has increased dramatically. Analysis of the emotions has drawn interest from researchers. Earlier research in the field of artificial-intelligence concentrate on identification of emotion and exploring the explanation the emotions can't recognized or misrecognized. The association between the emotions leads to the understanding of emotion loss. In this Work we are trying to fill the gap between emotional recognition and emotional co-relation mining through social media reviews of natural language text. The association between emotions, represented as the emotional uncertainty and evolution, is mainly triggered by cognitive bias in the human emotion. Numerous types of features and Recurrent neural-network (RNN) as deep learning model provided to mine the emotion co-relation from emotion detection using text. The rule on conflict of emotions is derived on a symmetric basis. TF-IDF, NLP Features and Co-relation features has used for feature extraction as well as section and Recurrent Neural Network (RNN) and Hybrid deep learning algorithm for classification has used to demonstrates the entire research experiments. Finally evaluate the system performance with various existing system and show the effectiveness of proposed system.

Keywords— Aspect mining, text processing, feature extraction, feature selection, classification, machine learning, NLP

1. Introduction

On social media peoples share a lot of data in the form of text, photos, and videos in memes, whether personal, everyday scenes, or their opinions. The Internet is a massive forum for global and instantaneous communication and knowledge exchange, providing users with a decent selection of people's viewpoints and thoughts on a huge range of topics. There are rarely any textual captions in most social media posts, but they are filled with images. This largely leads to a spectrum of views and emotions being articulated quite implicitly only through visual content.

Wang Xinzhi et. Implemented Expression Connection Processing on Natural Source Language Via Deep Learning Models, which undermines the connection of emotions depending on the success of speech signals from state-of-the-art deep learning techniques.

---

[1] Mr. Ganesh N, Ph.D. Research Scholar, Department of Computer Science and Engineering, Bhabha University
Email. ganesh.jorvekar83@gmail.com.

Some subjective individual by the dataset are expressed in the system and models are minimized by designing three kinds of features and two artificial learning models. Via an emotion misunderstanding law, which is arbitrary, and an expression evolution legislation, which is guided, the expression connection is mined. Studies are conducted on open to interpretation and long-short Online news messages. There are some promising emotional association studies

Via email, text or images, one may convey feelings. Although several works have used techniques to decode feelings from user posts in the past, especially on social media [2, 3], text sentiment analysis has yet to be explored. In the current situation, social media use to convey feelings is growing, so it is an important field of study. The new technologies are aimed at improving accuracy. For text sentiment analysis, there are several algorithms and techniques proposed.

We aim to find out if there are benefits to applying CNN to the study of visual emotions.

How to fine-tune training data identified on a limited scale distinct from the dataset to identify text, shall be explored. It may only be possible to perform learning in top layers during back propagation, as they have less specific data set features. The experimental findings show that this domain-specific fine-tuning for heterogeneous text classification effectively improves neural network performance.

2. Literature Survey

Xinzhi Wang et.al. (2020) [1], In the past few years, the proposal of emotion analysis has garnered a lot of attention among researchers. Many AI research papers deal with how to identify the feelings, but very few explore why it is that anyone might mistake his or her own for anything else. The fact that the association among emotions contributes to the problems associated with the inability to recognise them. in this study, we are working to bridge the gap between emotional sentiment identification and traditional content analyses by providing more data from articles harvested from the Web. Humans often express their emotions as the tendency to change in emotional expressions, thereby creating the confused and evolving nature of those expressions. Emotion is detected by three kinds of features and two deep neural network models, two of which are deployed in tandem to expand on their meaning. It's estimated that people experience two thousand types of emotions in a movie. From three viewpoints, the hypothesis of emotion evolution is supported. 1) change only happens in one direction at a time, 2) in small steps, and 3) on the shortest path The technique is rigorously examined in three investigations. 1) title examples; 2) and news article body examples of statements spanning both positive and negative terms; and negative content ranging from short ones to long-form texts (long and short). The experimental evidence indicates that people quickly misinterpret emotional statements when they are unaccompanied by other cues. When people are emotionally touched by comments, such feelings appear to create feelings of love or rage. It is obvious in print whether fear or joy is being portrayed; in print, it is simple to identify the emotion of fear or joy being recorded. According to the study's results, some applications, such as affective interactions, social media and public opinion analysis, as well as well as human-computer interaction, this could be very helpful.

Renata L. Rosa et.al. (2019) [2], The author argued that social networks (e.g., OSNs) ought to include information about various user-related topics that was pertinent to the individuals' lives. This application gathers and analyses this type of data while allowing it to track and suggest other applications, such as application monitoring and application recommendations. In this paper, we describe a Knowledge-Based System (KBS), which involves a mental wellbeing monitoring to find those who may be suffering from psychological issues, in particular. Depending on the outcome of the monitoring, one's findings, we could send messages that assist users suffering from anxiety, for better understanding of concepts, the system will rely on KBRS, KB knowledge, and sentiment analysis. To ensure long-term success, there is a feature to alert trusted users in the event of any deviation from the set parameters is built in, particularly with regards to potential itself with building models that identify the cause and effect relationships between different items of data. However, the observation that online social networks tend to expand a user's ability to make connections could mean they could result in a decreased amount of actual face-to-to-face interactions. A massive outbreak is the result of these phenomena, and these people now have the words such as Phubbing (This word, according to my opinion, is about laziness) and Nomophobia (This word, for me, denotes fear of being out of cell phones) to identify them.

3. Proposed System Design

The proposed research paper on text sentiment classification uses a strong machine learning technique called deep learning Based on the examples given, the application illustrates numerous text extraction and data building options and principles are added to the train's understanding. Feature extraction techniques including aspects like form, texture, alpha density, and colour have been implemented in order to derive other attributes such as amplitude and tone. Most times, it is used to define text meta-data to include categorization of thoughts as well. In order to attain higher classification accuracy, normalizing the data has the data has the biggest influence.

Features have been taken from the training data collection and applied during training, and a training model has been developed accordingly. The model that allows each textual function to be introduced with the same amount of strength was added to the testing results. the weight estimation approach takes into account the attributes of research and preparation, as well as one other related attributes It's a part of the sub-process of describing the similarities between two sets of characteristics. by using weighting factors that build on the factor in their appetite, the factors according to their respective emotional responses A perfect weight is completely subjective; it is entirely up to the customer to the individual to decide. You will find more detail here.

1. Text resizing. Evaluate each text's height and width accordingly and adjust it to the appropriate size.
2. Often text contains some noise or a certain kind of noise already contains unique input text. We remove noise from images using the Gaussian noise filter.

4. Results and Discussions

The To measure the accuracy for the method, the matrix has to be expanded a multi-tier java 3. 2.2 GHz and 512 MB RAM on the implementation using the OPENstep Java Runtime Environment; we have made it the calls to the web server using 1.3MB and 2MB of Java SE class directories. After the evaluation of various systems and alternatives has been done, the preference of a more sophisticated system could have been reached.

This experiment we analyse the classification accuracy of ReLU using twitter dataset, the similar experiments has done with various cross validation and results has illustrates in table 1. According to this analysis we conclude 10-fold cross validation provides highest 95.30% and 96.10% for 10-fold cross validation classification accuracy for RNN.

Table 1. Average classification accuracy with confusion matrix RNN

RNN	Fold 5	Fold 10	Fold 15
Accuracy	94.20	95.30	96.10
Precision	94.30	95.70	95.30
Recall	94.15	95.80	96.40
F1 Score	93.20	95.60	96.50
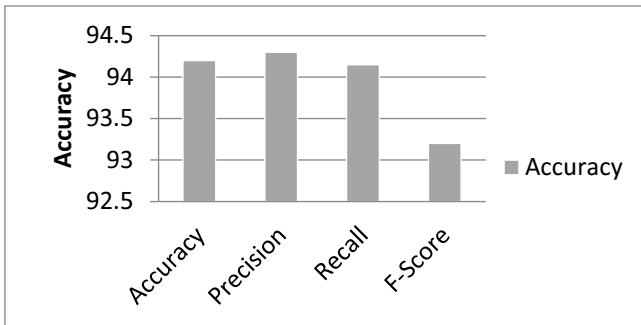
Figure 1. System classification accuracy using RNN with 5-fold data cross validation
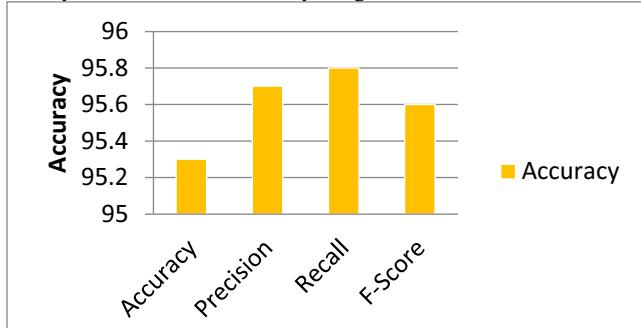
Figure 2. System classification accuracy using RNN with 10-fold data cross validation
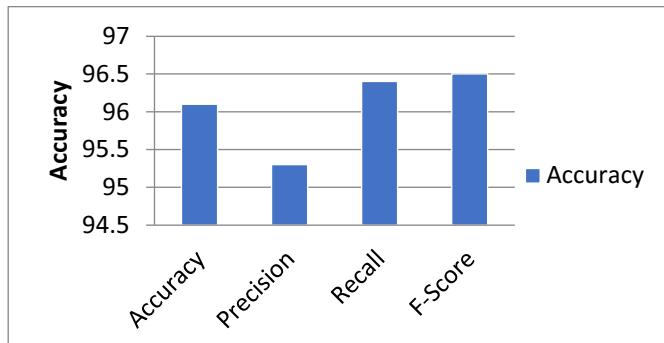
Figure 3. System classification accuracy using RNN with 15-fold data cross validation

5. Conclusions

Our research examines the techniques of deep learning in text sentiment classification schemes. in programmatic classification, videos provide several representations of all three of the three of these emotions in a row in a specific order, thereby categorising them as thrillers, romantics, and comedy, with about a third being labelled as action. According to the survey, it was found that there are three different types of RNN (over-regularization, under-regularization, and over-regularization) as the method works on different datasets. can be expressed as a total state expansion of the following DNN utilising more than 96% of the available Deep4 resources, making it the most efficient. Our observational study indicates that, along with the classification of emotions, the Deep Learning algorithm provides promising text data results. The future study involves conducting tests in a manufacturing environment on a wide scale. In addition to pictures, video scenes can be accommodated to further categories feelings into various genres such as happy, thriller, humorous, romance etc.

References

[1].   Xinzhi Wang , Luyao Kou , Vijayan Sugumaran , Xiangfeng Luo, and Hui Zhang, " Emotion Correlation Mining Through Deep Learning Models on Natural Language Text", IEEE TRANSACTIONS ON CYBERNETICS, May 2020.

[2].   Renata L. Rosa, Gisele M. Schwartz, Wilson V. Ruggiero, and Demostenes Z. Rodr ıguez, Senior Member, IEEE, "A Knowledge-Based Recommendation System that includes Sentiment Analysis and Deep Learning", IEEE Transactions on Industrial Informatics Vol. 15, April 2019.

[3].   B. Liu,"Many facets of sentiment analysis, a practical guide to sentiment analysis," Springer International Publishing, pp. 11–39,Jan 2017.

I.   B. Weiner and R. L. Greene, "Handbook of personality assessment," in John Wiley and Sons, N.J, EUA, 2008.

[4].   H. Lin, J. Jia, J. Qiu, Y. Zhang, G. Shen, L. Xie, J. Tang, L. Feng, and T. S. Chua, "Detecting stress based on social interactions in social networks," IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 9, pp. 1820–1833, Sept 2017.

[5]. J. T. Hancock, K. Gee, K. Ciaccio, and J. M.-H. Lin, "I'm sad you're sad. Emotional contagion in cmc," in Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, pp. 295– 298, Year 2008.

[6]. Afzaal M, Usman M and Fong A 2019 "Tourism Mobile App with Aspect-Based Sentiment Classification Framework for Tourist Reviews", IEEE Transactions on Consumer Electronics (IEEE). 65. 233-242. doi.10.1109/TCE.2019.2908944.

[7]. Rana T A, Cheah Y N 2017 "Improving aspect extraction using aspect frequency and semantic similarity-based approach for aspect-based sentiment analysis", In. International Conference on Computing and Information Technology, Springer, Cham, doi.10.1007/978-3-319-60663-7_30, pp. 317-326.

[8]. I.-R. Glavan, A. Mirica, and B. Firtescu, "The use of social media for communication.", Official Statistics at European Level. Romanian Statistical Review, vol. 4, pp. 37–48. Dec. 2016,.

[9]. M. Al-Qurishi, M. S. Hossain, M. Alrubaian, S. M. M. Rahman, and A. Alamri, "Leveraging analysis of user behavior to identify malicious activities in large-scale social networks," IEEE Transactions on Industrial Informatics, vol. 14, no. 2, pp. 799–813, Feb 2018.

[10]. R. L. Rosa, D. Z. Rodr´ıguez, and G. Bressan, "Music recommendation system based on user's sentiments extracted from social networks," IEEE Transactions on Consumer Electronics, vol. 61, no. 3, pp. 359–367, Oct 2015.

[11]. R. Rosa, D. Rodr, G. Schwartz, I. de Campos Ribeiro, G. Bressan et al., "Monitoring system for potential users with depression using sentiment analysis," in 2016 IEEE International Conference on Consumer Electronics (ICCE). Sao Paulo, Brazil. IEEE, pp. 381–382, Jan 2016.

[12]. Y. P. Huang, T. Goh, and C. L. Liew, "Hunting suicide notes in web 2.0 – preliminary findings," in Ninth IEEE International Symposium on Multimedia Workshops (ISMW 2007), pp. 517–521. Dec 2007.

[13]. W. H. Organization, "World health statistics 2016. Monitoring health for the sustainable development goals, world health statistics annual," World Health Organization, p. 161, 2016.

[14]. Hu M, Liu B 2004, "Mining and summarizing customer reviews.", In. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, doi.10.1145/1014052.1014073, pp. 168-177.

An Intrusion Detection System for Network Security Using Recurrent Neural Network

Kishor P. Jadhav [a,1] and Dr Mohit Gangwar [b]

[a]*Ph.D. Research Scholar, Department of Computer Science and Engineering, Bhabha University, Bhopal, Madhya Pradesh, India*
[b]*Dean-Engineering, Bhabha University, Bhopal, Madhya Pradesh, India*

Abstract. To maintain the security of vulnerable network is the most essential thing in network system; for network protection or to eliminate unauthorized access of internal as well as external connections, various architectures have been suggested. Various existing approaches has developed different approaches to detect suspicious attacks on victimized machines; nevertheless, an external user develops malicious behaviour and gains unauthorized access to victim machines via such a behaviour framework, referred to as malicious activity or Intruder. A variety of supervised machine algorithms and soft computing algorithms have been developed to distinguish events in real-time as well as synthetic network log data. On the benchmark data set, the NLSKDD most commonly used data set to identify the Intruder. In this paper, we suggest using machine learning algorithms to identify intruders. A signature detection and anomaly detection are two related techniques that have been suggested. In the experimental study, the Recurrent Neural Network (RNN) algorithm is demonstrated with different data sets, and the system's output is demonstrated in a real-time network context.

Keywords. Recurrent Neural Network, KDDCUP99, Intrusion Detection System, Network security.

1. Introduction

The IDS is responsible for detecting a connection form of attack, such as a fragment of unknown attack, a DoS attack, a U2R attack, or an R2L attack. It then deploys a series of such components one by one in a sequential fashion. This accomplishes two objectives. For starters, each sub-phase can only train a limited number of characteristics that detect a specific form of attack. Second, the sub-size unit is still small enough to be helpful. A common disadvantage, similar to our system, is that it increases the amount of time it takes for modules to communicate. However, in our system, this can be easily prevented by making each sub-phase independent of the other layers. As a result, specific characteristics can be observed in more than one sub-phase. If an offense is committed without a centralized decision-maker, any thread will block it, depending on the channel's security policy. Numerous sub phases mainly function as filters blocking suspicious associations as long as they are formed during a specific layer, allowing for a quick response to the intrusion while also reducing analysis time in successive phases. It should be noted that in different sub-phases that rely on sight-trained attacks, completely different responses are often initiated.

---

[1] Kishor P. Jadhav,Ph.D. Research Scholar, Department of Computer Science and Engineering, Bhabha University, Bhopal, Email. kishor268@gmail.com.

At the first layer and in subsequent stages, the amount of system-analysed auditing information decreases further as more and more attacks are detected and blocked.

In the worst case, if no attacks are detected prior to the last sub-phase, all staggered sub-phases in phase 2 have the same load. However, as attacks are detected and blocked in any subsequent method, the average load is expected to be significantly lower. On the other hand, when the sub-phases are arranged in parallel rather than in a series, in a sequence configuration on a subsystem, the load is equal to the worst case. The initial step can be repeated in the sequential configuration to perform load balancing to improve performance.

2. Literature Survey

This article uses the ANN (Artificial Neural) of an Operating System Sensor to monitor malicious activities in Android and ios devices, based on the Flow anomaly system [1], based on the flow anomaly Detection Platform for Android mobile devices. The detection rate of this approach is 85 percent and 81% accuracy, respectively. Impersonation is considered in terms of CPU, space and better view, which helps to characterize a small, scalable and effective IDS after an Integration node to combat public attacks by various services. By using powerful data mining algorithms, the data sources are analysed. Improving the accuracy and classification rate requires the future scope.

PRADEEP and Dr. Yogesh Kumar [2] Effectual Secured Approach for the Internet of Things with Fog Computing and Mobile Cloud Architecture Using Ifogsim, this work cloud computing performance is assessed Simulation model world using iFogSim, where artifacts and Cloud services provide a greater degree of consistency and Precise.

Javier A. et al. proposed in [3] information security boosting using malware detection in a network environment. The platform designed would be an efficient algorithm for malware detectors for Ghana limited application security due to extensive Framework. In the final research, the participants are already confident and pleased with their reliability and functionality. The research revealed this device met that experiment's goal. "High Quality" analyzed the processes and solution to the proposed method. The development of the malware detection system for Asia Technology Security to maintain its position was successful.

Bholanath Mukhopadhyay et al. [4], cloud-based task scheduling and Protection using SSL for IaaS Application, implemented a new approach wherein we built both protection and authorization access policies. We also implemented the functionality of an Endpoint Protection choice search. In our configuration, numerous profiles can be built, one with its own different access policy, for various network applications. For illustration, for dynamic access point connections, and internet connectivity authentication policy can be developed. Using our unorthodox technique, it is possible to quickly classify the user, customer location, existing network situation at the time of connection, and server status.

Algorithm Design

Training Process

Input. Training dataset TrainData[], Various activation functions[], Threshold Th

Output. Extracted Features Feature_set[] for completed trained module.

Step 1. Set input block of data d[], activation function, epoch size,

Step 2 . Features.pkl ← ExtractFeatures(d[])

Step 3 . Feature_set[] ← optimized(Features.pkl)

Step 4 . Return Feature_set[]

Algorithm for system testing

Input. Training dataset TestDBLits [], Train dataset TrainDBLits[] and Threshold Th.

Output. Resulset <class_name, Similarity_Weight> all set which weight is greater than Th.

Step 1. For each testing records as given below equation; it works in convolutional layer for both training as well as testing

$$testFeature(k) = \sum_{m=1}^{n} (. \ featureSet[A[i] \dots \dots A[n] \ \leftarrow TestDBLits)$$

Step 2. Create feature vector from testFeature(m) using below function.

$$Extracted\_FeatureSet\_x \ [t \dots \dots n] = \sum_{x=1}^{n}(t) \leftarrow testFeature \ (k)$$

Extracted_FeatureSet_x[t] is the outcome of each pooling layer that is extracted from each convolutional layer and forward to net convolutional layer? This layer holds the extracted feature of each instance for testing dataset.

Step 3. For each train instances as using below function,

$$trainFeature(l) = \sum_{m=1}^{n} (. \ featureSet[A[i] \dots \dots A[n] \ \leftarrow TrainDBList)$$

Step 4. Generate new feature vector from trainFeature(m) using below function

$$.Extracted\_FeatureSet\_Y[t \dots \dots n] = \sum_{x=1}^{n}(t) \leftarrow TrainFeature \ (l)$$

Extracted_FeatureSet_Y[t] is the outcome of each pooling layer that is extracted from each convolutional layer and forward to net convolutional layer? This layer holds the extracted feature of each instance for training dataset.

Step 5. Now evaluate each test records with entire training dataset, in dense layer

$$\text{weight} = \text{calcSim (FeatureSetx} \mathbin{||} \sum_{i=1}^{n} \text{FeatureSety[y])}$$

Step 6. Return Weight

3. Proposed System

Machine learning methods were used to identify and avoid intrusions in the current research methods. The runtime packets data block will conduct training, including packet selection for remote data monitoring. The role collection for a particular packet operation will then be submitted. Send it forward as a group if all is well. Misconduct samples will be examined for feature selection for different attributes in order to identify individual attacks. Figure 1 illustrates the system's entire execution using specified algorithms. To produce train modules and conduct research, various machine learning techniques were employed. The proposed network intrusion detection mechanism aims to increase the detection accuracy, cut the number of false positives, and minimise the amount of time to wait for a detected intrusion. This means that there are two distinct phases to the proposed system. during system preparation, NSLKDD data will be used; afterwards, it will be used to conduct system tests.

The proposed system would be arrayed with comprehensive support. Individuals with these same features would generate two or more models known as an ensemble model. This ensemble model has taken in the different classifier classifications from many sources and produced a single result. We've developed our classifier definition of numbers based on this information. The first step is that the programme obtains data from different sources, whether online or offline. Once all the data has been entered, classification algorithms are in place, other techniques for data mining will be employed.

System initially collects the input packet from various sources like KDD CUP, NSL KDD, ISCX and real time network packets. The entire execution holds three different phases which are listed as below.
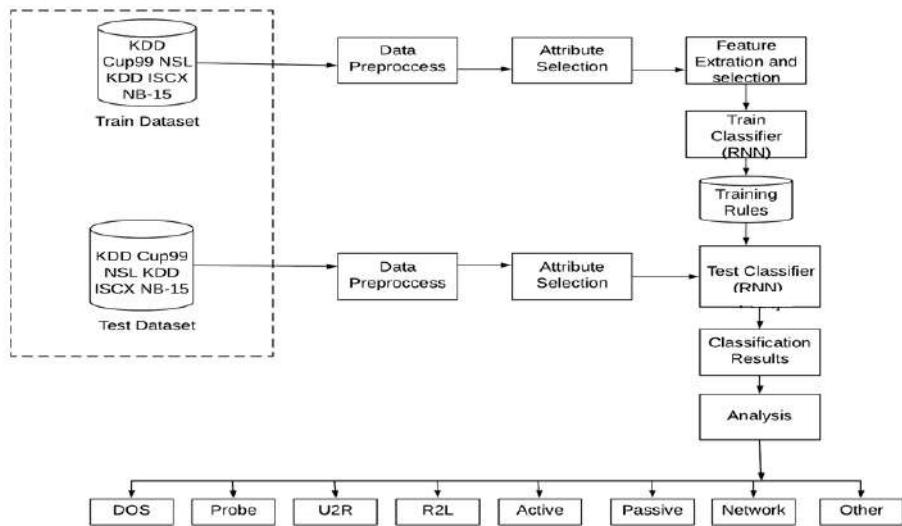
Figure 1. Proposed system architecture

4. Results and Discussion

We measure the performance evaluation of the system after it has been successfully implemented. The outcome of system collected on real time as well as synthetic traffic data and validates it with machine learning algorithm. The outcome of system has been shown in Figure 2 with multiple attack detection both environments. Figure 3 shows depict various methods, such as the RNN algorithm, were used to identify and predict the classification accuracy of the proposed system.
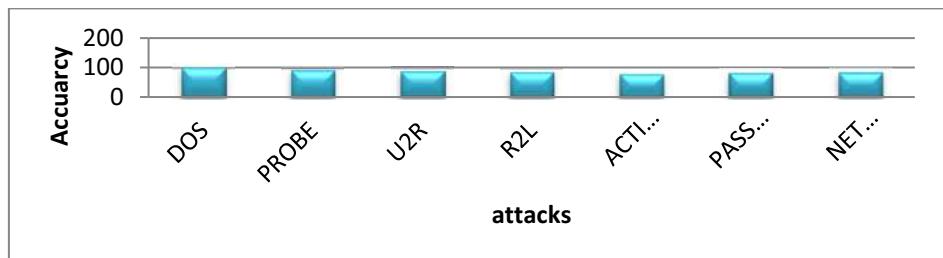
Figure 2. Detection accuracy with various attacks using RNN classification

According to the results of the second experiment, RNN with sigmoid has a higher classification accuracy than the other two activation functions, ReLU and TanH (see Figure 3). Based on the results of the above experiment, we may infer that the proposed framework improves the accuracy of trust computation in the IoT in-service environment. The entire study is driven by a collection of simulation environmental conditions and a mix of machine learning techniques. With regards to machine learning algorithms, a variety of computation specifications have been used clusters distinction and id.mi.com.
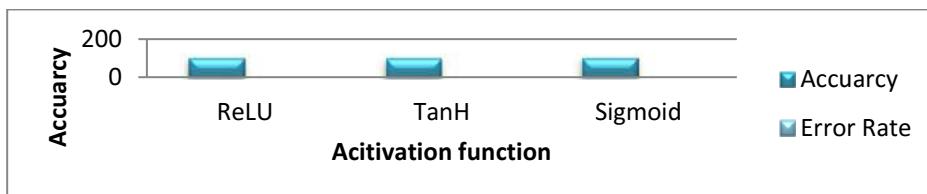
Figure 3. Experimental analysis of RNN with three activation function

5. Conclusion

In this research we proposed an efficient IDS scheme, this research proposes an RNN-IDS approach focused on deep learning. We used the numerous real time networks as well as some synthetic dataset to evaluate anomaly detection and classification accuracy. We also use deep learning to apply IDS in the cloud environment in the future. In addition, we examine and compare different deep learning approaches, such as. During the data search, the software basically functions as an RNN classification and soft computing algorithms to evaluate the unknown type of connection and attacks. To improved classification and high-class identification are possible important to the powerful rule structure. Several studies have been used for experimental investigation for evaluate the algorithm's effectiveness using a variety of methods, and we came to the conclusion that we were getting satisfactory results.

References

[1]. Panagiotis I. Radogloa-Grammatikis; Panagiotis G. Sarigannidis, "Flow anamoly based Intrusion Detection System for Android Mobile Devices", 2017 6th International Conference on MOCAST, May 4-6, 2017, Kazani, Greece.

[2]. PRADEEP, S.; SHARMA, Dr Yogesh Kumar. Effectual Secured approach for Internet of Things with Fog Computing and Mobile Cloud Architecture Using IFogSim. WE C-2019-London, UK, DOI, 2019, 978-988.

[3]. Jaevier A. Villanueva, Luisito L. Lacatan, Albert A. Vinluan, Information Technology Security Infrastructure Malware Detector System, International Journal of Advanced Trends in Computer Science and Engineering, pp. 1583-1587 ,Volume 9, No.2, 2020.

[4]. Bholanath Mukhopadhyay , Dr. Rajesh Bose, Dr. Sandip Roy, A Novel Approach to Load Balancing and Cloud Computing Security using SSL in IaaS Environment, International Journal of Advanced Trends in Computer Science and Engineering, pp. 2130-2137,Volume 9, No.2, 2020.

[5]. Arthur, Menaka Pushpa. "Detecting Signal Spoofing and Jamming Attacks in UAV Networks using a Lightweight IDS." 2019 International Conference on Computer, Information, and Telecommunication Systems (CITS). IEEE, 2019.

[6]. Otomo, Safa, Burak Kantarci, and Hussein T. Mouftah. "On the feasibility of deep learning in sensor network intrusion detection." IEEE Networking Letters 1.2 (2019). 68-71.

[7]. Vinayakumar, R., et al. "Deep learning approach for the intelligent intrusion detection system." IEEE Access 7 (2019). 41525-41550.

[8]. Sheu, Ruey-Kai, et al. "IDS-DLA. Sheet Metal Part Identification System for Process Automation Using Deep Learning Algorithms." IEEE Access 8 (2020). 127329-127342.

[9]. Abou Khamis, Rana, M. Omair Shafiq, and Ashraf Matrawy. "Investigating Resistance of Deep Learning-based IDS against Adversaries using min-max Optimization." ICC 2020-2020 IEEE International Conference on Communications (ICC). IEEE, 2020.

[10]. Panagiotis I. Radogloa-Grammatikis; Panagiotis G. Sarigannidis, "Flow anamoly based Intrusion Detection System for Android Mobile Devices", 2017 6th International Conference on MOCAST, May 4-6, 2017, Kazani, Greece.

[11]. Trae Hurley, Jorge E. Perdomo, Alexander Perez-pons, "HMMBased Intrusion Detection System for software-defined networking", 2016 15th IEEE Conference on Machine Learning and Application, Dec 18-20, 2016, Miami, Florida.

[12]. Loganathan, Gobinath, JagathSamarabandu, and Xianbin Wang. "Sequence to sequence pattern learning algorithm for real-time anomaly detection in network traffic." 2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE). IEEE, 2018.

[13]. Kumar, A., Vengatesan, K., Vincent, R., Rajesh, M., & Singhal, A. (2019). A novel Arp approach for cloud resource management. International Journal of Recent Technology and Engineering (IJRTE), 7(6).

[14]. Mayank Agarwal, SankethPurwar, Santosh Biswas, Sukumar Nandi, "Internal Detection System for PS-Poll DOS attack in 802.11 networks using real-time discrete event system", IEEE,vol.4,issue4,2017.

Enhanced IoT Based Child Missing Alert System Using RSSI

Simran Kaur Gill [a,1], and M. Safa [a]

[a] *Dept of Information Technology, School of Computing, SRM Institute of Science and Technology, Chennai, TamilNadu, India*

Abstract. Our proposed system discusses the concept of a smart wearable device connected to their parent's mobile phone for children and their parents respectively. In this project we propose that to let the system be divided into three parts, namely the safe, intermediate and danger zones. If the child is within the safe zone, then no buzzer is sounded whereas if the child is in the intermediate range a buzzer alert will be sounded. If the child crosses the 'danger' zone, the buzzer is sounded with an immediate notification sent to the parent. In case the child goes out of danger zone, a GPS module is attached that would help parent know the exact location of the child once he/she is outside the 100meters of radius from the parent. This project also has features to sense the child's temperature and heartbeat along with notifying the child's parent in case the child has an accident using the temperature, heartbeat and pressure sensors respectively. The RSSI is used for distance sensing whereas GSM is used for notification sending to the parent's mobile phone.

Keywords. Free Sensing, Tracking, Radio Signal Strength Information (RSSI), Global System for mobile Communications (GSM), Global Positioning System (GPS).

1. Introduction

We often come across billboards filled with their faces smiling at us. In most cases, these children are found safely by the government officials, while the others are lost forever. In such cases we imagine such reunions filled of tearful happiness with hugs and kisses. But fail to notice that while these children and their families, where they have found each other had a happy ending. The odd senses we feel of what they might have went through bring a chill to my spine. On occasions where we do find them, some of them are found missing a part of their body. Thus, a proper child location tracking is necessary. Therefore, we would like to propose an enhanced RSSI based Child Tracking System using IoT of independent and efficient tracking of the child while the child is nearby to the parent and aided with a GPS module when the child is far away from the reach of parents, preferably outdoors in a crowded place.

---

[1] Simran Kaur Gil,Dept of Information Technology, School of Computing, SRM Institute of Science and Technology, Chennai, Email: ss8723@srmist.edu.in.

2. Proposed System

In our enhanced RSSI system we propose that to let the system be split into three parts, the safe, intermediate and danger zones. If the child is within the safe zone, then no buzzer is sounded whereas if the child is in the intermediate range a buzzer alert will be sounded. If the child crosses the 'danger' zone, the buzzer is sounded with an immediate notification sent to the parent. In case the child goes out of danger zone, a GPS module is attached that would help parent know the exact location of the child once he/she is outside the 100meters of radius from the parent. The RSSI is used for distance sensing whereas GSM is used for sending notification to the parent's mobile phone. The system uses the Aurdino Uno microcontroller and takes the input from the two sources. One of which is the environmental parameters Heartbeat, Temperature and Pressure Sensors. The second takes the received signal of the child from the parent using the RSSI module. And then the data is sensed in the form of dbms to check whether the child is in range of the parent mobile hotspot or not, that is 100 meters, to send the precise reading in case the child is going far away from the central point. Meanwhile using the DHT11, transducer and heart rate sensors to sense the body temperature of child along with his/her heart rate and in case if the child has fallen down, it sends the appropriate notification via sms to the parent device. Meanwhile a buzzer goes 'off' to notify the people around the child of the same situation. For a checking the precise location of the child once he/she is out of the 100metres radius, a GPS module is also installed. The location of the child through which would be sent as a link to the parent via sms. For the purpose of sending the messages to the parent, GSM   or   the   Global   System   for   Mobile   Communications   is   used.
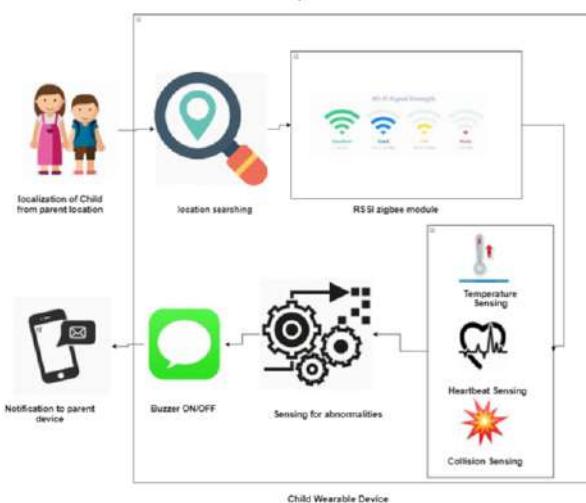
Figure 1. work flow diagram

3. Project Dataflow and Architecture
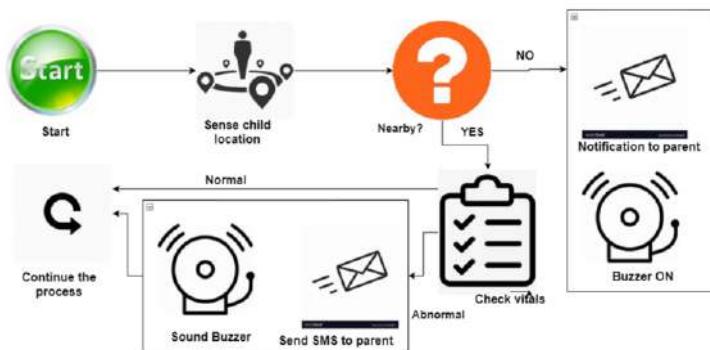
Figure 2 System Data Flow Diagram

This system consists of two separate pieces. A child wearable device and the parent mobile phone. The child wearable device consists of all the sensors, microcontroller, GPS Module, GSM, a Buzzer, an LCD, LEDs and the RSSI module. On the parent mobile phone, the wifi hotspot is used as the centre within which the RSSI senses the child location, stepping out of which a buzzer is sounded. The hardware first senses the current real time location of the child. If the child is right beside the parent, a green light is turned on whereas if the child is a little further away from the parent the yellow light is turned on else if the child goes out of the 100 meters radius, the red LED is turned on with a notification sent about the same to the parent and Buzzer is turned ON. Next, if the child has high temperature, heartbeat or if he/she fell down, an appropriate message is sent to the parent device and simultaneously a buzzer is turned ON. The LCD setup on the child wearable device also displays the same message on it. The whole process plays continuously on a loop to sense the abnormalities and report them accordingly.

4. Implementation
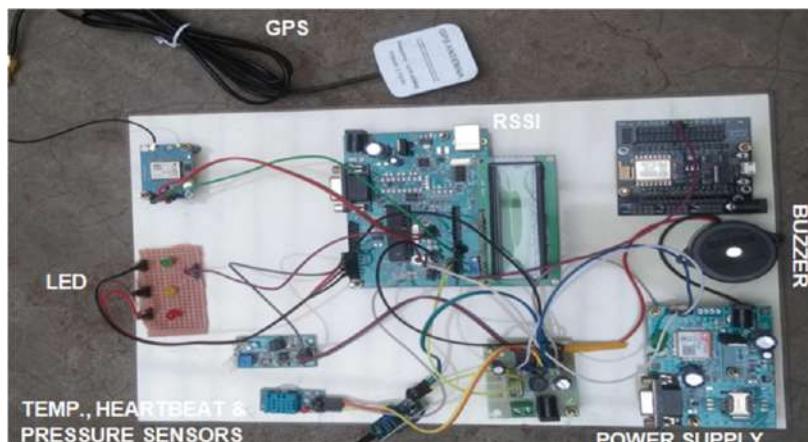
Figure 3. System Architecture

The microcontroller contains one or more Central Processing Units (CPU). It is the unit that undertakes the administration of the various input and the output constraints. Arduino code is launched on the microcontroller. It processes the input and sends it to another module and also to control the output devices i.e., the buzzer and LCD on the child wearable device along with the LEDs and Parent notification. It accepts the temperature, heartbeat and the intensity of pressure faced by the child wearable device as the input constraints from the DHT11 and transducer sensors. In the next step, the source code is launched on the microcontroller. From here, the microcontroller begins to take the input constraints and forwards it to the next module. DHT11 sensor is used to measure the temperature and heartbeat at the moment, which is used as a constraint to decide the type and nature of the notification and whether or not to send it.
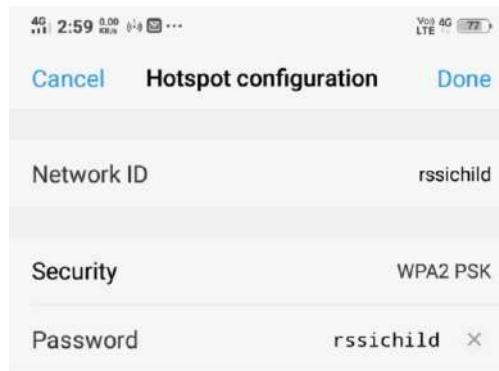
Figure 4. Wifi Configuration

The child device and the parent device are connected to each other via the parent mobile hotspot. The COM 19 available on the Aurdino Uno is used is used to sense the wifi hotspot which in turn connects and initiates localization via RSSI which then uses the parent's mobile phone as the central point to which the child's distance is calculated.A GPS module is setup on the child wearable. This is done because the Received Signal Strength Indicator only indicates the relative distance between the child and the parent. In the GPS Output, the message is made up of three major components. This message is known as the navigation message. Thus it is important to extract the specific longitude and latitude values.

Sending message alert to parent mobile phone via GSM

The proposed framework contains a set of notifications to be sent to the parent mobile, such as the child 'fell down' or 'irregular heartbeat and temperature', or even in case the child is 'out of range' when the child crosses the 100metres radius distance from the parent location. The code snippet is necessary to transfer the notification to the registered parent contact details. The notification is sent by inserting the parent's contact number in the 'phno' area whereas the '+91' indicates the country code. This overall notification or alert sending work is done via the GSM module.

5. Input/output Result
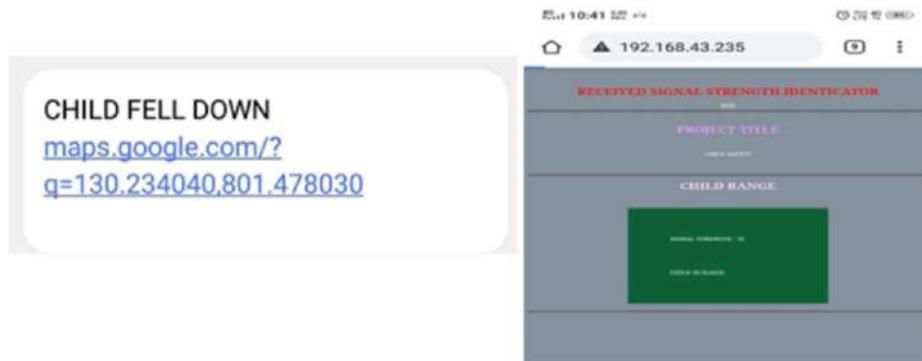
Figure 5. a). Notification with location of child and b). RSSI based Localization output for parent

6. Conclusion

Hence, the most useful part of this system over the other alternatives is that it can be used in any smart phone irrespective of its cost. Moreover, this system doesn't require the parents to be of any technical background. Hence can easily be used by every parent, unbeknownst of their study background. The aim of this project is to aid the parents pinpoint their child's location and detect when he/she is fallen astray, in the most unchallenging way. This device can help the government officials by pointing out all the possible hotspots from which the children are abducted with the timings around which these cases frequently take place. With the hotspots and timings noted, more attention and security during this time range when these cases occur, this project can help lessen the cases of child missing and kidnappings. On the other hand, The Device can fail to prove to accuracy during really high temperatures, as during such weathers the body temperature is significantly higher and lower during winters.

References

[1] Daniel Konings, Fakhrul Alam, Frazer Noble, and Edmund M-K Lai (2019). "Device-free Localization Systems Utilizing Wireless RSSI: A Comparative Practical Investigation".

[2] Kishore Kumar Reddy. N.G, Ramakrishnan.G and Rajeshwari.K (2017). "Ensuring Fishermen Safety through a Range Based System by Trizonal Localization using Low Power RSSI"

[3] Pradipta Ghosh, Jason A. Tran, and Bhaskar Krishnamachari (2019). "ARREST: A RSSI Based Approach for Mobile Sensing and Tracking of a Moving Object"

[4] Y. Guo, K. Huang, N. Jiang, X. Guo, Y. Li, and G. Wang, "An exponential-rayleigh model for rss-based device-free localization and tracking," IEEE Transactions on Mobile Computing, vol. 14, no. 3, pp. 484–494, 2015.

[5] D. Konings, N. Faulkner, F. Alam, F. Noble, and E. M. K. Lai, "The effects of interference on the rssi values of a zigbee based indoor localization system," in 2017 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), 2017, Conference Proceedings, pp. 1–5.

[6] L. Oliveira, H. Li, L. Almeida, and T. E. Abrudan (2014). "Rssi-based relative localisation for mobile robots,".

[7] D. Konings, N. Faulkner, F. Alam, F. Noble, and E. M-K Lai, "Do rssi values reliably map to rss in a localization system?" in Workshop on Recent Trends in Telecommunications Research (RTTR). IEEE, 2017

[8]   I. 18305:2016, Information technology – "Real time locating systems – Test and evaluation of localization and tracking systems", ser. Standard ISO/IEC 18305:2016. ISO/IEC JT, 2016. [Online]. Available: https://www.iso.org/standard/62090.html

[9]   D. Vasisht, S. Kumar, and D. Katabi (2016). "Decimeter-level localization with a single wifi access point," .

[10] Simo Särkkä, Ville V. Viikari, Miika Huusko, and Kaarle Jaakkola (2012). "Phase-Based UHF RFID Tracking With Nonlinear Kalman Filtering and Smoothing"

[11] Emidio DiGiampaolo and Francesco Martinelli (2014). "Mobile Robot Localization Using the Phase of Passive UHF RFID Signals"

[12] Mauro Boccadoro, Francesco Martinelli and Stefano Pagnottelli (2010). "Constrained and quantized Kalman filtering for an RFID robot localization problem"

[13] Ricardo Tesoriero, José A. Gallud, María D. Lozano, Víctor M. R. Penichet (2009). "Tracking Autonomous Entities using RFID Technology"

[14] SIMON J. JULIER AND JEFFREY K. UHLMANN (2004). "Unscented Filtering and Nonlinear Estimation"

[15] Piyush Raj, M. Safa, Nitish Kumar, P J Sahrudh, Shivaditya Singh. (2020). Enhanced Smart Music Controller by Applying CNN in IoT. International Journal of Advanced Science and Technology, 29(06), 2739 - 2749.

[16] Aditya Vadagave, M.Safa, Sanjit Singh Sokhi, Bhaawan Mathur, Bhagavatula Anirudh. (2020). Enhancing Smart Precision Agriculture in IoT Using Agro ecologicali Analysis. International Journal of Advanced Science and Technology, 29(06), 2750 - 2760.

[17] Abhirup Bose, M. Safa, Sanjay Bhargav Siddi , Manas Raj Anand, Vivek Kumar. (2020). Implementation of Dynamic Lighting & Augmented Reality (DLAR) Smart Home System for Deaf and Hard-of-Hearing (DHH) Residents. International Journal of Advanced Science and Technology, 29(06), 2724 - 2738.

[18] Aadhineni Ganesh, M. Safa, M.Roopanjali, P.V.K Sai Sri Harsha, D. Anjana Tulasi. (2020). An Approach for Monitoring and Analyzing Parking Occupancies using IoT. International Journal of Advanced Science and Technology, 29(06), 3536 - 3545.

571

Prediction of Heart Disease Severity Measurment Using Deep Learning Techniques

Rina S. Patil [a,1] and Dr. Mohit Gangwar [b]

ªPh.D. Research Scholar, Department of Computer Science and Engineering, Bhabha University, Bhopal, Madhya Pradesh, India
ᵇDean-Engineering, Bhabha University, Bhopal, Madhya Pradesh, India

Abstract. Machine learning enables AI and is used in data analytics to overcome many challenges. Machine learning was the growing method of predicting outcomes based on existing data. The computer learns characteristics from the test implementation, then applies characteristics to an unknown dataset to predict the result. Classification is an essential technique of machine learning which is widely used for forecasting. Some classification techniques predict with adequate accuracy, while others show a small precision. This research investigates a process called machine learning classification, which combines different classifiers to enhance the precision of weak architectures. Experimentation using this tool was conducted using a database on heart disease. The collecting and measuring data method were designed to decide how to use the ensemble methodology to improve predictive accuracy in cardiovascular disease. This paper aims not only to enhance the precision of poor different classifiers but also to apply the algorithm with a neural network to demonstrate its usefulness in predicting disease in its earliest stages. The study results show that various classification algorithmic strategies, such as support vector machines, successfully improve the forecasting ability of poor classifiers and show satisfactory success in recognizing heart attack risk. Using ML classification, a cumulative improvement in the accuracy was obtained for poor classification models. That process efficiency was further improved with the introduction of feature extraction and selection, and the findings show substantial improvements in predictive power.

Keywords. Heart disease, Machine learning, Internet of Things, Feature extraction, Feature selection

1. Introducion

Insufficient asset status of tribal communities and unawareness are the critical problems in the rural healthcare system. For example, 650 million people live in a rural area Bharat to tell the story 33 per day with this financial benefit. It will be difficult for a private individual to access innovative medical facilities to have associate degree access. There is, therefore, a great need to build an outsized health monitoring system that is low-cost, economical, and easy to use. Here since we would like to provide medical assistance to even the poorest segment of society, the emphasis on low prices may not be tasteful. Exhaustive research has been devoted to researching various

---

[1] Rina S. Patil, Ph.D. Research Scholar, Department of Computer Science and Engineering, Bhabha University Email. rinapatil190@gmail.com.

developments, such as advances in knowledge in supplementation and strengthening established medical facilities. The Internet of Things (IoT) has generally been linked to interface open therapeutic assets. It offers older adults and clinicians any persistent, secure, viable, and keen medical services management. Indeed, before making the IoT vision a fact, many test protection problems must be tended to. We need to address essential questions about further empowering the IoT while maintaining all aspects, such as faith, protection, and security. We have guided this work to assist people who are keen on growth and improvement in this arena. Various overviews have just been suggested, but they rely for the most part upon a broader vision that involves "Things"-located, "Web"-arranged, and "Semantic"-arranged thoughts, or on a layered sight. Our vision remains a good decision for fundamental leadership, regardless of the limits of its hypothetical meticulousness, because we are speaking about the system-level mission.

2. Literature Survey

According to the authors [1], cardiac disease is one of the world's leading causes of death, and precipitate disease prediction is essential. The computer-aided machines assist the doctor as a gimmick in detecting and treating heart disease. This paper aims to broaden cardiovascular disease related to the heart and briefly explore open decision support systems through data mining and smart hybrid techniques for computing and researching heart disease. With many methodologies, a lot of DSS remains to be projected on heart disease.

Explain that according to estimates, heart disease is one of the most fundamental reasons behind deaths worldwide [2]. In challenging scientific support systems, data mining techniques are prevalent and can identify concealed patterns and associations in medical data. Data mining classification methods have so far been used to test the different forms of heart-based issues. This essay aims to build a data mining techniques framework using classification methods of analysis mining. The various clustering methods, nearest neighbor, EM, and the most restricted hypothesis methodology for first heart disease are discussed in this article.

Using different machine learning algorithms in [3] [4], the multi-classification problem has been solved. Both systems deal with massive unstructured data and continue to use the methodology of supervised learning. The results obtained showed that classification accuracy was higher than conventional machine learning algorithms.

According to Chengjin Yu. et al. [5], an adversarial training approach is proposed to multitask learning to estimate multi-type Cardiac Indices in MRI and CT. By using multitask learning networks, these task dependencies are shared and learned. Lastly, they transferred parameters learned from MRI to CT. A series of experiments were performed. The authors first optimized system efficiency over 2900 cardiac MRI images through ten-fold cross-validation. Then the network was run on an independent data set with 2360 cardiac CT images. The results of all experiments on the proposed reverse mapping indicate excellent performance in estimating multiple cardiac indices.

3. Propsoed System Design

The framework first gathers the input sequence states of each device, then uses ADC to transform it from analog - to – digital. when the transformation has been completed, it

will be received by the microcontroller and saved in the database same moment. The dynamic monitoring method parallels work for entire application events and displays them to the front wave user interface. Then, the suggested machine learning algorithms operate in the system's central ware, often checks all model parameters from the target threshold, if any time account level below minimum confidence and max resistance, then the output devices are automatically executed. Simultaneously, as the system calculates the hazardous amount of the operation state, the program also means the time count with the same state, and the timer and GPS messaging system will run whenever it reaches the expected to meet.
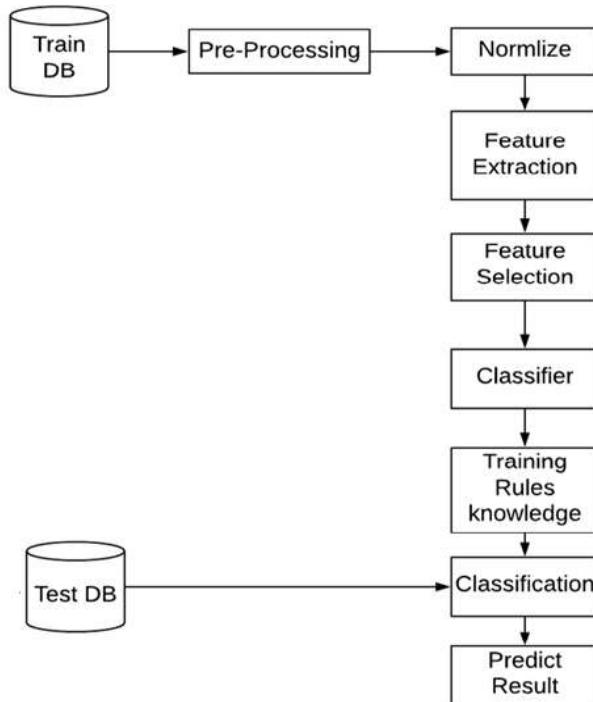
Figure 1. Proposed system architecture

4. Results and Discussions

With different Java and Mobile and web platforms, the proposed architecture has been transmitted. The Figure 2 below shows that the recognized part causes data loss and ECG sensor (it may be useful for loss and increased rate sensors). In contrast, stress dramatically increases the number of proposed devices that can be used all the time for the given packet size. Here x-axis shows the no. of users, and Y shows the average packet loss from various IoT devices below, Figure 2 shows the system accuracy with no experimental test analyses.
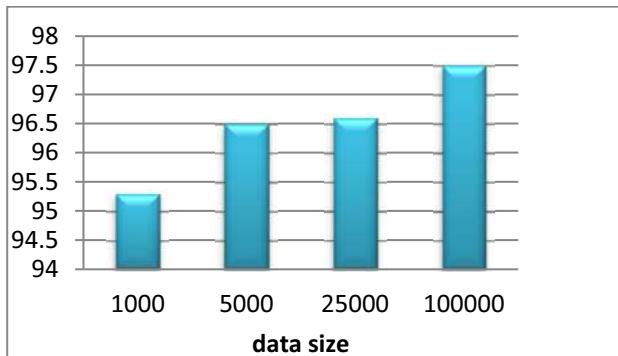
Figure 2. System accuracy of proposed system with false ratio.

Figure 3 provides the thorough overview of the next experimental investigation to measure the effectiveness of the proposed method.
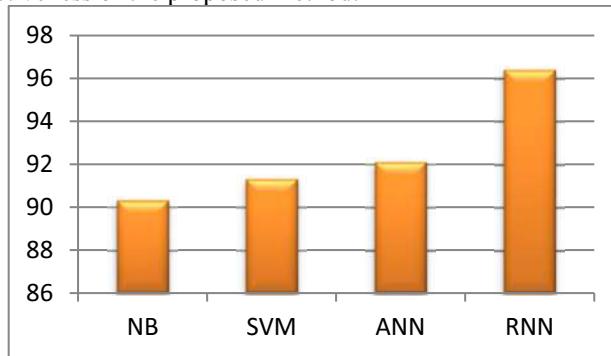
Figure 3. System accuracy of proposed system with various existing systems.

The overall experimental analysis describes a comparative analysis of proposed system vs various existing systems. The system supports Real-time health tracking and disease predictions over the network. Detection probability is more substantial than different learning styles, and it can function based on synthetic and real-time datasets. The device also has the option to convey alarm when there is some criticality over the internet.

5 Conclusions

In this research, we introduced an ideal framework for prediction models in real-time, and users with the coronary disease can use that. Unlike many other systems, it is capable of tracking and predicting both. The system's diagnostic method will predict cardiovascular disease utilizing ML algorithms, and the predictive conclusions are based on the dataset example of heart disease. On the other hand, the device is very economical. we used an enthused pulse sensor and sent the data via the Arduino suite microcontroller to mobile devices. To check the variances and raise the alarm if the patient's heart rate increases above the usual heart rate. In order to prove the system's efficacy, we performed tests with both the tracking and diagnostic method. We conducted experiments with supervised machine classification methods such as NB, SVM and ANN. The procedure was done with the validation set test, and 96% efficacy

of the proposed method was obtained with the Random Forest. We performed two tests for the monitoring system. We conducted with various healthy patients on the first sample, and we interacted with 20 individuals with a cardinal illness on the second. In both cases, the control system's accuracy was 100%. We intend to use the proposed device in the future and to omit by use of a sensor.

References

[1].   Vijayashree, J., and N. Ch Sriman Narayana Iyengar. "Heart disease prediction system using data mining and hybrid intelligent techniques. A review." Int. J. Bio-Sci. Biotechnol 8 (2016). 139-148

[2].   Singla, Meenu, and Kawaljeet Singh. "Heart Disease Prediction System using Data Mining Clustering Techniques."

[3].   B.Pandu Ranga Raju, B.Vijaya Lakshmi, C.V.Lakshmi Narayana, Detection of Multi-Class Website URLs Using Machine Learning Algorithms, International Journal of Advanced Trends in Computer Science and Engineering, pp. 1704-1712 ,Volume 9, No.2, 2020.

[4].   Roobaea Alroobaea, An Empirical combination of Machine Learning models to Enhance author profiling performance, International Journal of Advanced Trends in Computer Science and Engineering, pp. 2130-2137, Volume 9, No.2, 2020.

[5].   Sun, Yingnan, Frank P-W. Lo, and Benny Lo. "A deep learning approach on gender and age recognition using a single inertial sensor." 2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN). IEEE, 2019.

[6].   Lo, Frank P-W., et al. "A novel vision-based approach for the dietary assessment using deep learning view synthesis." 2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN). IEEE, 2019.

[7].   Uddin, Md Zia, and Mohammad Mehedi Hassan. "Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network." IEEE Sensors Journal 19.19 (2018). 8413-8419.

[8].   M.Ganesan and Dr.N.Sivakumar, "IoT based heart disease prediction and diagnosis model for healthcare using machine learning models ", Proceedings of International Conference on Systems computation automation and networking 2019, IEEE- 978-17281-1524.

[9].   AKM Jahangir Alam Majumder, Yosuf Amr ElSaadany, Roger Young Jr., and Donald R. Ucci, "An Energy Efficient Wearable Smart IoT System to Predict Cardiac Arrest", Hindawi Advances in Human-Computer Interaction Volume 2019, Article ID 1507465, 21 pages.

[10]. C. Yu et al., "Multitask Learning for Estimating Multitype Cardiac Indices in MRI and CT Based on Adversarial Reverse Mapping," in IEEE Transactions on Neural Networks and Learning Systems, DOI. 10.1109/TNNLS.2020.2984955

[11]. Liu, X., Faes L., Kale A et al. A comparison of deep learning performance against healthcare professionals in detecting diseases from medical imaging.9* a systematic review and meta-analysis, Vol 1(6), 271-297, 2019.

[12]. Byrne, N., Velasco Forte, M., Tandon, A., Valverde, I., & Hussain, T.. A systematic review of image segmentation methodology, used in the additive manufacture of patient-specific 3D printed models of the cardiovascular system, JRSM Cardiovascular Disease, 2016, https.//doi.org/10.1177/2048004016645467

[13]. Damen Johanna A A G, Hooft Lotty, Schuit Ewoud, Debray Thomas P A, Collins Gary S, Tzoulaki Ioanna et al. Prediction models for cardiovascular disease risk in the general population. systematic review, 2016, BMJ 2016; 353. i2416.

[14]. Matthew Marsh A, Literature Review of Image Segmentation Techniques and Matting for the Purpose of Implementing "Grab-Cut", Department of Computer Science at Rhodes University, South Africa, Available http.//citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.103.1334&rep=rep1&type=pdf

[15]. Mamta Sharma, Farheen Khan, Vishnupriya Ravichandran, "Comparing Data Mining Techniques Used for Heart Disease Prediction, f Engineering and Technology, vol.5, Issue 6, 2017

Analysis of Brain Tumor Disease Detection Using Convolutional Neural Network

Yogesh Deshmukh [a,1], Samiksha Dahe [a], Tanmayeeta Belote [a], Aishwarya Gawali [a],
Sunnykumar Choudhary [a]

[a] *Dept. of I.T. Sanjivani College of Engineering, Maharashtra, India*

Abstract. Brain Tumor detection using Convolutional Neural Network (CNN) is used to discover and classify the types of Tumor. Over a amount of years, many researchers are researched and planned ways throughout this area. We've proposed a technique that's capable of detecting and classifying different types of tumor. For detecting and classifying tumor we have used MRI because MRI images gives the complete structure of the human brain, without any operation it scans the human brain and this helps in processing of image for the detection of the Tumor. The prediction of tumor by human from the MRI images leads to misclassification. This motivates us to construct the algorithm for detection of the brain tumor. Machine learning helps and plays a vital role in detecting tumor. In this paper, we tend to use one among the machine learning algorithm i.e. Convolutional neural network (CNN), as CNNs are powerful in image processing and with the help of CNN and MRI images we designed a framework for detection of the brain tumor and classifying its Different types.

Keywords. Brain Tumor, MRI, Machine Learning, Convolutional Neural Network (CNN), Classification.

1. Introduction

As there is huge improvement in the medical field, medical images of various parts of human body plays a vital role in diagnostic and treatment and also in medical research. The medical imaging consists of various imaging technologies like X-ray, Fluoroscope and Magnetic resonance imaging (MRI). In all those images a tumor disease is frequent and complex. And hence Brain tumor disease detection is one of the important topics in the medical field. The detection of brain tumor is basically depending on medical imaging data analysis. The Accurate and Precise analysis of tumor is a important step in determination of patient's state. In these various factors are important as far as the correct diagnosis of tumor is concern i.e., doctor's knowledge, experience and visual weariness. And this various factor affects the accurate analysis of tumor image analysis. Hence detection of brain tumor images is very important. Magnetic Resonance Imaging (MRI)[5] provides the useful information of size, shape and position of tumor. The images obtained using MRI is accurate and exact. The efficiency of diagnosis is greatly improved with the help of MRI [5]. The MRI act as a guide for lesion localization and

---

[1] Yogesh Deshmukh, Assistant Professor, Dept. of I.T. Sanjivani College of Engineering
Email:yogeshdeshmukhit@sanjivani.org.in.

surgical treatment and it avoids the operations of thoracotomy. MRI of Brain uses three-dimension multi-band imaging technology [5]. In multi model MRI images we can obtain the different structures of the same tumor using the unused development sequence. Different modes can display different brain tumor features. Basically, brain tumor has classified into two different types that is Cancerous and non-cancerous. Non-cancerous brain tumor is called as benign these types of tumors stays in only one place they do not spread in the body or in brain which is not that much dangerous to our body. cancerous brain tumor is called as malignant these types of tumor does not stays in only one place they spread in our body or in brain very rapidly which dangerous to our body. That's why early prediction of brain tumor is very important so that the patient can take the appropriate treatment according to its brain tumor type (i.e., cancerous and non-cancerous) [4].

Magnetic Resonance Imaging (MRI) can provide information about the shape of tumor, its size, and position of the tissues and organs with no high ionizing radiation [3]. MRI of the brain is useful in identifying problems such as blurry vision or seizures, dizziness persistent headaches, weakness, and it helps in detecting certain chronic diseases of the nervous system that is multiple sclerosis. In some of the cases, we cannot able to see the clear images of parts of brain with the help of an X-ray, CAT scan, or ultrasound. that's is the reason we use MRI images here because MRI scan can provide us the precise and clear images of the different parts of the brain and makes it valuable for diagnosing. In addition, MRI imaging also can acquire totally different structures of identical tissue mistreatment the unused development sequence. That is, multimodal MRI image [3]. Different types of modes will show different tumor options.

2. Related Work

In the basic research, the image pre-processing is done on basis of height, width& number of channels like RGB i.e., RED, GREEN and BLUE that means the colors, It also work with the spatial attributes. Ming Li, Lishan Kuang , Shuhua Xu & Zhangvo Sha [1] has provide a mechanism to identify & classify the tumor, in that they use MR image that provides information like shape, size & position of human tissues. By using the MR image with provided information the prediction & classification carried out. Multi-model MR image-based brain tumor detection has used by many researchers. Brain tumor detection technique is used for dividing the tumor tissues into different types such as tumor tissues, edema tissues, necrotic tissues and normal tissues. Problem of brain tumor detection has a challenge of tumor to itself, because of that it received widespread attention for tumor detection in the last 20 years. By human intervention, the image detection of brain tumor is classified into 3 categories: Manual detection, Semi-Manual section and fully-automatic Detection. The manual detection is completely manual based it is capable of manually depicting the contour of tumor. The semi-manual detection is based on the manual initialization. And the fully automatic detection is done without human interference.

3. Problem Description and System Architecture

Now a day's various images in medical field plays important role in diagnosis of various disease and research as well. Therefore, the research on the medical diagnosis

data is now become very important as a brain tumor with periodic occurrences and with their complexity, now a day's brain tumor detection has become an important topic in the research of medical field. The detection of brain tumor is depending on the various medical images available on tumor. The patient's condition is determined by doing the analysis of brain tumor images accurately. So, the data available on brain tumor in the form of medical images are processed with the help of various Machine learning algorithms like Convolutional Neural Network (CNN). The MRI image provides the tumor information about size, shape and position of tumor. Those images will be provided as a input to the various machine learning algorithms. Machine learning plays vital role in prediction of brain tumor. In the Preprocessing phase the MRI images will be filter and smooth images will be given as input to the machine learning algorithm.
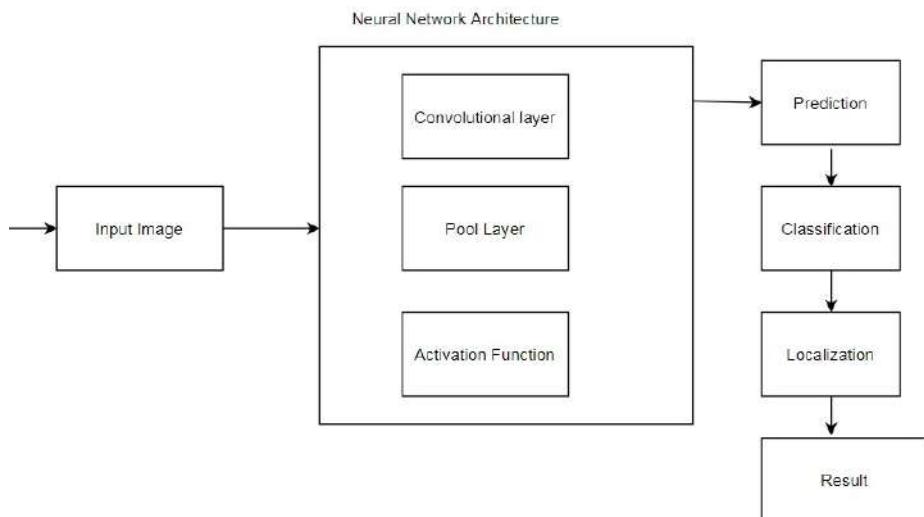
Figure 1. Architecture of the proposed system

Multi-model MR image-based brain tumor detection has used by many researchers. Brain tumor detection technique is used for dividing tumor tissues into different types such as tumor tissues, edema tissues, necrotic tissues and normal tissues. Brain tumor detection has a challenge in tumor detection to itself because of that it received widespread attention for tumor detection in last 20 years. By human intervention, the image detection of brain tumor is classified into 3 categories: Manual detection, semi-manual and fully automatic detection. The manual detection is completely manual based, it is capable of manually depicting the contour of tumor, the semi-manual detection is based on the manual. Brain tumor MR images uses 3D multiband imaging technology and chest x-ray and etc. By comparing the 2D images with 3D images, 3D multiband MRI provide the coordinate position of lesion area that helps doctors to locate the lesion area accurately.

3.1. Input Layer:

In this model it takes input as an 3D MR image. This MR image is the key point on basis of that MRI it will predict the presence of tumor or not. After the completion of

pre-processing of MR images, it will be provided to the Convolutional Neural Network (CNN).

3.2. Convolutional Neural Network:

CNN is a machine learning algorithm used very commonly in the signal processing. The name of the Convolutional Neural Network is also derived for image signals from special convolutional operations. The main purpose of convolution operation in the CNN is to extract features from images. Basically, CNN consisting the multiple convolutional layers, Pooled layers and fully-connected layers. Usually in the front part of the network the convolutional layer and pooled layer appearing alternately, while the latter part consisting the Fully-connected layers.

3.3. Activation Layer:

Activation function is a branch that is concatenated after the 1st and 2nd layer of CNN. The standard activation functions are namely sigmoid, ReLu and softmax.

3.4. Pooling Layer:

Poling is used to reduce the number of computations and it is an aggregate statistical operation of images. The main use of pooling is to remove the duplicate data or information and to reduce the dimension of feature map. After the image passes the feature map is obtained through the convolutional layer. The main purpose of pooling operation is to reduce the dimension of the feature map and to remove some redundant information thereby it is reducing the number of calculations and also avoid the over-fitting. The CNN model predicts the tumor after the image processing. After tumor detection it simply classifies the detected tumor of appropriate type. By observation or prediction, it will classify the tumor among the four types of tumor. Namely these types are Benign, Malignant, Gliomas and Meningioma. It will help medical staff for providing the accurate treatment to patient. After the classification of tumor, the location of tumor is identified by the system which is effectively help medical staff for providing the proper treatment to the patient, or there is a need of tumor surgery then it will provide huge help to doctors.

4. Results and Discussion

The proposed technique was implemented using Python programming language. The Convolutional Neural Network (CNN) algorithm was implemented for the classification of various images. We have taken 1000 MRI images from Kaggle dataset and we have used 75% of input images for training and 30% for testing.

The performance of the proposed technique was evaluated using different measurement parameters. The Accuracy was calculated using following equations.

$$Presicion = \frac{TP}{TP + FP} \qquad \text{(i)}$$

$$Recall = \frac{TP}{TP + FN} \qquad \text{(ii)}$$

$$Specificity = \frac{TN}{TN + FP} \qquad \text{(iii)}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad \text{(iv)}$$

TP-True Positive, TN-True Negative, FP-False Positive and FN-False Negative
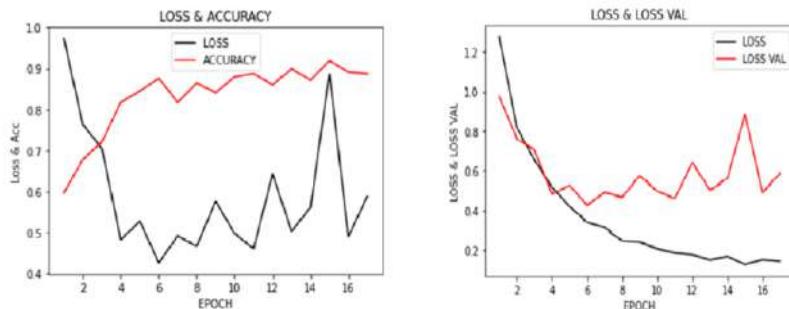
Figure 2. Loss and Accuracy value of the training model for the data set
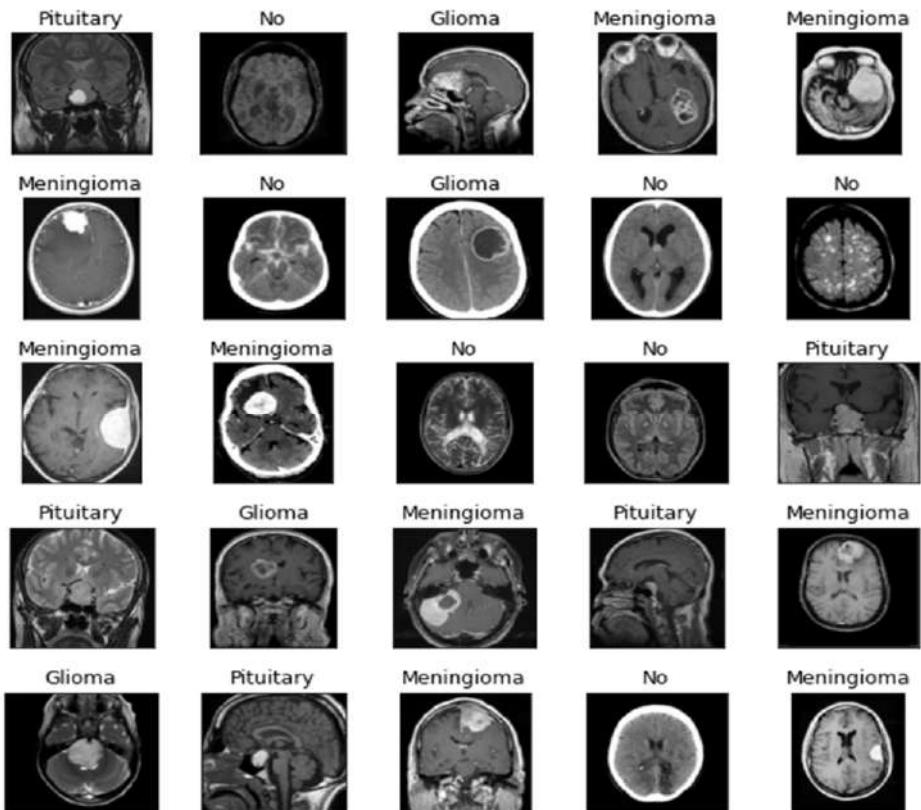
Figure 3. Classification of input data based on the classes i.e. Glioma, Pituitary, Meningioma and Normal.

5. Conclusion

As we all know the brain tumor is one of the dangerous diseases. It can cause the human death. This disease is not possible to detect early using manual processing. So, in this paper we used some Deep Learning models to detect tumor as early as possible. In this paper we used CNN algorithm which is very important for image processing and classification. In CNN there are three main layers i.e., convolutional layer, activation layer and pooling layer. These all layers are interconnected so that CNN can process and perceive data in order to classify images. Based on classification prediction is done. Another module is used i.e., localization using specific object detection. Defected area of brain will be highlighted using localization.

References

[1]. Ming Li, LishanKuang, Shuhuaxu and ZhanguoSha, ''Brain Tumor Detection Based on Multimodal Fusion and Convolutional Neural Network,'' IEEE Access,dec 9, 2019.

[2]. S. Khajanchi and S. Banerjee, ''A strategy of optimal efficacy of T11target structure in the treatment of brain Tumor,'' J. Biol. Syst., vol. 27, no. 1, pp. 1–31, 2019.

[3]. M. A. Khan, I. U. Lali, and A. Rehman, ''Brain Tumor detection and classification: A framework of marker–based watershed algorithm andmultilevel priority features selection,'' Microsc. Res. Technique, vol. 82, no. 6, pp. 909–922, 2019.

[4]. T. Cao, W. Wang, S. Tighe, and S. Wang, ''Crack image detection based on fractional differential and fractal dimension,'' IET Comput. Vis., vol. 13, no. 1, pp. 79–85, Feb. 2019.

[5]. M. Ramalho, A. P. Matos, and M. Alobaidy, ''Magnetic resonance imaging of the cirrhotic liver: Diagnosis of hepatocellular carcinoma and evaluation of response to treatment—Part 1,'' RadiologiaBrasileira, vol. 50,no. 1, pp. 38–47, 2017.

[6]. Q. Li, Z. Gao, and Q. Wang, ''Glioma segmentation with a uni-fied algorithm in multimodal MRI images,'' IEEE Access, vol. 6, pp. 9543- 9553, 2018.

[7]. W. Wu, A. Y. C. Chen, and L. Zhao, ''Brain tumor detection and segmentation in a CRF framework with pixel-pairwise affinity and superpixel-level features,'' Int. J. Comput. Assist. Radiol.Surg., vol. 9, no. 2, pp. 241–253, 2014.

[8]. Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore

[9]. Y.S. Deshmukh, P. Kumar, R. Karan and S.K. Singh, "Breast Cancer Detection based Feature Optimization using Firefly Algorithm and Ensemble Classifier", 2021 International Conference on Artificial Intelligence and smart system(ICAIS) IEEE, pp. 1048-1054.

[10]. Kumar, A., Vengatesan, K., Vincent, R., Rajesh, M., & Singhal, A. (2019). A novel Arp approach for cloud resource management. International Journal of Recent Technology and Engineering (IJRTE), 7(6).

[11]. Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.

Online Attendance Marking System Using Facial Recognition and Intranet Connectivity

Vibhanshu Pant [a,1], Deepak Sharma [a], Annapurani. K [a], R. Sundar [b]

[a] *Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, India*
[b] *Scientist-D, National Institute of Ocean Technology, Chennai, India*

Abstract. Keeping up the attendance record with everyday exercises is a difficult task. The conventional method of marking staff attendance is by tapping their ID card and then using fingerprint scanner. But due to COVID-19 pandemic the attendance system of using fingerprint scanner is stalled and currently not in use. The following system depends on face recognition and intranet connectivity to keep up attendance record of facilities and staff. The paper discusses the attendance marking system that is passive (no direct contact with the scanner or sensor) and restricting the users within certain network. The main goal of this system is divided in two steps, in initial step face is snare from the front camera of the smart phone and it is then recognized in the picture and in the second step these distinguished appearances and features are contrasted with stored information in data set for confirmation.

Keywords. Attendance, appearances, COVID-19, face-recognition, intranet connectivity, front-camera, data-set, database

1. Introduction

Maintaining attendance is a very important task in all the institutes and organizations, whether it is of students or for faculties, staff or employees. Every organization has their own method for marking attendance. Some of them are taking attendance using the old paper or file based method while others are using bio metric techniques. These methods take time for marking attendance and this could lead to large queue every day. Bio metric systems, used for attendance marking system, consists of two major process, first special highlights of an individual are stored in data set and secondly attendance is marked based on ID and confirmation of the individual. When an individual gives his attendance to the machine, it is contrasted with the data that is in the data set and in the event that it is coordinated, the attendance is marked effectively. Numerous strategies have been proposed for facial identification and acknowledgment like Ada Boost calculation, the Float Boost calculation, the S-Ada Boost calculation Support Vector Machines (SVM), and the Bayes classifier.

---

[1] Vibhanshu Pant, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Email:vs3137@srmist.edu.in.

 Face recognition procedures can be either Appearance Based, in which utilization of surface highlights is applied to entire face or some particular districts is executed, or Feature Based, in which utilization of mathematical highlights like mouth, nose, eyes, eye temples, cheeks and connection between them is actualized.

   The attendance system proposed in this paper depends upon two steps, in initial step face is snare from the front camera of the smart phone and it is then recognized in the picture and in the second step these distinguished appearances and features are contrasted with stored information in data set for confirmation. A successful message is displayed if the verification process is successful.

2. Literature Survey

     Everybody needs to keep track of the employees that are working under them. In this area we have seen many changes, starting from using pen and paper to many RFID machines for marking daily attendance. Some organizations uses RFID cards for swiping to mark their attendance, it has its own shortcomings, like the staff have to stand in queues for a long time and there can be chances of proxy attendance also. Many organizations use fingerprint scanners in addition to RFID cards to eliminate these shortcomings.

     In [4] the authors have proposed a system which uses HAAR Classifier and computer vision algorithm to implement face recognition System. The result was that when the system is integrated with the already existing system it uses the existing feature from Learning Management System, but the shortcomings of this paper were that the recognition rate is 56% In paper [5] the authors use PCA and ANN for facial recognition. They first apply the PCA algorithm for training data and reduce its dimensionality and then use ANN to classify these input data which then helps to find the pattern. The result they concluded is that the accuracy is high due to the combination of PCA with ANN as ANN used for classification is more accurate than PCA with Eigen face. The only limitation of this system is that it requires a high computational time and thus the implementation cost increases. In paper [6] the authors use PCA along with MATLAB to implement a face recognition system. The result from this system was that the pre-processing of the image and cropping the region of interest from it made the accuracy of the system higher. The drawback of this system was that this system was only tested with a single image. In paper [3] the authors use Eigenvector and Eigenvalue for designing a face recognition system. Besides using these, the system was also able to prevent the fake attendance mishap by the implementation of clock time. It is used for checking whether the student was there in the class for the whole time or not. The proposed system has many upsides but the accuracy of recognition of faces is high only for the frontal face of the student. In case the face is in the direction of 54 degree the accuracy is only 58%. In paper [7] the authors uses PCA with Histogram due to which additional noises were removed from the data. The proposed system provides better accuracy due to the implementation of the Ada-Boost algorithm on Face Detection. In addition to this skin classification was also utilized for improving the precision of the system. The database is also updated up to date with new images. Since it does not implement the concept of clock time therefore there will be an issue of fake attendance if the student does not attend the whole class and leaves in the middle.

3. Proposed Face Recognition based Attendance Marking System

We are building a mobile application to provide a hassle-free and proxy free attendance marking during the rush hour and also try to provide an alternative to fingerprint scanner for daily attendance of staffs or employees of an organisation.
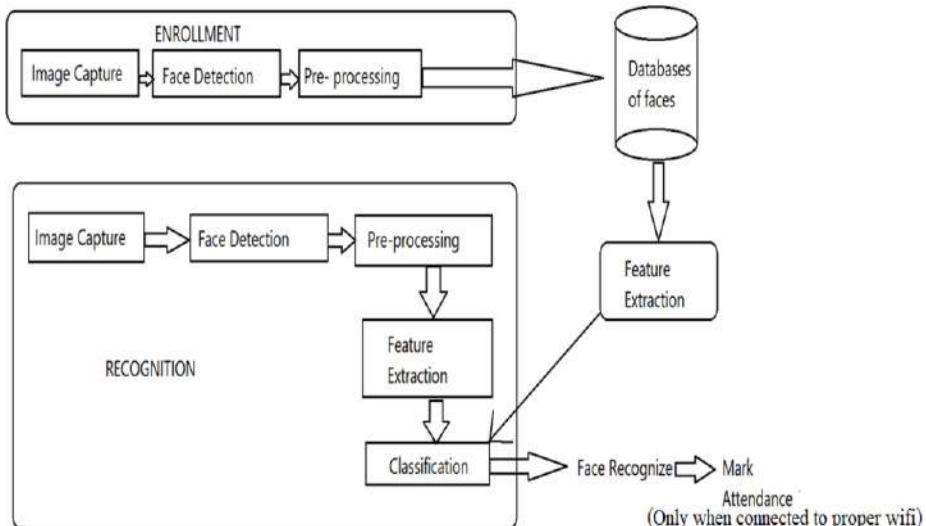
Figure 1. Architectural diagram of the system

Figure 1 shows the architectural diagram of the system demonstrating how the attendance system will work.

3.1. Conceptual Working

The working of the system is of two steps. The first step includes the database creation. In this, the staffs of that particular organization have to get enrolled in the system. Here an admin enrols the staff into the database, when the database is ready then only the training of the model is done over these data sets. The model is then tested with real time images.

Now the staffs have been successfully registered in the system. In the next step, they have to open the mobile application and login with the given credentials. If it is validated, the system goes to the photo capturing stage where staff clicks his photo using the front camera and the face is detected and recognized from it. This is done with the pre-existing data sets in the database. If the recognition stage is successfully passed, a message stating that the attendance has been marked is displayed on the screen.

3.1.1. Dataset

For training our model we have taken fifty people selfies from thirty different

angles so totally 1500 images are taken. Live data set is being used to train our model. A copy of each selfie is then converted to grey scale. Small data set [2] used for analysis of skin tumour.

3.1.2. Face Detection and Recognition

For detecting and recognizing the face in the image captured by the camera Haar Cascade [1] is used. Haar Feature Selection it divides the group of pixels into two groups consisting of a black and white group. It has three variants which are edge features, line features, and four rectangular features. In this, these rectangle features are superimposed on the image array that is obtained and then calculate the data value and then move to the next one. By doing this 'delta' value is calculated for each feature which is the difference be-tween the white region from black region. In the HAAR features, the dark pixels or region describes lower values whereas the bright pixels denote higher values. In Adaboost training misclassification value is calculated by adding all the correctly classified parts and the incorrectly classified parts, these are shown by the plus sign and negative sign respectively. In Cascade classifier, all the three stages are combined and then the resultant is trained. There are four feature and each feature have different sizes and positions therefore in total HAAR Cascade Classifier has around 180000+ features

3.1.3. Intranet Connectivity

To restrict the application to a certain network, WifiManager API is used which a library is provided by Android Studio. Through this, IP address of the currently connected WiFi is obtained when IP Address function from connection Info class of System Service package is invoked. It is checked against the IP Address of the network we want the application to work in. If it doesn't match, an error message is shown otherwise login activity is continued.

3.1.4 Database

Firebase Firestore is used as a database. We save the attendance in name and timestamp format. The attendance can be later download as json file. We selected this as our database since it is scalable and follows ACID properties and it is also integrated with google cloud for better storage.

4. Results and Discussion

To build the android application Android Studio is used and for building face detection and recognition model Pycharm is used as Integrated Development Environment (IDE). JAVA, Python and Kotlin are used as programming languages.

4.1. Face Detection

Face detection is the first step for face recognition and it can be done by four different methods as Knowledge based, Feature based, Template matching and Appearance based. In our work we have gone with Feature based technique.

The Feature Based technique utilizes a pre-processor of the picture based technique and aides the hunt of picture based strategies utilizing GUI that inspect the face competitor locales rather than performing colossal inquiry in all aspects of the test picture. In this we are using Haar cascade classifier for face detection and it is Feature based, and with edge detection, centre detection, line detection, centre detection and facial features like eyes nose mouth. Figure 2 shows the outcome when python scripts are run for detecting faces.
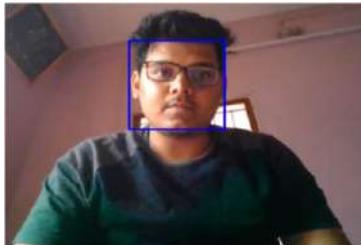
Figure 2. Results of Face Detection   **Figure 3**. Results of Face Recognition

4.2. Face Recognition

In order for system to start recognizing faces, cv2 and numpy libraries are imported. cv2 library contains a Local Binary Pattern Histogram (LBPH) Face Recognizer class. In this class, create method is used to create a recognizer for the model, the recognizer will read the trainer.yml file and based on the training file the output is predicted using the predict function of the recogniser. From the current model we got an accuracy of 43%. Figure 3 shows the results of face recognition. For testing of our model we are using live images.

4.3. Intranet Connectivity

In Android Studio, firstly WifiManger, Formatter and AppCompatActivity packages are imported, In order to get the IP Adress of connected WIFI we first have to invoke the WiFi Manager using getSystemService() method. After invoking, connectionInfo method is used to get all the information regarding the current connected WIFI. IP Address is extracted from the received information and it is then formatted using formatIpAddress() method. The resulted IP Address is then checked against the predefined IP Address and if it's a match the intent of the system moves to Login activity otherwise an error is displayed.

For marking daily attendance the user opens the application, during the splash screen the requirement for proper connectivity is checked if the user is in proper network area then the intent goes to main activity where the user captures his/her photo using front camera, then this photo is checked with the existing images in the databases and if any image is matched then the name of the person is shown in the screen. Now the user has to click on the submit button. One the button is clicked the intent of the application goes to success page where a message is shown to the user. Now the user's attendance is marked and he/she can close the app. If during splash screen the proper network condition is not matched then the intent of the application goes to error page

and an appropriate message is shown to the user. So the user has to go in the prescribed area to mark the daily attendance.

5. Conclusion

In the conclusion, we can say that the online attendance marking system that we build can ease the hassle that occurs in the morning and evening and it can also prevent proxy attendance. The most important benefit of this is that it can be used as an alternative to the fingerprint scanner in this COVID-19 pandemic and even after it. But we do have some computational limitations and with higher computational power more accurate and better results can be found. This application currently runs on Android based smartphone but can be extended further on other platform based smartphones. Since we have taken a small dataset of 1500 images in consideration and used HAAR Cascade for face detection and recognition process therefore we got an accuracy of 43% but with larger data set and advance algorithms like CNN, accuracy of the system can be further increased.

Acknowledgements

References

[1]   L. Cuimei, Q. Zhiliang, J. Nan, and W. Jianhua. Human face detection algorithm via haar cascade classi-fier combined with three additional classifiers. In 2017 13th IEEE International Conference on Electronic Measurement Instruments (ICEMI), pages 483–487, 2017.

[2]   Y. Fujisawa, Y. Otomo, Y. Ogata, Y. Nakamura, R. Fujita, Y. Ishitsuka, R. Watanabe, N. Okiyama, K. Ohara, and M. Fujimoto. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. British Jour-nal of Dermatology, 180(2):373–381, 2019.

[3]   Nirmalya Kar, Dr. Mrinal Deb Barma, Ashim Saha, and Dwijen Pal. Study of implementing automated attendance system using face recognition technique. International Journal of Computer and Communi-cation Engineering, pages 100–103, 01 2012.

[4]   Visar Shehu and Agni Dika. Using real time computer vision algorithms in automatic attendance man-agement systems. pages 397 – 402, 07 2010.

[5]   S. Thakur, J. K. Sing, D. K. Basu, M. Nasipuri, and M. Kundu. Face recognition using principal com-ponent analysis and rbf neural networks. In 2008 First International Conference on Emerging Trends in Engineering and Technology, pages 695–700, 2008.

[6]   E. Varadharajan, R. Dharani, S. Jeevitha, B. Kavinmathi, and S. Hemalatha. Automatic attendance man-agement system using face detection. In 2016 Online International Conference on Green Engineering and Technologies (IC-GET), pages 1–3, 2016.

[7]   P. Wagh, R. Thakare, J. Chaudhari, and S. Patil. Attendance system based on face recognition using eigen face and pca algorithms. In 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), pages 303–308, 2015.

Analysis of Security in Wireless Network: A Survey

Mr. Abhimanyu. D. Sangale [a,1], Dr. Sanjeev Kumar Sharma [a]

[a] *Dept of CSE, Oriental University, Indore, India.*

Abstract. Information security maintenances with the truthfulness, accessibility, and secrecy of knowledge irrespective of the arrangement of the information might take. Data security is frequently branded in different ways like in connected and non-connected security. Non-connected grid security varies with execution prerequisites. The proposed study of worries and safety mechanism stages accessible in non-connected networks. The anticipated representations for non-connected networks are grounded on diverse significances for sanctuary and effortlessness.

Keywords. Security maintenance, non-connected network, data security.

1. Introduction

Sanctuary in non-connected devices means to protect the information or the data from the unauthorized access, discloser, usage, alteration, disturbance, scrutiny, inspection, recording or obliteration. The difference of opinion like computer security, knowledge assurance or guarantee and information security are constantly used in a mixed way or interchangeably. These turfs are interconnected often and stake the mutual goal for shielding the secrecy, truthfulness and accessibility of the info packets, nevertheless there is certain difference between them. [1]

There seems to be a great deal of proprietary knowledge accessible to military, federal departments, administrative offices, hospitals, corporate offices as well as several large enterprises about their staff, clients, goods, analysis and financial status. All of this evidence is collected in a digital format and sent to other organizations throughout the network. So, these days wireless communication is used to transmit this sensitive data, which makes it easy target to other organizations to hack it or modify the data. This paper provides an introduction to assaults as well as the vulnerability of access point in the non-connected network.

---

[1] Mr. Abhimanyu. D. Sangale, Dept of CSE, Oriental University, Indore, India

Email: abhimanyu.sangale@gmail.com.

2. Security Policy in Wirelessnetwork

Non-linked networks aid as equivalent as wired network. In wireless network, as the data is mid-air data transmission, its more susceptible for attacks and illegal uses. So, to protect this data, higher security system needs to be implemented, which is come with lofty price tags. To keep the benchmark, companies have to pay against their operational needs and use the non-connected connectivity. Several kinds of safe guard strategies are also used to secure the wireless network. In sight of each wireless network backdoor, setting up a series of protection plans would essentially prevent security problems such as illegal terminal entry, bogus access points, midway data interception, etc. [2]

2.1. Distinct types of wireless communication attacks:

Often, wireless attacks are categorized as aggressive, passive, near, insider intrusion and service provider assault. Personal data systems and networks provide tempting opportunities and should be prone to diseases from the wide spectrum of malicious hackers, from ransomware to nation-state attackers. There should be a framework that reduces harm and gets better soon when incidents happen. [3]

2.2. Five varieties of attacks:

Aggressive attack: A aggressive attack searches unencrypted traffic and finds confidential data and clear-text codes that can be used in other types of attacks. Traffic collection, insecure communications recording, weakly encrypted traffic decryption, and information capture of password-like authentication are passive risks. Passive network process interception helps adversaries to see future actions. Without the permission or knowledge of the user, passive attacks end in the transmission of information or data files to an attacker.

Active Attack: The intruder seeks to circumvent or hack through secured networks in an active assault. With stealth, malware, worms, or Trojan horses, this can be achieved. Active threats affect attempts to disable or crack security features, implement malicious code, and steal or change data. These attacks are installed against a network backbone, exploit data in transit, electronically break an enclave, or threaten an authenticated remote user during an attempt to connect to an enclave. Discovery or distribution of data files, DoS or alteration of active attack material.

Disseminated Attack: A disseminated intrusion allows the competitor to add malware to a "trusted" component or software, a Trojan horse or back-door program, for example, which will later be spread to several other organizations and clients. Distribution attacks in the warehouse or during distribution depend on the malicious exploitation of hardware or software. These threats cause security vulnerabilities in order to gain unauthorized access at a later date to a sensitive or device feature, such as a back door to a product [4].

Internal Attack: An internal attack includes somebody from the within, including a disgruntled former employee. It may be harmful or not harmful to attack the system malicious activities. Intentionally, dishonest insiders eavesdrop, intercept, or destroy information; use information in a dishonest manner; or deprive all authorized user's

entry. Usually, no malicious attacks arise from carelessness, lack of information, such purposes as executing a mission, deliberate circumvention of privacy.

Close-in Attack: A near-in attack requires someone trying to get physically close to network elements, data, and systems in order to learn more about a network. In order to change, collect, or deny access to information, close-in attacks consist of ordinary individuals who achieve close physical proximity to networks, devices, or services. Unrestricted access or both through surreptitious intrusion into the network, or both, close physical proximity is accomplished. [5]

3. Security Goles for Wirelss Network

   The From this list, we can use different categories of stages that are compatible with our need for protection and utility.

Stage 1 Security: Security at the primary stage is necessary and is translated into every wireless system that may be bought today. It is backed by a 'Wired Equal Privacy' algorithm that is intended to beat most security threats. Details can only be decrypted by the receiver with the right WEP key. Furthermore, WEP is not used to block unintended wireless network connectivity. WEP (wired equivalent privacy): It gives two main elements of defines (authentication and confidentiality).
   WEP used a joint key feature in addition encryption of the RC4 algorithm and used CRC-32's consistency confirmation. Initially, WEP tried to use four data encryption operations, initially 24 - bit vector combination behaves as encoding or decoding just like the key used in 40 - bit WEP algorithm.
   The initiation of a virtual-random number originator serves as resulting key (PRNG). Second, the honesty of the algorithm is shown by the plaintext and again concatenates with the plaintext. The main sequence and ICV results will be handled by the RC4 algorithm. By adding the IV ahead for the encoded text, the final encryption message is generated. WEP tries to decrypt five operations by using them.
   The pre shared key (PSK) and the iteration vector (IV) were initially merged to form a hidden key. Then, the encoded text and concealed key are present in the CR4 algorithm, resulting in a plaintext. Thirdly, it can distinguish the ICV and plaintext. Lastly, Integrity Algorithm gets plaintext to form a replacement ICV and eventually the latest ICV is matched with old ICV.
   WEP feebleness: Imitation of packets can't be vetoed by using WEP. Repetitional attacks can't be vetoed. Attackers can simply record and replay packets as desired and that they are going to be acknowledged as authentic. RC4 used by WEP is inappropriate. Very feeble key is used, in hours to minutes, using readily available tools such as 'air cracking' might be over forced on regular systems. WEP reuses vectors for initialization. Without understanding the encryption key, a dispersion of available cryptanalytic methods will decrypt data. WEP allows an attacker without understanding the encryption key to undetectably change a message.
   However, even with stage 1 protections, there are still many unresolved security risks remaining within a wireless network system.

Stage 2 Security: In stage 1 security, there are some treats or its not completely secure. Its easily compromised.

3.1. Easy Access: If adequate security procedures are not enforced on the network, non-connected LANs are incredibly informal to find and join. Attackers will invade the network without having to enter the facility physically. If the SSIDs are transmitted completely through the system, they could be stopped and enable unsanctioned access (Secure Device Identifiers are allocated to a wireless network).

3.2. Rough access points: In an organization if an access point is installed without the permission of administrator are called rough access point. Access points are most conveniently bought anywhere and mounted. However, appropriated security measured would not be enforced on the system, depending on the person upgrading the access points, fixing an entrance point for attackers and hackers. [7]

3.3. Eavesdropping: It says that, it's an intervention over the non-connected networks of the evidence being communicated. Eavesdropping is also achieved by wireless sniffers, such as software for air dumping.

3.4. Traffic analysis: Facilitates the collection of data transfer and network operation information through monitoring/intercepting wireless communication patterns.

3.5. Data tampering: This explains the possibility that wireless data is frequently collected and discarded during transmission.

3.6. Denial of service (DoS): The communication channel will be chunked by the intruders by using a powerful frequency generator, thus disturbing access to the network.

Stage 3 Security: In Stage 2 security, there are still several risks. Using WPA (Wi-Fi Safe Access), WPA facilitates increased regulation of network access, enhances encryption technologies, and enforces data integrity. Without the users needing to modify the hardware, the WPA arrived with the intention of solving the problems inside the WEP cryptography process. WPA boosts software/firmware over WEP (no new hardware required).

3.1 Personal WPA or WPA-PSK: It is used for domestic usage authentication in SOHO (small office/home office). Personal WPA does not use a server for authentication, so the data encryption key can be up to 256 bits. Keys are often any alphanumeric string and are only used to barter the initial access point session (APs). Both the app and the AP also have this key here. WPA offers reciprocal authentication, meaning the keys are never sent over the air.

3.2 Enterprise WPA: An 802.1x authentication server is formed during this method of authentication mode, generating superb access and protection within the users' wireless network traffic. For authentication, this WPA uses 802.1X+EAP (Extensible Authentication Protocol) but replaces WEP with the more sophisticated TKIP (Temporary Key Incorporation Protocol) encryption again. There is no Pre-Shared (PSK) key used here. TKIP uses the RC4 strategy in an analogous WEP but produces a hash before the RC4 algorithm improves. A duplication of the initialization vector is generated. [8]

One copy is shipped to subsequent step, and therefore the other is hashed (mixed) with the bottom key. After performing the hashing, the result generates the key to the

package that's getting to join the primary copy of the initialization vector, occurring the increment of the algorithm RC4. Then, from the text you just want to cryptography, there is the generation of a sequential key with an XOR, then generating the cryptographic text. Lastly, the message is ready to be received. By inverting the operation, the encryption and decryption will be carried out.

Stage 4 Security: Some more threats and weakness in stage 2 security.

3.7. Encryption vulnerabilities: WPA has certain encryption vulnerabilities; code tampering and masquerading are still not entirely fixed by protection stage 2.

3.8. Compromising performance: device performance

degrades and data transfers and communication speeds are dropped due to intense authentication and encryption protocol computations.

4. Recently Discovered Insecurities in Wireless Network

The 'hole 196' flaw recently found in the WPA2 authentication protocol opened WPA2-secured Wi-Fi networks to insider attacks. This helps insiders to send secure group addressed data traffic using the Group Temporary Key. This data traffic was only meant to be conveyed by Access Point, not by client nodes. But if this GTK traffic is submitted by a malicious insider, he/she can be prepared to upgrade other ARP cache nodes with the aid of ARP request broadcast with GTK. This allows the insider to scent all the user's private information. [9]

5. Variables Addressed During Wireless Network Architecture

We would like to learn about the considerations below when erecting a wireless network in order to design a secure wireless network.
Concentrate on the physical position of the connection point: within the breeze, the wireless signal is spreading; its border definitely does not seem obvious because of the wired network; at any time, the signal will appear beyond the exact coverage. With respect to physical secure, the location of the AP must be measured by means of a special purpose method. We ought to reduce the probability of disclosure of cellular signals beyond the coverage of the network as much as possible.
Access Point control and Management: During a wireless network, where there are multiple APs, the way to control the maintenance and hence the tracking of such APs seems to be highly significant. If there is no cost-effective AP control scheme, the network protection void will be triggered and the attacker will be able to interrupt.
The verification and user identification: While constructing a wireless network, the authentication of wireless network connections to wired network services must be taken into account. To ease user management, enterprise IT network administrators can incorporate the wireless local area network into the existing RADIUS building. It also helps to classify remote consumers in large- scale enterprises. [10]

6. Conclusion

The paper presents numerous security policies in various wireless networks, different from their wireless network and their security concern. Each network has to deliver different protection problems thanks to different features of different networks. Thanks to general security priorities, though, there are several common concerns. This is why a common algorithm is always generated to defend these numerous wireless networks.

References

[1]. Bodhe, Abhijit, Mayur Masut and A. S. Umesh." Wireless LAN security at tacks and CCM protocol with some best practices in deployment of services." International Research Journal of Engineering and Technology (IRJET) 3.1 (2016): 429-436.

[2]. Umesh, Abhijit S. Bodhe1 Dr AS." Rouge Access point:A Threat to Wireless Society."

[3]. Bodhe, Abhijit S., and N. Jagdisha." Security Enhancement Scheme in Mobile Wireless Sensor Networks Using RAPD Approach." 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). IEEE, 2019.

[4]. B.Daya, (2013) "Network Security History, Importance, and future" International Journal of Advance Foundation and Research in Science & Engineering, Volume 1, Issue 3.

[5]. Bodhe and A. Sangale," Network Parameter Analysis; ad hoc WSN for Security Protocol with Fuzzy Logic," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 960-963, doi: 10.1109/ICIRCA48905.2020.9183311.

[6]. http://interscience.in/IJSSAN Vol1Iss1 Zpaper25.pdf

[7]. Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.

[8]. http://www.airtightnetworks.com/WPA2-Hole196

[9]. http://www.wifi.org/knowledge_center/ wpa2

[10]. Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.

[11]. S. Alabady, (2009) "Design and Implementation of a Network Security," Technology, Vol. 1,p. 11.

[12]. Jain, M.K., 2011. Wireless sensor networks: 10. Ning, P., A. Liu and W. Du, 2008. Mitigating DoS Security issues and challenges. International Journal attacks against broadcast authentication in wireless of Computer and Information Technology, sensor networks. ACM Transactions on Sensor 2(1): 62-67

[13]. Dr. Gurjeet Singh (April 2012) "Security Issues in Wireless Broadband Networks" Volume 12 Issue 5 Version 1.0, Double Blind Peer Reviewed International Research Journal.

[14]. Sikkens B. (2008). Security Issues and proposed solutions concerning authentication and authorization for WiMax. 8th twenty student Conference on IT.

Restoring and Enhancing Degraded Underwater Pipelines for Identifying and Detecting Corrosion

Vaibhav A. Parjane [a,1] and Dr. Mohit Gangwar [b]

aPh.D. Research Scholar, Department of Computer Science and Engineering, Bhabha University, Bhopal, Madhya Pradesh, India
bDean-Engineering, Bhabha University, Bhopal, Madhya Pradesh, India

Abstract. Detection of corrosion from underwater images is necessary for oil and gas pipelines to eliminate the internal leakages and hazards. The tests utilized a broad range of underwater pictures of various situations. A modern technique for estimating subsea pipeline corrosion based on the colour of the corroded pipe. For corrupted underwater videos, an image reconstruction and enhancement algorithm is created as a preliminary phase. The created algorithm reduces blurring and improves picture colour and contrast. The improved colours in the imaging details aid in the method of corrosion estimation. In this work we proposed a underwater corrosion detection using image processing techniques. Some machine learning and deep learning techniques have been used for classification of corrosion. In experimental analysis various features have been evaluated for detection of corrosion and it introduces better classification accuracy than traditional approaches.

Keyword. Corrosion detection, Image process, Deep Learning, DCNN

1. Introduction

Therefore, we need detection and testing techniques based on deep learning methods that alert us to the first access to water in surface images and insulation. Deep learning to prepare rust estimates and develop perspectives. Create different corrosion picture parameters of the pipeline under different tests and estimate which utility does not count the execution of trains and testing with different fold assumptions. Exterior sunshine is caused by internal pipeline oxidation, resulting in general corrosion Due to thin internal walls (homogeneous rust), pitting (local rust), and fracture rust, and microbiological rust. The most common problem is insulation repair (ICC) in the oil and gas industry. In the proposed research work, using a deep learning algorithm, design and implement a system for detecting and predicting corrosion on the data set of the underwater pipeline image. The system uses different deep learning algorithms to classify algorithms and predict the possibility of detecting corrosion.

---

[1] Vaibhav A. Parjane, Ph.D. Research Scholar, Department of Computer Science and Engineering, Bhabha University
Email. parjanevaibhav77@gmail.com.

2. Literature Survey

There are many techniques that are being used today to analyze CUI. None of the techniques was applied on their own. For the best performance, several strategies should be used in conjunction with one another. Corrosion under insulation can be detected in many ways, though most is by checking the placement of the connection between the cable and the device. In order to enable ultrasonic testing, it must be expunged. Radiography and total separation are two others widely used approaches. These approaches require the use of a Pulse Eddy Current. The following are some of the most popular techniques for examining CUIs.

2.1. Existing Methodologies

Red-dark channels were previously defined and removed for background light and estimation of the outbreak of this disease. Visualization compensation for object-camera distance retrieval Background and color of objects by analyzing the physical properties of We developed a simple but efficient low-pass filter to debug the point spread function, Debiler Underwater imagery. Different types of water surface images were used under different conditions for experiments. The experimental results indicate that the proposed algorithm effectively Underwater images were recovered while absorbing and dispersing effects [1-4].

The reason for damaging the surface image of the water is to survey sophisticated intelligence algorithms. As a sophisticated method of deicing and refining underwater images, underwater image decorating performance, and color restoration in various ways, the underwater image identifies color evaluation metrics and provides an overview of the critical underwater image applications. Underwater environment, which contains numerous organisms Resources and energy are the main factors needed to sustain sustainability Human development. People often use videos or images to get valuable information when studying the underwater environment. Underwater imagery is the enhancement of contrast widely used techniques for color correction. Contrast is an enhancement of contrast that has attracted much attention in recent years [5-9].

Exploitation is advantageous because they have the ability to use side-scan sonar (SS) for testing underwater meteoric like, i.e., explode (cables and pipelines). When designing houseboats, builders propose (supposedly) to use water vessels that have their own purposes and goals. The issue is fixing the contact lines and under water depth barriers when trying to identify the fixed points of the robots [10]. The original study details of the cable quest results and of the Pipeline simulation experiments are reviewed. An android body that can be programed the work area is small because they have to account for the effect of other elements, and because additional characters are required (supporting characters are required). The effect of realising thematic seabed scenes using AUV Cable measurements (measurements obtained by the AUV) According to the testing of this device, we know that this can be used in an acoustic control system in underwater robots, the use of the pipeline monitoring system in the investigation of movement is supported pressed for change.

3. Proposed System Details

3.1 Problem Statement

In the proposed research work, using deep learning algorithm, design and implement a system for detecting and predicting corrosion on the data set of the underwater pipeline image. The system uses different deep learning algorithms to classify algorithms as well as to predict the possibility of detecting corrosion.

3.2 System Functions

- Creating and developing approaches for corrosion prediction models for oil and gas pipelines using underwater imagery and deep learning approaches.
- Design and build rust detection systems using water-influencing pipeline images.
- Developing an algorithm to detect harmful patches from input images using the DCNN algorithm deep learning foundation.
- Defining the image restoration technique using the CNN approach.
- Exploring and validating the proposed efficacy of the program with certain existing systems.

3.3 System Architecture
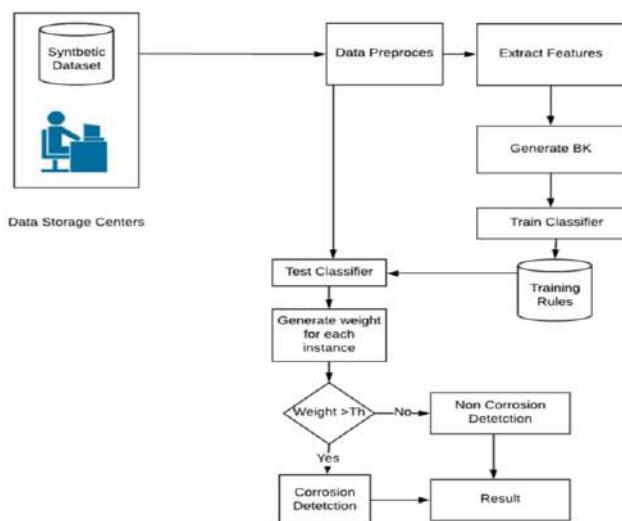
Figure 1. Proposed system architecture

The BK is nothing but a background knowledge that is generated based on extracted values from sensing systems. According to the proposed algorithm, each event gives reward or penalty respectively, based on that each event changes weight state, and based on that system generates BK rules during the execution.

3.4 Algorithm Design

Input. Test Dataset which contains various test instances Test_DBLits [], Train dataset which is built by training phase Train_DBLits[] , Threshold Th.

Step 1. For each read each test instances using below equation TestFeature (m) = $\sum_{m=1}^{n}$(featureSet[A[i] … … …. . A[n] TestDBList)

Step 2. Extract each feature as a hot vector or input neuron from test-Feature(m) using below equation.

Extracted_Feature [t……..n]= $\sum_{m=1}^{n}$((t) ← testFeature)

Extracted-Feature-Set x[t] contains the feature vector of respective domain

Step3. For each read each train instances using below equation

Train_Feature (m) = $\sum_{m=1}^{n}$(featureSet[A[i] … … …. . A[n] TestDBList)

Step 4. Extract each feature as a hot vector or input neuron from test-Feature(m) using below equation.

Extracted_Feature [t……..n]= $\sum_{m=1}^{n}$((t) ← trainFeature)

Extracted-Feature-Set x[t] contains the feature vector of respective domain.

Step 5. Now map each test feature set to all respective training feature set

$Weight = Calscore(FeatureSetX || \sum_{m=1}^{n}(featureSety[y])$

Step 6. return instance [label [weight]

Output. HashMap <class label, Similarity_Weight> all instances which weight violates the threshold score.

4. Results and Discussions

For experiment analysis of proposed system evaluate entire execution in two different open-source platforms. First system creates network simulator environment to generate sensor nodes, the entire simulation log has used as IoT communication log which is generated by various analogue sensors. The techniques basically define in base approach which is carried out to calculate the various parameters between two sensors called train and test feature vectors weight. For each transaction system automatically calculate some values which is denoted in matrix X. K-means clustering unsupervised learning approach has used to generate the data labels and DCNN has used as a supervised learning algorithm.

Table 1. Performance evaluation with DCNN and SVM

	SVM	DCNN
Accuracy	98.92	99.25
Precision	98.67	98.97
Recall	99.33	99.63
F-Score	98.99	99.29
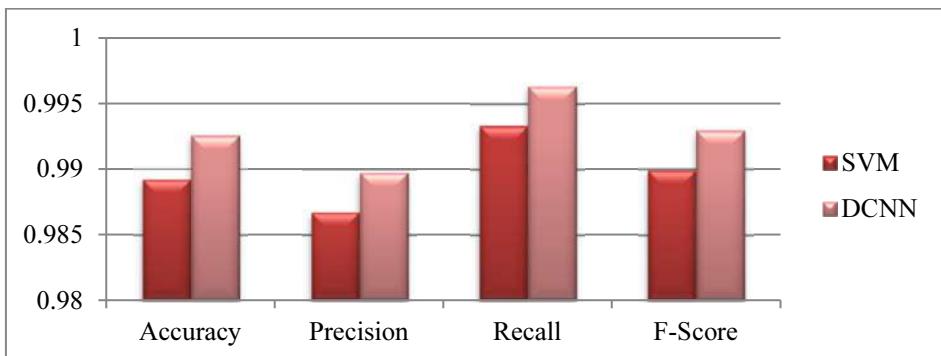
Figure 2. Performance evaluation using Naïve Bayes and SVM

According to above Figure 2, we demonstrate Deep CNN provides better detection accuracy over the Naïve Bayes machine learning algorithm. The CNN has used multi convolutional layers and collaboration with activation function.

5. Conclusions

The primary purpose was to use underwater image processing techniques. For the pipeline network, the same error was found in each file. This research is to be presented. Corrosion on the inside and out has a multitude of different causes, but the structural integrity is required in order to ensure that all of them are considered. Anomalous pipeline wear deposits on this pipeline construction pipeline can be examined to identify pipeline flaw carriers that originate from the excavation of galvanized steel. The mechanism of corrosion is the formation of a galvanic couple between microbe-produced iron sulphides and steel. Using the available information, we can apply this Machine Learning methodology to determine the corrosion. For sub-pip corrosion analysis, a new approach for pip drawing will be implemented using images. Before performing a self-collected reconstruction and improvement, an initial self-check was performed to rule out self-diagnose issues. It can be inferred from the publicly accessible underwater imagery dataset, and it appears that the two are very promising. The experiment was done alongside the exploration of the color drift of the traditional picture measurements. The research will provide data for risk assessment models used for maintenance repair and functions of the pipeline system. The bald question asks the rest of the question some possible perspectives on the safe life of the tube can be applied or developed Studying rust-related cracking and pitting events.

References

[1]   Khan A, Ali SS, Anwer A, Adil SH, Mériaudeau F. Subsea Pipeline Corrosion Estimation by Restoring and Enhancing Degraded Underwater Images. IEEE Access. 2018 Jul 13;6.40585-601.
[2]   Rankin A, Ivanov T, Brennan S. Evaluating the performance of unmanned ground vehicle water detection. InProceedings of the 10th Performance Metrics for Intelligent Systems Workshop 2010 Sep 28 (pp. 305-311).
[3]   L.Developments in sonar technologies and their applications. In2013 IEEE International Underwater Technology Symposium (UT) 2013 Mar 5 (pp. 1-8). IEEE.
[4]   Szyrowski T, Sharma SK, Sutton R, Kennedy GA. Developments in subsea power and telecommunication cables detection. Part 1-Visual and hydroacoustic tracking. Underwater Technology. 2013 Jul 1;31(3).

[5] Bonnin-Pascual F, Ortiz A. On the use of robots and vision technologies for the inspection of vessels. A survey on recent advances. Ocean Engineering. 2019 Oct 15;190.106420.

[6] Motamedi M, Faramarzi F, Duran O. New concept for corrosion inspection of urban pipeline networks by digital image processing. InIECON 2012-38th Annual Conference on IEEE Industrial Electronics Society 2012 Oct 25 (pp. 1551-1556). IEEE.

[7] Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.

[8] Gao, Y.; Liu, Y.; Ma, Y.; Cheng, X.; Yang, J. Application of the differentiation process into the correlation-based leak detection in urban pipeline networks. Mech. Syst. Signal Process. 2018, 112, 251–264.

[9] Yin, S.; Weng, Y.; Song, Z.; Cheng, B.; Gu, H.; Wang, H.; Yao, J. Mass transfer characteristics of pipeline leak-before-break in a nuclear power station. Appl. Therm. Eng. 2018, 142, 194–202.

[10] Scott, S.L.; Barrufet, M.A. Worldwide Assessment of Industry Leak Detection Capabilities for Single & Multiphase Pipelines. Offshore Technology Research Center College Station. 2003. Available online. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.118.6455&rep=rep1&type=pdf (accessed on 17 December 2018)

Architecture for Secure Communication Among IoT Devices with Ethereum Blockchain

Bawankar Chetan D [a,1], Dr. Sanjeev Kumar Sharma [b]

[a] Sanjivani COE, Dept. of IT & Research Scholar Dept. of Computer Science and Engineering Oriental University Indore, India
[b] Dept. of Computer Science and Engineering Oriental University Indore, India

Abstract. The paper aims to clarify the relationship between Internet-of-Things devices and Ethereum blockchain. It proposes the arrangement to ensure information transmission among parties in an open system of IoT must be secure using Ethereum. The accompanying joining strategy utilized terminal gadgets as system innovation and Ethereum blockchain stage that delivered back-end, which guarantees high security, accessibility, and protection, supplanting conventional back-end frameworks. The following issues should be considered to prevent the malicious hub from attacking, resist distributed denial-of-service attacks, and prevent firmware backdoor access. This paper proposed a system in which the Peer-to-Peer authentication model, where every IoT node in the system must be authenticated and verified by the proposed framework. The paper provides empirical insights into IoT nodes manufactured in bulk, and they are remaining with their default username and password.

Keywords. Performance Measurement Authentication, Botnet, Blockchain, Data Protection, Ethereum, Smart contacts

1. Introduction

Ethereum Blockchain and Internet of Things (IoT) are used for creating new ideas and innovations. All the while, they change contemplations and make extra chances, each in their individual conditions, and there is an opportunity for applications that can be shared with standard attributes of IoT and Ethereum, the significant advantage of decentralized for IoT hub/nodes is with Blockchain.[1] Right now, Blockchain likewise falls along this line of research may be used to authorize, authenticate and audit the data generated by IoT hub/nodes. Like- wise, thinking about decentralized nature takes out the need to trust in the untouchable and doesn't have a particular clarification behind failure. In proposed IoT Hub/hubs are relied upon to be online and ought to speak with one another.

---

[1] Bawankar Chetan D, Research Scholar Dept. of Computer Science and Engineering Oriental University Indore, Email: chetan251htc@gmail.com.

In Private blockchains, hubs/nodes are foreordained, while in Public blockchains, any hub/node can leave or join the network. Ethereum is a Public blockchain variation where each exchange has a cost estimated in the wording of" Gas." Ethereum receives the PoW consensus algorithm.[2]

To manage the security issues of IoT, one needs a few stages which could give both security and integrity of the information being imparted between the IoT gadgets. In the Beginning, the advancement which is exhibiting an enormous achievement in the area of assurance and security of data for clients in Blockchain is improved. Blockchain innovation has a decentralized network with trustless in more minor conditions. It doesn't require outside interference. Having this sort of condition, the information is still verified in the Blockchain as it uses Cryptographic hashing to verify information in the distributed ledger. Blockchain as a passed-on record advancement makes a trustless situation that can thoroughly clean the dependence of trusted third parties. Each IoT hubs has the stability to cutoff information through a Consensus computation in the Ethereum blockchain network. The structure is secure if the reasonable center points control inside and out more hash rate than any interest social affair of aggressor hub. The informational index aside of Blockchain can't be changed, regardless of whether the aggressor begins from the internal framework.[3]

The blockchain advancement gives reasonable courses of action to the verification and security protection of IoT hub/node. The edge arrangement and board (management) of Blockchain is moderately below average.[4] Regardless of whether intelligent gadgets have constrained processing assets, Blockchain may be developed. Using Blockchain in the IoT makes the system decentralized, which can significantly improve the system's security. It is essential to spare contraption ID, public key, and the hash value of critical information in the record before gadget access to the IoT network. Using cryptography, shared certification with partners to check the rightness of the data out between a public key and a contraption ID. [5] Comparing hash values of contraption ID information, any difference in the data can be perceived promptly.[6]

The paper is arranged as mentioned:

2.   Section Presents Background and Related work done in the field of Botnet Internet of Things (BIoT)

3.   Section Presents the Problem Identification and Proposed System architecture

4.   Section Proposed System Implementation details

5.   Section Addressed the challenges in implementations

6.   Section Concludes and the future scope.

2. Background and Related Work

Blockchain is a decentralized appropriated database of changeless records, where exchanges are ensured by robust Cryptographic calculations, and the system status is

kept up by the Consensus algorithm. Examination and investigation of security gaps in the brilliant home system layer and proposed arrangements. It is conceivable to oversee and confirm gadgets utilizing an ISP and oversee intelligent home gadgets in an external web condition. In any case, because of the absence of client information for examination, this methodology isn't successful in guaranteeing the security of the inside Internet condition. Use Wireshark to check home horizons and lack of confidential data for devices such as free detectors when building intelligent home devices. This approach offers software-specific solutions but cannot compete with software other than Wireshark.

3. Problem Identification and Proposed Methodology

1. A novel authentication method for IoT hub/node based on Ethereum for secure communication over IoT network.
2. To Implement smart contract and validation of node in proposed authentication framework.

The Proposed system comprises of two main parts as   follows

3.1. Peer-to-Peer Authentication Method

Blockchain is an open, secure, and distributed exchange ledger innovation, can flexibly adjust to complex also changing system conditions. The failure of a few hubs/nodes doesn't influence the ongoing activity of the framework. Distributed authentication between hubs prevents malicious hubs from attacking the system. Regardless of whether few hubs are undermined, the ledger won't be altered. In a multi-node network, the identity data of the hub/node should be enlisted in the blockchain   each time a new hub/node is included. Every hub/node ID, public key, a hash of basic information, and other data are put away in the blockchain ledger.

Simultaneously, every IoT hub/node is a major component in the Ethereum blockchain network, and the consensus component ensures that every hub stores similar data. At whatever point, Peer-to-Peer Authentication Method correspondence happens via public-key cryptography, which can be utilized for validation in IoT networks.[5] The framework process is mainly divided into three stages. All gadgets need to finish the enlistment in the blockchain before authentication. When a new node wishes to access, it will first check the list mentioned in the blockchain. Once getting authentication from the blockchain gadget will be checked for trustworthiness with the hash value of data to find the potential intrusion behavior.

3.2. Independent Framework to detects Botnet

The proposed framework contains three significant parts Ethereum Blockchain, Hosts, and Self-governing System (SS).

•   **Hosts:** Nodes which are associated with the web using an SS. Two types of hosts are their general and IoT nodes. IoT nodes have devices like sensors, actuators, and many more devices, which may play an important role in the network and send/receive data remotely. As botnet misuses the vulnerabilities of IoT has; consequently, they are increasingly helpless against attacks. [2]

•   **Self-governing System:** SS is one of the significant segments of the proposed framework. The botnet moderation component is executed in the SS. An SS is mindful

of advancing the packets outside of the system. SS stores the list of host and threshold values per IoT hub/nodes. Four records are looked after" Blacklist"," Whitelist"," Reckoned Attacker List" and" Conceivable Victim List". The SS are associated with one another using the Ethereum blockchain and exchange arrangements of IP address. Each SS monitors the number of packets sent and received by the host. This edge esteem is invigorated at regular intervals.

•   **Blockchain:** The SS is associated with the Ethereum blockchain. They communicate with one another in the type of blockchain ex- changes. For the most part, in Ethereum information is communicated when the block exchanges.

4. Implementation Details

This section addresses the particular usage of the framework, which includes the Ethereum blockchain, the procedure of the asymmetric key generation; at last, security analysis of the framework is given as   below. Session key KSAB is shared with both. This received key will be used by both parties for secure communication.

Algorithm 1: Proposed method for Key Exchange

1 Aryan will request to Ethereum for public-key of Agrani PuKB
2 Ethereum will check for authentication of IoT hub/node
3 if IoT hub/ node is in 'BlackList'   then
4 go to step 23
5 else if IoT hub/ node is in 'reckoned Attacket List' then
6 go to step 23
7 else
8 Send the Public-key of Agrani PuKB to Aryan
9 Aryan will encrypt a" Number only used once (Nonce)" N1 that solely identifies
   a transaction and Aryans IDA with the PuKB of Agrani is given by EPuKB (N 1,
   IDA)
10 Agrani will request to Ethereum for PuKA of Aryan
11 Ethereum blockchain will check the authentication of IoT hub/node
12 if IoT hub/ node is in 'BlackList'   then
13 go to step 23
14 else if IoT hub/ node is in 'reckoned Attacket List' then
15 go to step 23
16 else
17 Send the Public-key of Aryan PuKA to Agrani
18 Agrani will encrypts a" Number only used once (Nonce)" N2 that solely
   identifies a transaction and Agranis IDB with the PuA key of Aryan is given by
   EPuKA (N 2, IDB)
19 Agrani send to Aryan the already received Nonce (N 1) and second Nonce (N
   2) is encrypted with Aryan PuA key. EPuA (N 1, N 2)
20 Aryan acknowledge with received Nonce(N2) to Agrani which is encrypted by
   Agrani's public-key PuB. EPuKB (N 2)
21 Aryan will now be generated secure key KSAB which is signed with Aryan
   PrKA and the complete message is encrypted with Agrani PuKB EPuKB
   (PrKAKSAB, N2)

22 Agrani will decrypts the received message by accessing both key of Agrani PrKB and verifies the digital signature using Aryan PuKA key

23 Add the user in Black list and report the same.
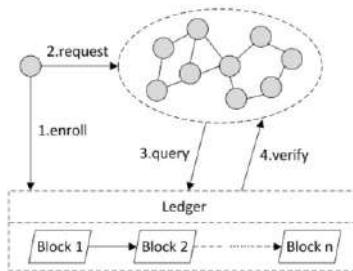
Figure 1. Working Process of Proposed method

4.1. Framework of blockchain network

Each SS keeps up four unique types of records. Since botnet explicitly focuses on the IoT hub/nodes, IP locations of IoT hub/node have SS, which includes the list of the reckoned attacker; meanwhile, other hosts' IP addresses are included in the whitelist. Different lists are updated to   SS using the Ethereum blockchain.

4.2. Private key generation

Random number generator likewise needs a seed from a source with ad- equate entropy count. An irregular seed is gathered by different data of the IoT gadgets, for example, free space, I/O delay, memory status, procedures running, CPU frequency, and many more. By using the RSA algorithm, the Public key is generated from the Private key.

5. Challenges of IoT hub/node with proposed blockchain solutions

Many applications use blockchain in IoT framework. A few difficulties are with security of IoT hub/nodes and their solutions with use of blockchain Privacy of IoT Hub/nodes IoT hubs are helpless against uncovering the private information of the user. To address this issue permission Ethereum used to provide security to IoT hub/nodes. [8] Traffic and Cost To deal with rapid changes in IoT gadgets. For this purpose, decentralization using Ethereum. All devices present in the system are directly connected and communicate with a peer rather than a central server. Shoddy architecture Every component of IoT gadgets has a state of washout that influences the network. Using blockchain, all nodes in the web will be verified. The data is also cryptographically secure. Verification is a framework by which a framework decides whether the customer has certain benefits. Verification arranged into a group of three as mention below: what-you-have (possession), who-you-are (ownership), and what-you-know (data).[10]

Data Protection/manipulation In IoT hub/nodes, guaranteeing the unwavering quality of the IoT gadgets gives security assurance. Regardless of whether a gadget has passed the authentication of different hubs, it, despite everything, has the danger of being assaulted by intruder clients because of programming or framework

vulnerabilities during the execution of the task. The intruder normally will change the system element to leave the second passage in the gadget to get ready for consequent invasion and adjust the key arrangement record in the gadget and prompt harm to the whole system. Due to the use of blockchain the Data protection/Manipulation, if any node updates the data, the framework will reject the manipulation.[11]

6. Conclusion and Future Scope

This paper addressed various challenges of IoT hub/node, which are having possible solutions using blockchain. Additionally, investigated the downside of the present IoT center/hubs in personality confirmation and security zone. The structure of multi-chain gives add-on security between various IoT hubs in a network. This paper proposes an Ethereum based response for the issue present in secure communication between IoT Hub/center points. Future work will concentrate more on the administration of IoT sensors versus the customary budgetary database in the blockchain. We aim to employ Edge computing, Fog computing, and Machine learning algorithms for the botnet detection method.

Acknowledgement

References

[1] S. G. Kumar, A. Murugan, B. Muruganantham, and B. Sriman, "IoT–smart contracts in data trusted exchange supplied chain based on block chain," International Journal of Electrical and Computer Engineer- ing (IJECE), vol. 10, no. 1, p. 438, feb 2020.

[2] Z. Ahmed, S. M. Danish, H. K. Qureshi, and M. Lestas, "Protecting IoTs from mirai botnet attacks using blockchains," in 2019 IEEE 24th Interna- tional Workshop on Computer Aided Modeling and Design of Communica- tion Links and Networks (CAMAD). IEEE, sep 2019.

[3] S. K. T. Mehedi, A. A. M. Shamim, and M. B. A. Miah, "Blockchain-based security management of IoT infrastructure with ethereum transactions," Iran Journal of Computer Science, vol. 2, no. 3, pp. 189–195, jul 2019.

[4] K. Košťal, P. Helebrandt, M. Belluš, M. Ries, and I. Kotuliak, "Manage- ment and monitoring of IoT devices using blockchain," Sensors, vol. 19, no. 4, p. 856, feb 2019.

[5] D. Li, W. Peng, W. Deng, and F. Gai, "A blockchain-based authentication and security mechanism for IoT," in 2018 27th International Conference on Computer Communication and Networks (ICCCN). IEEE, jul 2018.

[6] M. El-Hajj, A. Fadlallah, M. Chamoun, and A. Serhrouchni, "Ethereum for secure authentication of IoT using pre-shared keys (PSKs)," in 2019 International Conference on Wireless Networks and Mobile Communica- tions (WINCOM). IEEE, oct 2019.

[7] H. Akhundov, E. van der Sluis, S. Hamdioui, and M. Taouil, "Public-key based authentication architecture for IoT devices using PUF," in 6th In- ternational Conference on Computer Science, Engineering and Information Technology (CSEIT-2019). Aircc Publishing Corporation, nov 2019.

[8] O. Alphand, M. Amoretti, T. Claeys, S. Dall'Asta, A. Duda, G. Ferrari, F. Rousseau, B. Tourancheau, L. Veltri, and F. Zanichelli, "IoTChain: A blockchain security architecture for the internet of things," in 2018 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, apr 2018.

[9] M. D. A. Dhaya and R. Ravi, "Multi feature behavior approximation model based efficient botnet detection to mitigate financial frauds," Journal of Ambient Intelligence and Humanized Computing, jan 2020.

[10] L. Hang and D.-H. Kim, "Design and implementation of an integrated IoT blockchain platform for sensing data integrity," Sensors, vol. 19, no. 10, p. 2228, may 2019.

[11] J. H¨oglund, S. Lindemer, M. Furuhed, and S. Raza, "PKI4iot: Towards public key infrastructure for the internet of things," Computers & Security, vol. 89, p. 101658, feb 2020.

[12] O. Pal, B. Alam, V. Thakur, and S. Singh, "Key management for blockchain technology," ICT Express, aug 2019.

Natural Language Querying and Visualization System

Trishali Banerjee [a,1], Upasana Bhattacharjee [a] and Mrs. K. R. Jansi [a]

[a]*Student, Dept. of CSE, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India*

Abstract. Data is the new gold; everything is data driven. But it is impossible for everyone to possess technical skills to be able to write queries and know different python tools used for data visualizations. The process of extracting information from a database is a mammoth task for non-technical users as it requires one to have extensive knowledge of DBMS language. But these data and visualizations are required for various everyday presentations and interactions in the professional world. This application would enable the users to overcome these obstacles. Our project aims at integrating two systems, an NLP interface to fetch data from simple English queries, and a second system where the fetched data with the help of natural language processing is used to form visualizations as demanded by the users will be created. This system would essentially help the people who are not techno-savvy or are not in the field of tech to interact with data using simple English.

Keywords. Natural Language Processing; Database; Python; SQL; Visualization

1.   Introduction

Natural Language Processing basically helps us talk to a machine using the languages that are easily interpreted by human beings. It's an artificially intelligent technology. Imagine if a database were a person with whom you could actually talk to. Yes, NLP makes that possible. In recent times as the world has made a huge progress in this field, where processing huge amounts of textual information is done with an acceptable amount of efficacy. If we have to take an example, consider Web Search Engines, where these techniques form an essential component and thus served as an inspiration for us to make this querying and visualization system. Our application will provide the user with an interface to do just that. By using simple English sentences, interaction with the database and generation of queries will become possible. NLP integrated with Python will also help us to visualize data.

Visualizations will be created using simple English sentences. We have harnessed the immense capabilities of the python language as it is proven to be the best language for Machine Learning and artificial intelligence. A parsing model TAPAS which is out sourced by google has been used in training the machine, it extends BERT architecture

---

[1] Trishali Banerjee, Dept. of CSE, SRM Institute of Science and Technology, Chennai
Email. trishalibanerjee44@gmail.com.

which develops the Neural Network required for the purpose. The second approach is using Spacy which is again an open sourced module by python.

2.   Related Work

Author Garima Singh and Arun Solanki in their paper "An algorithm to transform natural language into SQL queries for relational databases" have talked about ambiguity removal, that is a very important thing that they have addressed, the removal of repetitive attributes and clarifying with the user in an interactive way improves the accuracy and in that way the dependency on the machine also reduces and the flexibility increases. [11]. Santosh Waghmode and his students in their paper "SQL Query Formation for Database System using NLP", have used a unique Multinomial Logistic Regression Algorithm, this algorithm aims at predicting the type of query from the information that the user provides in the previous steps and they have also mentioned about the regular dictionary update that helps in achieving maximum accuracy.[10] For this project, the dictionary will be updated regularly not only for accuracy but also to address ambiguity. But the drawback of this system is that there exist a lot of complex queries that they have not addressed yet. Swapnil Kanhe, Pramod Bodke, Vaibhav Udawant and Akshay Chikhale in their paper "SQL Generation and PL/SQL Execution from Natural Language Processing", have mentioned their software will be able to generate both PL/SQL queries. In their algorithm they mention about a "Morphological Analysis"- Morphology is the structure of word. It is also concerned with derivation of new words from existing ones, Ex. Lighthouse (formed from light house). [2] They have also made sure the tables in the database are normalized for more accuracy. They have used c# .net frame work for the development of this application. George Obaido, Abejide Ade-Ibijola and Hima Vadapalli in their paper "Synthesis of SQL Queries from Narrations", have addressed the fact that SQL is not an easy language to master and their system would help the non- technical end user to interact with the RDBMS. They have created a color coded table and to convert NLP to query entities with matching color is found. They have also performed a 2 fold evaluation of the tool. First was using the crowdsourced XNorthwind dataset and second was using human subjects. [9]

3.   Querying and Visualization System and Architecture

The natural language querying and visualization system is aimed at fetching answers to the queries then using that to visualize the data. The user inputs queries as simple English sentences, techniques such as matching, parsing, and tokenizing and dictionary mapping are utilized to generate the relevant query. This system will help in solving moderately difficult queries and as the system keeps learning the accuracy increases accordingly. The previous systems were based on forming queries and then fetching data. The formation of logical queries is difficult and according to the previously available data the accuracy is not satisfactory. Here the predictions based on denotations are done by selecting corresponding cells that are formed into vectors and also perform aggregations based on the user requirements. Also added is the visualization system, to give more meaning to the data and to give user an end to end

experience. An additional feature to this project being the use of Google Collab and Google Cloud for implementation.

In Figure 1, we can see the user gives input to the machine through the user interface. The data in the csv format is converted into data frame using pandas. Then the NLP pre-processing begins. It starts with breaking the input into individual words or sometimes phrases. These are then associated with a token id. Lemmatisation and stemming leads to the root word and clubs the words eliminating the suffixes in the word like 'ed 'or 's' or 'ing'. This helps reduce confusion. Then comes removal of stop words. Then tagging begins. Here basically the grammatical meaning is attached to the input. The tokens generated are tagged as nouns, pronouns, verbs etc. The data frame post this is converted into list of lists, this basically allows the tagging of user input to the data set to be easier and faster. Here the approach is different, the algorithm used is recurrent neural network and the modules used are tapas module open sourced by google and spacy which is also open sourced. The table is queried according to the meaning that has been attached to the user input and it is mapped to the required subset / value in the data set. In the result file the rows and columns are enumerated.

For the visualization again the data frame is formed, the user inputs the requirements and this is then passed through the matcher and we obtain the visualization appended in the GUI itself.
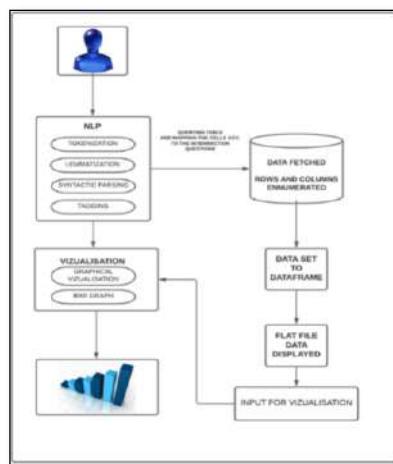
Figure 1. System Architecture

4.   Modules

4.1.   *User input and Natural Language Query Pre-processing.*

This module deals with the creation of an interface where the users will be able to give their inputs. Anyone who doesn't know how to write SQL queries can use this interface for the purpose of interacting with the database, the input as already mentioned will be a simple English sentence. It is very important to note that for NL processing everything has to be converted into string. Tkinter is used for the creation of graphical

user interface. The system first starts with pre-processing the natural language query entered, by carrying out the following.

Tokenization. The interface accepts input from the user in form of English sentences. The NLP system breaks the raw text input into small portions and generally separates each word. Here NLTK library is used to perform tokenization. Spacy Tokenizer has been used which is a modern technique for tokenization; much faster and easier, the advantage of using this method is one can specify special tokens that need not be segmented as well as those that typically have to be specified. Basically the tokens that one specifies takes precedence. There are many challenges during tokenization like there are way too many word boundaries available in the spread of English language.

Lemmatization. Lemmatization and stemming prune a word to its root. Let's say there is a phrase, "Display the names of managers who work in Bangalore" here, "names" becomes "name". Basically the words are converted to their base or root words. It is done to remove the endings of inflected words to return the base form of the word known as lemma. Lemmatisation and stemming contribute to reducing the number of common words (same words in different forms) and these in turn help amplify the accuracy of the tagging process.

Removal of Escape Words or stop words. Escape words are those words that are unnecessary for query generation and can be overlooked. Here is a list of few possible escape words that have been defined (tf_examples_utils is imported from tapas repository, it contains the stop words) A, An, The, Select, Find, Which, whose, Is, Of, A, With, To, for, Are, And, What.

Removing Ambiguous Attributes. Suppose there are 2 sentences, "display the students in the blue house" and "select the students who have the blue house as their reference location". In these 2 sentences "blue house" creates an ambiguity, in these cases an error alert is given out if the dictionary being used is unable to handle this.

Tagging. It basically pertains to converting a sentence to list of words or tuples in the form of "word, tag". It assigns grammatical information of a word. In simple words its parts of speech tagging, it tell us what part of the speech does the word belong to. Post tokenisation the tokens are classified into nouns, pronouns, verb or the type of variable mentioned. double, integer, string. Then the complete thing is passed through an IDENTIFIER where the generated tokens are classified into relations, attributes and clauses.

Dependency Parsing. It is a process of analyzing the grammatical structure or the grammatical construct of the sentence and also analysing what are the dependencies between the words in the sentence. There are various tags. .

Named-entity recognition. NER also called entity identification. It is a subset or sub task.

Name Stream. Proper nouns are crucial for query generation and visualization. Spotting them in English is also easy since they start with a capital letter, also if they comprise of a group of words the first letter is always capitalized such group of words are not broken down and treated as a single token. We also save these words token wise (dual approach) that helps us with the ambiguity just in case if any of the word is missing or if it is the same thing reframed in a different way.

4.2. Formation of a data frame

The first important step is to convert the data set into python data frame. This will allow the modules and python function to work on the data frame. The flow of the software is after the data is fetched post tagging, it is either displayed like a flat file or the user demands for visualization. Data is first converted into a data frame. From here there are two possibilities either using tensorboard for creating visualization or matplotlib. Here Pandas is used for the formation of the data frames. The original data is in the .csv file format that is converted into a data frame with rows and columns. The query entered by the user is tagged to these tuples. The data frame is converted into a "list of lists." That way the tagging becomes way easier and also increases accuracy.

4.3. Fetch data on the basis of the Natural Query

This system, as already mentioned uses TAPAS module and imports tensorflow from the python libraries (there are many imports but they are quite generic). Tensorflow incorporates the deep learning algorithm- **Recurrent Neural Network and to be more specific Long Short- Term Networks (LTSMs).**This helps in mapping the complete NL query or as the system calls it an "Interaction Question". Pre-processing the question is marked with an id. Essentially each cell in this algorithm is converted into a vector and the algorithm according to the user input maps it to the particular cell "answer coordinates" it wants to point towards. Also used is an "annotator", it keeps track of the question and table pair. This helps in case when questions are repeated or rephrased as something else but mean the same. There are 2 approaches that is used to fetch data, the first one is discussed above (tapas module). The second software library used is Spacy.

4.4. Input about the visualizations

Regardless of the imports being used, deciding the requirements of user-input for the visualisations is slightly difficult as each plot is built separately and it is a work in progress. The user requirements are tagged and the interface is updated accordingly. Seaborne and Matplotlib are the 2 python modules used for the visualisation and tagging to get the input to match the visualisation requirements.

5. Algorithm

Tapas uses a Bert file for the creation of the neural network. The Artificial Neural Network or the ANN works identical to a human brain. Specifically the software uses Long Short Term Memory Network Algorithm. NLP pre-processing is triggered using the artificial neural networks. There are different layers to the network generated. There are multiple inputs given to the system. In this case a stream of words and the output obtained is a single output, here the "query". The strategic reasoning is done with the help of a radial basis function. An algorithm which remembers the order of dependencies for the sequence of input, this helps in tagging the terms with accurate weight and in the long run helps in ranking of terms in accordance of their importance. LTSM it can choose to modify and remember or forget a piece of information selectively which makes sequential processing easier. LTSM works on a Cell State mechanism and there are 3 states to it. First, the previous cell state which describes the

information that was present in the cell in the previously timed step. Second, the previous hidden state describes the output of the hidden state previously. Third, the current input. The input is passed through the different layers of the Network. The network used is Feedback network because feed forward network thinks in one direction only and never looks back, that is an important functionality of the algorithm to be used because we are looking at a Sequential stream of data. There are 2 parts in an RNN first is an encoder it is responsible for encoding symbol sequences to fixed length representation in vector format. The data set, the list of list that was generated is converted into vector format ready to be tagged and queried. The decoder does the opposite, brings down the symbol to be displayed or used further for visualisation. In addition to that Matcher. Match is used in spacy .Here the user inputs an argument and a group of possible cases to match with. There are 4 types of cases available but we have used "values". Values because the argument is matched against a list of values specified. Spacy uses Deep Neural Network that is based out of Convolution Neural network. It is built over a frame work Embed, encode, attend and finally predict. Bloom filter and shift reduce parser is used, the hash codes of the words are generated and kept in the dictionary, this helps in maintaining a compact dictionary, the collision is also less hence ambiguity removal. Then DNN is used to encode list of words to a matrix of sentences and ids are generated. Attend is removal of stop-words and maintaining only the parts which contribute to the meaning. Processing through the remaining layers and tagging results are generated.

6.   Results

The data is being fetched successfully, and is accurate to the best of knowledge. Since, the machine is still in the learning phase, to reach desired accuracy the dictionary in use has to be regularly updated.

Figure 2. Output from Cricket Dataset

Figure 2 gives an idea of the output being generated. The system currently can seamlessly produce bar graphs to provide the users with a visualization of the data. Figure 2 shows the output carried out on a Cricket data set. It can be seen here that the user queries the database with variety of questions .It can also be seen here that the system is capable of answering progressive questions involving pronouns, like, "What is his highest score?" etc.

Figure 3. Output from Employees dataset.

Figure 3 shows the output from an Employee dataset

Figure 4. Generation of question ID.

Figure 4 shows how the Question id is generated. The role of generating a question id is the next time a similar question is asked by the user, the id is fetched and the results are returned in a shorter period of time. Hence, question id helps make the process more efficient.
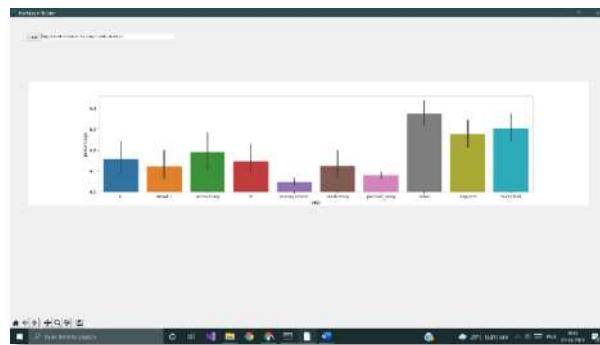
Figure 5. Visualization Output.

Figure 5 shows the visualization that is generated on the Employee dataset to satisfy the user's requirements.

7.   Conclusion

NLP is one of the few tools which can change the complete working of the computer program interface. This system can not only query the database and return the required results to the user but also provide the user with the option to visualize their data in a way that will make the results of their query even clearer. Dealing with ambiguity removal and constant update of the dictionary being used helps improve the accuracy of the system. Currently, this system is capable of producing bar graphs to visualize the data. Since all forms of visualization have separate requirements, the future scope of the project is to include varieties of visualizations and also refining the library of querie

References

[1] Dey, S., Gosh, P.K. and Sengupta, S. Automatic SQL Query Formation from Natural Language Query. International Journal of Computer Applications (0975-8887) & International Conference on Microelectronics, Circuits and Systems (MICRO-2014).

[2] Bodke, P., Chikhale, A., Kanhe, S. and Udawant, V. SQL Generation and PL/SQL Execution from Natural Language Processing. International Jornal of Engineering Research & Technology (IJERT) ISSN. 2278- 0181, Vol. 4 Issue 02, February-2015. URL. www.ijert.org.

[3] Nguyen, D.Q.[Dat], Nguyen, D.Q.[Dai] and Pharm, S.B. Ripple Down Rules for Question Answering. Semantic Web 0 (2015) 1-0 IOS Press.

[4] Goswami, A., Gupta, P., Koul, S. and Sartape, K. IQS-Intelligent Querying System using Natural Language Processing. International Conference on Electronics, Communication and Aerospace Technology ICECA 2017.

[5] Srinivasan, A. and Stasko, J. Natural Language Interfaces for Data Analysis with Visualization. Considering What Has and Could Be Asked. Eurographics Conference on Visualization (EuroVis) 2017.

[6] Patil, S.J. Python- Using Database and SQL. International Journal of Science and Research (IJSR) ISSN. 2319-7064 Impact Factor (2018). 7.426.

[7] Haribhakta, Y., Kamle, V., Kariya, S., Pagrut, A. and Paknode, I. Automated Sql Query Generator By Understanding A Natural Language Statement. International Journal on Natural Language Computing (IJNLC) Vol. 7, No. 3, June 2018.

[8] Kombade, C., More, M., Pujari, A. and Patil, S. Natural Language Processing with some abbreviation to SQL. International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN. 2321- 9653; IC Value. 45.98; SJ Impact Factor. 7.177. Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com.

[9] Potharaju, S. P., & Sreedevi, M. (2017). A Novel Clustering Based Candidate Feature Selection Framework Using Correlation Coefficient for Improving Classification Performance. Journal of Engineering Science & Technology Review, 10(6).

3D Animation and Virtual Reality Integrated Cognitive Computing for Teaching and Learning in Higher Education

Abhishek Kumar [a,1], Rini Dey [a], G. Madhukar Rao [b], Saravanan Pitchai [c], K. Vengatesan [d], V. D Ambeth Kumar [e]

[a] *School of Computer Science and IT, JAIN (Deemed to be University), Bangalore, India*

[b] *Department of Computer Science & Engineering, Koneru Laksmaiah Education Foundation, Hyderabad*, India

[c] *Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur*, India

[d] *Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon*, India

[e] *Department of Computer Science & Engineering, Panimalar Engineering College, Anna University, Chennai*, India

Abstract. This paper proposes the and justify how we can enhance the quality of medical education through immersive learning and AI (Artificial Intelligence) use in education. A Multimodal Approach for Immersive Teaching and learning through Animation, AR (Augmented Reality) & VR (Virtual Reality) is aimed at providing specifically medical students with knowledge, skills, and understanding. It is important to understand the current challenge involved in medical education. This paper reports the findings of a novel study on the technology enable teaching with Animation, AR and VR by and MR impact. A case study was conducted involving 521 participants from different states of India. The data was analyzed by their feedback after using this Virtual reality-based teaching procedure in classroom. Recommendations from this paper that are expected to effectively improving the quality of medical education in faster way.

Keywords. 3D Animation, Cognitive Computing, AI, Virtual Reality

1. Introduction

At present, Innovation and technology are playing an essential role in the enhancement of the standard of education and students' learning outcome. Creativity and Innovation in education have improved a lot recently. One of the rising advancements that have started more prominent enthusiasm for educational technology is Virtual Reality, Augmented Reality, and Mixed Reality. Future educators must have a satisfactory result when implementing this immersive technology in education. Since this is a matter which concerns the current enthusiasm on a universal level to reduce use the technology in education [2–4]. "Learning will be significantly changed by AI," says UNESCO Director-General Audrey Azulay.

[1] Abhishek Kumar, School of Computer Science and IT, JAIN (Deemed to be University), Bangalore
Email: abhishek.maacindia@gmail.com.

"Showing apparatuses, methods for learning, access to information, and instructor preparing will be upset." [5] The Fourth Industrial Revolution serves as a comprehensive unit for terminologies of 3D – Printing, Artificial Intelligence, BIG data, and the Internet of things, in the education sector. For partnerships, it implies re-examining procedures and auto-cannibalization of plans of action. Its impacts on work and business are a gauge to be mind-boggling, conceivably elevating disparity by diminishing interest for low degrees of abilities. Utilizing savvy AI-fueled frameworks can enormously trigger the fruitfulness of innumerable instructive foundations, reduce their operative costs and provide noteworthy perceivability into payments and expenses. Implementing these technologies in education improves the general responsiveness of students and the excellent performance result."[6].

2. AR, VR AND MR In Higher Education

Virtual Reality is starting to change the classroom from 2017 worldwide. Harder, Better, Faster, Stronger (HBFS) are the four key words that can most efficiently depict the potential effect that the usage of VR and AR can make in the framework of instructions. Innovations are required within the training framework to remain relevant in the ever-changing world [7-8].
Virtual and Augmented Reality is the following legitimate advance in the development of the Education System. This Reality of training will be significantly affected by the presentation of new AI-based advances, and that is the Virtual Reality-based real-time learning experience [9-12]. Higher Education has seen an upsurge in the application of Artificial Intelligence (AI) with huge consideration over the recent past. In the 2018 Horizon report the undenied progression of Artificial Intelligence and versatile learning have been highlighted as noteworthy advancement, with an opportunity to attain appropriation within 2 to 3 years. In conformity with the report, experts have envisioned AI to scale by 43% in training sector between 2018-2022, although the 2019 Horizon report Higher education edition anticipated that AI applications associated with instructional learning would sizeable become rather crucial than this.
Education, as a summary, tremendously affects our entire world and is one of the establishments of the human turn of events. Expanded utilization of artificial intelligence in education without a doubt carries immense potential for improving education and instructing, yet are these enhancements starting to make a superior network and a superior world? School education that squeezing needs to change under fast improvements in AI strives on post-mandatory educational developments, as well [13-14]. Business AI mastery through advanced degree and request is one of the significant methodologies utilized by governments to impart their various expertise holes. In the endeavor to expand their various capacities in AI and transform into pioneers in this field, a few countries are attempting to make callings in AI research and learn additionally engaging.

3. Future of AI Enable Education

The fourth industrial revolution in education has operationalized the Education 4.0 framework that validates the transformation in education through technological Innovation. As indicated by forecasts from lexalytics report, "Artificial intelligence in
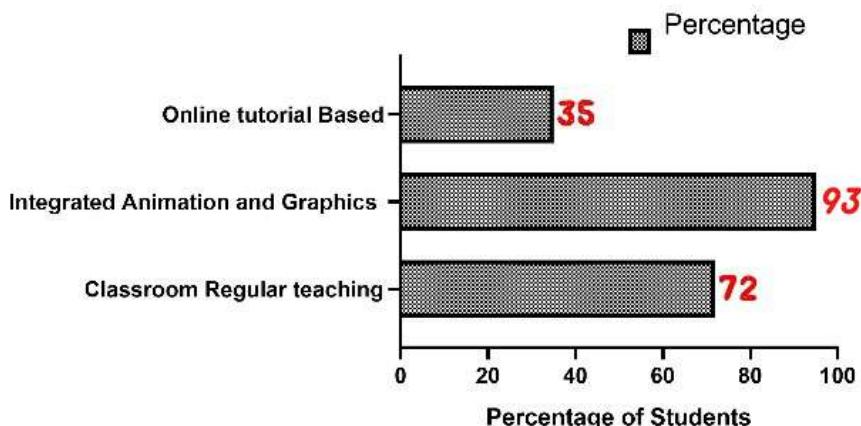
Figure 1. Student preference percent for technology-integrated learning

education will detonate throughout the following five years and is relied upon to arrive at worldwide consumption of $ 6b by 2025. [1]

By application, the market is divided into K-12, advanced education, and professional preparation. The advanced education section is required to create a more substantial income share in the market during the conjecture time frame. Virtual Reality offers consistent learning chances to the understudies. Raising support, understudy enrolment, vivid learning is a portion of the significant utilization of VR in advanced education.

Advanced education is relied upon to be the primary use of Virtual Reality in the education division. As of advanced media has made many open doors regarding the entrance to the active substance. Test learning, dominance-based learning, and customized learning are the most significant suitable advantages of VR in the education business. For Online Employee Training, there is no lack of stages and innovations. One of the approaches to connect with and train individuals could be augmented Reality (AR) and Virtual Reality (VR). Nowadays, Organizations have been utilizing VR for worker preparation. Walmart is accustomed to it to prepare workers for the fever Black Friday shopping season (they successfully made the scariest awfulness round ever – merely joking). Walmart additionally applied it to train its workers to stock retires all the more proficiently.

4. Research and Analysis

According to a recent survey by me in 2020, about AI enable learning, it was discovered that about 85% of Undergraduate and Postgraduate students in India have a preference for Animation and technology-integrated learning, as shown in Figure 1.

Student feedback at the time of Animation and VR integrated learning method for Teaching are Animation and VR is awesome in education. I could not imagine anything better than to see it presented in human life structures because the vast majority need visual articulation to get Reality. Excellent and fabulous representation to strengthen ideas. It simplifies the results to fathom. Amazing way to help instigate

learning. I would not mind to bear an extra cost for this type of learning. Memory retention of teachings are quite high due to the visual aid of 3D Animations. In addition, the pleasure of learning this way by far transcends various other schemes. This must be made available for understudies; the sooner, the better. Arrange learning materials in a context where they appear better. Visual depiction of learning materials corresponding to various parts helps to identify the proximity of the ideas and hence is a great way of learning. Coping up with and assisting thought process and fascination. With the ease of understanding I would be able to impart exceptionally to the learning process, and ameliorate the relationship with complex and difficult themes.

The experience with VR technology was exceptional. Retention of data has unquestionably become simpler. Learning with the visual aid of models instead of the course book is much convenient. The kind of learning if made accessibility to people groups with learning abilities would definitely enhance learning.
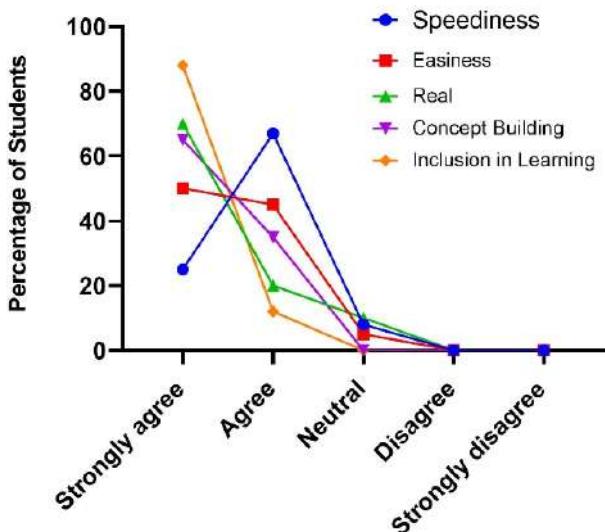
Figure 2. AR-VR learning experience

See Figure 2, Concerning VR learning experience, learning experience speed, materials were connected emphatically to this present Reality. The statistical analysis has unveiled the overflowing responses from students who strongly concur to the factors of Perception Enjoyment, Easiness, Real Concept Building Inclusion in learning.

The survey was done with google form with 521 participant pan India, and the participant responds in India AR, VR and MR are the fastest growing technology specifically in games, but in education, the use of this AI enables technology is very less.
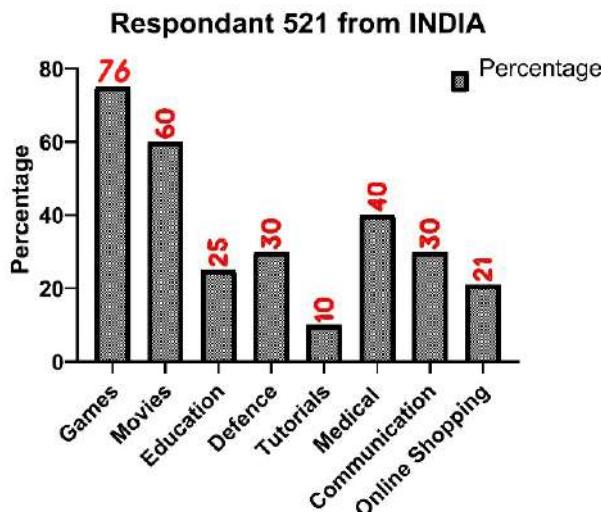
Figure 3. VR, AR, MR impact in multiple sectors

5. Conclusion and Future discussion

The outcome of the analysis and study clearly reveals the exceeding interest in immersive VR technologies for educational purposes. This has been indicated and verified by varieties of research domain that have implemented this technology in Teaching. The use of VR technology for higher education has been acknowledge by majority authors to be a promising learning tool. However, what is still questionable is the maturity of its use in Higher education. There is no denial of the fact that in some domains like engineering and computer science, have regularly been using specifically designed VR applications for delivering individual skills, particularly the ones that focus upon analytical, procedural and empirical knowledge. However, in several mother domains, the VR technology is still naively experimental, and is not systematically employed to best practices.

References

[1] https://www.lexalytics.com/lexablog/ai-in-education-present-future-ethics
[2] Barnes, N.; Fives, H.; Dacey, C.U.S. teachers' conceptions of the purposes of assessment. Teach. Teach. Educ. 2017, 65, 107–116.
[3] Darling-Hammond, L.; Bransford, J.D. Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do; Jossey-Bass: San Francisco, CA, USA, 2005.
[4] https://en.unesco.org/news/how-can-artificial-intelligence-enhance-education
[5] Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.
[6] Kesavan, S., Kumar, E. S., Kumar, A., &Vengatesan, K. (2019). An investigation on adaptive HTTP media streaming Quality-of-Experience (QoE) and agility using cloud media services. International Journal of Computers and Applications. https://doi.org/10.1080/1206212X.2019.1575034

[7]   Kumar, A., Vengatesan, K., Rajesh, M., & Singhal, A. (2019). Teaching literacy through animation & multimedia. International Journal of Innovative Technology and Exploring Engineering, (5), 73–76.

[8]   Gurjot Singh, Stephen R. Ellis, and J. Edward Swan II. The Effect of Focal Distance, Age, and Brightness on Near-Field Augmented Reality Depth Matching. IEEE Transactions on Visualization and Computer Graphics, 26(2):1385–1398, Sep 2018. DOI: 10.1109/TVCG.2018.2869729, preprint: arXiv:1712.00088v1.

[9]   A. Kumar, K. Vengatesan, A. Singhal and D. K. Verma, "3D Lighting Courseware development for 3D Motion Picture Science," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), 2018, pp. 2321-2323, doi: 10.1109/ICRIEECE44171.2018.9009258.

[10]   A. Kumar, A. Singhal, K. Vengatesan and D. K. Verma, "Study and Research of 3D Animation Courseware Development," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), 2018, pp. 2514-2516, doi: 10.1109/ICRIEECE44171.2018.9008670.

[11]   Nate Phillips, Kristen Massey, Mohammed Safayet Arefin, and J. Edward Swan II. Design, Assembly, Calibration, and Measurement of an Augmented Reality Haploscope. In Proceedings of PERCAR: The Fifth IEEE VR Workshop on Perceptual and Cognitive Issues in AR, 2019 IEEE Conference on Virtual Reality and 3D User Interfaces, pp. 1770–1774, Mar 2019. DOI: 10.1109/VR.2019.8798335, preprint: arXiv:1908.08532.

[12]   Kumar, A., Singhal, A., &Sheetlani, J. (2019). 3D Stereoscopy Video Production : Z-Depth Extraction and Optimized Rendering using Foundry Nuke. (1), 6658–6664. https://doi.org/10.35940/ijeat.A1926.109119

[13]   J. Adam Jones, J. Edward Swan II, Gurjot Singh, Josh Franck, and Stephen R. Ellis. The Effects of Continued Exposure to Medium Field Augmented and Virtual Reality on the Perception of Egocentric Depth. In Poster Compendium, Proceedings of ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (APGV 2009), pp. 138, September 2009. DOI: 10.1145/1620993.1621032.

[14]   Joseph L. Gabbard, Divya Gupta Mehra, and J. Edward Swan II. Effects of AR Display Context Switching and Focal Distance Switching on Human Performance. IEEE Transactions on Visualization and Computer Graphics, 25(6):2228–2241, May 2018. DOI: 10.1109/TVCG.2018.2832633.

Deep Psychological Learning of Japanese Anime Graphics in India Using Cognitive Computation

Pranjal Singh [a,1], Gondi Surender Dhanunjay [a], Thakur Santosh [b], Bhavana S [a],
K. Vengatesan [c] & Abhishek Kumar [a]

[a]*School of Computer Science and IT, JAIN (Deemed to be University), Bangalore, India*
[b]*Department of Computer Science & Engineering, Koneru Laksmaiah Education Foundation, Hyderabad*, India
[c]*Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon*, India

Abstract. This paper analyzes the Japanese Animation (Anime), their art style, how it is created, about popular Anime series and movies, their growth, and adaptation in India and mainly about the growing Anime fans. Majority of Anime series and movies still use two-dimensional (2D) animation style even though there are constant technological advancements in the field of media and entertainment. Though there are setbacks in Anime, we can observe that the growth in Anime viewership is constantly rising. The animation pipeline system that is involved in the making of an Anime will also be explained in detail for a better understanding of the animation process. A research is conducted through a questionnaire form to collect the required information for the study. The data collected is examined methodically and reported. The respondents are the Anime fans, Anime viewers, Anime influencers of all age groups. The survey is mainly to understand why they prefer watching Anime, how often they watch anime? what do they like about it? how were thy influenced to watch Anime? and if Anime fans influenced other individuals to watch Anime and how many people have, they influenced to watch Anime.

Keywords. Animation, Cognitive Computing, psychological Learning, Computer Graphics

1. Introduction

Anime in itself has a wide reach of viewers. There are a lot of other animated films or series that use 3D animation and is developing rapidly to make fantasy visuals look much more striking and/or realistic. However, it is understood from the preliminary survey that most Anime have not reached its next stage in technology implementation, but the number of viewers and ratings are constantly growing! People have tried to capture a sense of motion in art by placing two or more images beside each other like in the cave paintings and drawings. The animation is like a smooth sequence of a set of drawn artworks which brings it to life as a single animation.

---

[1] Abhishek Kumar, School of Computer Science and IT, JAIN (Deemed to be University), Bangalore
Email: abhishek.maacindia@gmail.com.

When fast moving images overlap each other, it forms a deception of the images to be moving. The field of animation is now appreciated by a set of people and it has grown into industry collaborating with allied media industries, and different styles of animation are adopted for its creation.

The animation style that originated in Japan is popularly known as Anime. It is a form of entertainment which is widely accepted by viewers around the world. This study is focused on Japanese Animation (Anime) and the growing popularity of Anime and its rapid growth in the field of animation. Millennials and Generation Z grew up watching Anime like Pokémon, Beyblade, Dragan Ball Z, Cardcaptor Sakura, and many more on TV without even realizing they were Japanese. Anime which wasn't initially created for kids but aimed for a wider age group of audience. Their art style, storyline and the dramatic introduction of characters challenged the animation markets and expanded the audience to various age groups; animation is highly appreciated even after continuing to use two-dimensional style.

This study is to understanding the growing popularity of Anime in the Indian context and how much people are influenced by the anime research was conducted to understand the popularity of Anime and Anime study are sparse in India.

2. Research Methodology

The animation style that originated in Japan and is usually associated with Japan is known as Anime. In Japanese, any form of animation from around the world is called as 'anime.' The term 'animation' is written in Japanese as 'アニメーション' (animēshon) and in short as 'アニメ' (Anime). Starting from the 20th century when it released the earliest commercial Anime, the Japanese animation production has continued to have a steady growth. Anime is distributed widely over the internet and television broadcasts. It targets a diverse, widespread audience and also niche audiences through its various genres.

Animation in Anime: The animation is a fundamental part of an Anime. There are different types of animations like 2D animation, 3D animation, and stop motion animation. 2D animation has other techniques like hand-drawn (Traditional Animation), Rotoscope, and the vector-based 2D animation where it adopts the methods of traditional animation but is computer generated. Anime is mostly created using Traditional animation techniques. Traditional animation is where each frame is hand-drawn by artists, and then the images are displayed sequentially to form animation, this is like a flipbook. Anime like Naruto and Dragon Ball Z have a hand-drawn background and sceneries. However, nowadays the art is done in the digital form, i.e., through the computer, as the systems have all the equipment to make it easier. But in certain circumstances, hand-drawn art is still preferred. Anime is improving its art style and animation over time, making it much better and more appealing to the audience. Most of the Anime however, utilizes limited animation, where pieces from each scene are used in the next scene if they are the same. Rather than drawing an entirely new scene every time, this process allows the animator to draw only the new details in the scene. This style of animation has its benefits; it takes lesser time to complete and costs less than other animation techniques resulting in an appealing anime that is created faster and with a lesser expense.

Art style: Japanese animation is known for their unique art style; each Anime has its distinctive art style which is mostly unnoticeable. Creative artists draw artwork differently from each other. The art style in an anime doesn't always have to be merely appealing for it to be great, in some cases, the art style represents the Anime. For example, anime like Ergo Proxy and Psycho-pass has a dark art, which is appropriate for the dark series. Styles can change drastically in just a few years, triggered by something as small as a single movie or series. There are clear patterns that emerge when you look at the medium in chronological order. Watching how these trends repeat and develop is something fascinating. A character design or animation cut will without a doubt have its roots in from 30 or 40 years ago. It's both essential and interesting to see how everything links together in one giant web of influence. The 1960s was very much a stylistic kickstart for Anime. They saw the birth of the Anime TV series; multiple long-running series started to appear every year, which gave opportunities for directors and animators to express their own styles, this began the stylistic evolution of Anime. Certain titles started to influence each other, and distinctive trends began to emerge as more shows were made.

Anime in India: Indians are by and large winding up more intrigued by Japan as its culture impacts India. While Indians have not traditionally been affected by any East Asian culture, there are indications of developing Japanese cultural impact in India. Generally, the East Asian cultural impact has been low in South Asia. South Asia's external cultural implications in the historical setting have been to a great extent from the Middle East and the West. The average spread of Buddhism from South Asia to East Asia was to a great extent one-way and was accomplished through delegates in Central and Southeast Asia; there is no confirmation of Confucian or Taoist impact in India. Because of an assortment of chronicled and social reasons, next to no East Asian impact saturated South Asia until the nineteenth century, however, there was an exchange. Until this time, nations, for example, Korea and Japan scarcely enlisted on the Indian consciousness, if by any means.

Japanese culture has affected India more than any other East Asian culture. The explanations behind this are intricate. Indian scholarly people had appreciated Japan since the nineteenth century when Japan effectively modernized; many considered Japan to be a good example. Driving Indian learned people, for example, the Nobel-laureate essayist Rabindranath Tagore kept up extensive contacts with Japanese intelligent people. Present day India and Japan have a minimal recorded history. Numerous in India sees Japan as an example of overcoming the difficulty of the East Asian model that they can securely gain from, dissimilar to China, which has disagreeable relations with India. As opposed to the talk of "civilizational ties" that occasionally underlies the Sino-Indian relationship, there is almost no civilizational learning or enthusiasm for either India or China concerning each other.

3. Result and Analysis

The A questionnaire was sent out in the month of November 2020 to people through social media, it was only for the Anime audience. The questionnaire was built on Google forms and sent as a link that was forwarded to many; this reduced the manual paperwork and made collecting data simpler. 155 valid responses were received,

the respondents are Anime fans or who watch Anime. These responses form the basis of the results, analysis, and findings presented in this report. The questionnaire consisted of

- The personal information of the individual.
- How long the individual watches Anime and how many have they watched?
- Their rating of Anime to understand how much they like and what they like in Anime.
- How they were influenced and if they influence to watch Anime.

The data was analyzed, and statistics of the responses were presented in the form of graphs using Google forms, these graphs are used for better understanding and interpreting of the survey. Reference to the questionnaire is used in the paper.
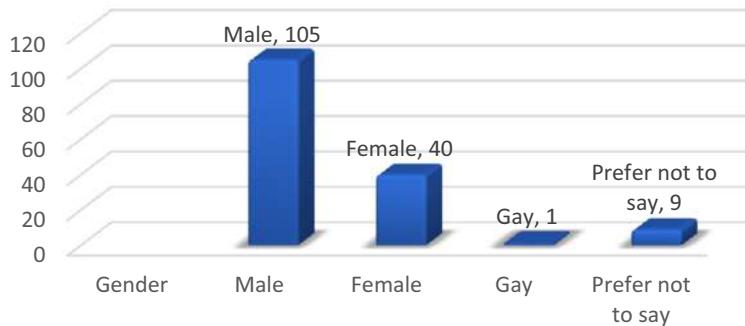
Figure 1. Viewer's Responses for Anime

The above Figure 1, shows that there are more male viewers than females who watch Anime, it is more popular among the male as there are more Shonen Anime that is popular and aired more often in India. The animated series and movies that are made in Japan have so many different genres and has so much to offer to people having different tastes, from niche community to the wide-ranging audience it has good concepts that appeal to many. Hence, people who watch Anime tend to consume it more often.
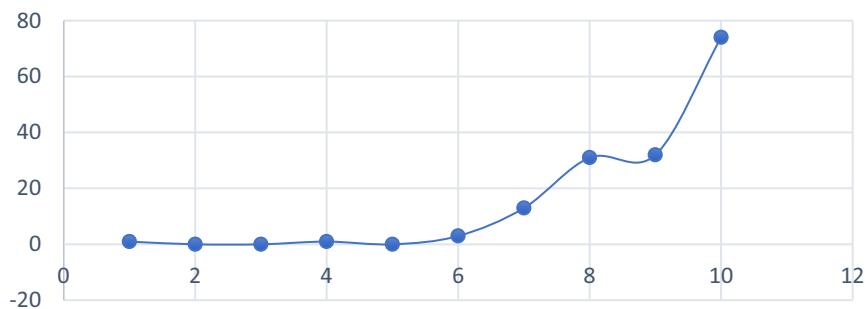
Figure 2. Rating Response of Anime

The above Figure 2 shows rating on how many survey participants enjoy watching anime. Almost 47.7% respondents have rated 10 which says that majority enjoys watching Anime, around 20.6% have rated 9 and 20 % have rated 8. The balance population in the sample has rated from 1-6 of which 8.4% have rated it 7 and lesser than 4% have rated it lesser than 7. This shows that the people who have watched Anime enjoy it a lot and are interested in it.

I watch only Anime

I watch Anime and few other Animated films/series

I watch more of other animated shows more than anime

I watch both equally
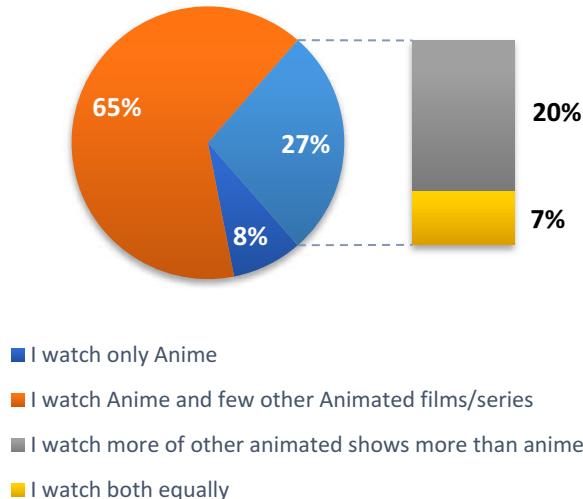
Figure 3. Respondents answering how often do you watch Anime compared to other animated films

13 respondents said that they only watch Anime and no other animated shows and 100 respondents said that they watch Anime and other animated shows. So, we can observe that people who have watched or watch Anime prefer watching more of Anime over other animated films. There are 31 people who watch more of other animated shows more than Anime and rest 11 have said that they watch both equally.

By other people/person's influence   **Television**

Advertisements   **Games**

Other sources
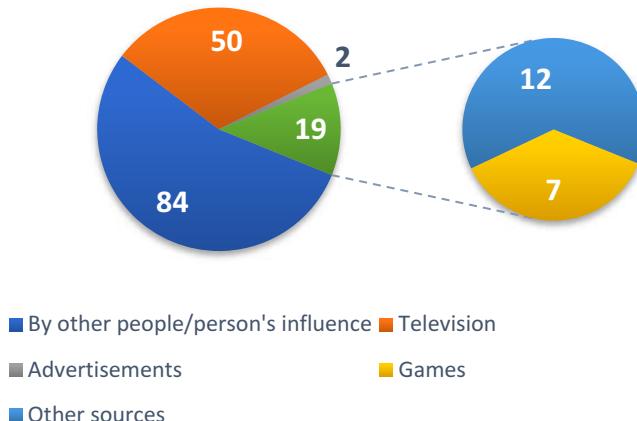
Figure 4. Ways of participants are influenced to watch anime

From the above Figure 4 we can clearly observe that Anime is being popularized mostly by other people's influence and secondly by television. There are a lot of Anime series like 'Pokemon', 'Shinchan' and 'Doraemon' that are telecasted in Indian cartoon channels. till 2017 TV Channel named ANIMAX was broadcasting in India where many viewers were influenced to watch Anime in the channel. even though ANIMAX stopped broadcasting in India Fans Now majorly watch on the web. Only a very few audiences were influenced by advertisements, games and other means as mentioned by the participants.
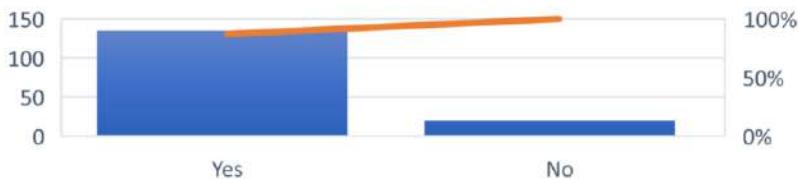
Figure 5. People influence others to watch Anime.

As seen in the above Figure 5 most people who watch Anime recommend and influence others to watch too. As we can see 87% of participants who watch Anime have agreed that they influence others to watch Anime and around 12 % said that they do not influence others.
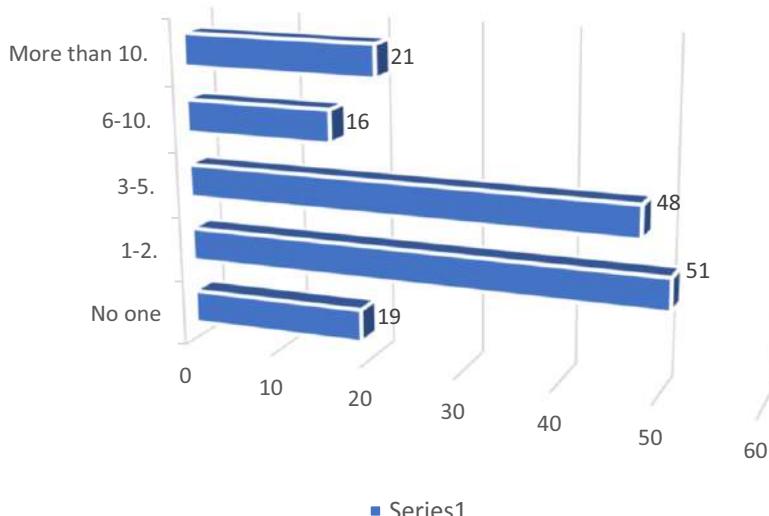
Figure 6. Peers Pressure Analysis

As seen in the above Figure 6 most of the people who watched Anime have influenced at least one other person to watch Anime. Only around 12% have not persuaded others to watch Anime. Around 30% people have influenced about 1 – 2 and the same percent have influenced about 3 – 5. The others have influenced more than 5 people to watch Anime.

4. Conclusion and Future discussion

Anime is an animation style that originated in Japan and is usually associated with it. The production of Anime started back in the 20th century and continued to grow steadily. Anime mostly uses the Traditional animation method. Initially, each frame and background of the Anime was hand-drawn, and as technology developed, Anime also evolved, now digital art is used to create beautiful artwork to bring realism in Anime. Anime has its animation techniques like the 'Obari Punch,' 'Itano Circle,' and the 'Yutapon Cubes' that are used by animators in the creation of many shows; these techniques have also evolved making them more appealing to the audience. Anime has a unique art style; unlike the other animations we see. Anime has various genres that reach a diverse range of audience. The Japanese culture has had a significant impact on the Indian culture in many ways. We have adapted and learn from them and their culture. Anime is one of the things that has widely impacted India and its people. The youth of India are engaged in activities related to Anime and conventions and actively participate in them. The Japanese government has also come forward to help spread its culture and in turn, improve the relationship between the two countries. As anime is gaining more and more popularity, it is influencing the people to get involved in it and grow as a community.

The research to understand 'The Anime Evaluation' had 155 responses, most of these responses were from Anime fans and the others who watch Anime. The responses showed that there were more male viewers than the others. Also, most of them watch Anime regularly, and only a few watches it rarely. Most of the participants have been watching Anime for more than 3 years and have watched quite a lot of Anime shows since they started. More than 88% participants have rated Anime more than 8. We can also conclude that most people like the 2-dimensional animation style that most Anime use. A lot of Anime fans influence other people wo watch Anime, and they have started watching Anime by their influence and they themselves have been influenced mostly by other people to watch Anime. According to this survey we can confidently conclude that Anime is mostly popularized by Anime fans.

References

[1] Poitras, Gilles (2000). Anime Essentials: Every Thing a Fan Needs to Know. Stone Bridge Press.
[2] Poitras, Gilles (1998). Anime Companion. Berkeley, California: Stone Bridge Press.
[3] Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470..
[4] Kumar, A., Vengatesan, K., Rajesh, M., & Singhal, A. (2019). Teaching literacy through animation & multimedia. International Journal of Innovative Technology and Exploring Engineering, (5), 73–76.
[5] A. Kumar, K. Vengatesan, A. Singhal and D. K. Verma, "3D Lighting Courseware development for 3D Motion Picture Science," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), 2018, pp. 2321-2323, doi: 10.1109/ICRIEECE44171.2018.9009258.
[6] Kumar, A. Singhal, K. Vengatesan and D. K. Verma, "Study and Research of 3D Animation Courseware Development," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), 2018, pp. 2514-2516, doi: 10.1109/ICRIEECE44171.2018.9008670.
[7] J. Adam Jones, J. Edward Swan II, Gurjot Singh, Josh Franck, and Stephen R. Ellis. The Effects of Continued Exposure to Medium Field Augmented and Virtual Reality on the Perception of Egocentric

Depth. In Poster Compendium, Proceedings of ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (APGV 2009), pp. 138, September 2009. DOI: 10.1145/ 1620993.1621032.

[8] Anna, A. (2007). The social implications of enjoyment of different types of music, movies, and television programming. Western Journal of Communication, 71(4), 259-271.

A Deep Analysis of Higher Education Cognitive and Psychological Learning Impact During Covid 19 Pandemic

K. Srinivas [a,1], Saravanan Pitchai [b], Udayabhanu N P G Raju [c], Abhishek Kumar [d],
B. Muthu Kumaran [e], K. Vengatesan [f]

[a1]*School of Computer Science and IT, JAIN (Deemed to be University), Bangalore, India*

[b]*Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, India*
[c] *Department of Computer Science and Engineering, Aditya College of Engineering, Surampalem, India*
[d]*Department of Computer Science & Engineering, Koneru Laksmaiah Education Foundation, Hyderabad, India*
[e]*Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Kattankulathur, India*
[f]*Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon, MH, India*

Abstract. Current scenario around the globe we can find that physical or face to face learning got a very big full stop for a long period of time. Virtual learning took its place, somewhat leaving behind both its positive and negative impact on the education sector. E-learning is playing a chief part in maintaining the decorum of education sector. The research and surveys found that young learners got many benefits through this type of education but also it is undeniable that it has negative aspects too, which needs to be solved. Mainly private higher education suffered less as compared to institutions in rural areas. This research proposes how to bring out the quality of output through e-learning for all the learners equally. It has become a challenge for private and government institutions to make this smart or virtual learning as the best integral part of educational system.

Keywords. Cognitive Computing, Phycological Learning, Animation, Online education, virtual learning, COVID-19

1. Introduction

Virtual or online learning can work as a salvation or lifeline for those who find obstacles in learning or getting the proper education. It is like raising a tide to lift up all the boats. The best instance of it can be seen at the hour of covid-19 phase. All the sectors of the economy suffered from the huge downfall across the globe. But one of the most important sectors, which decided the future of our youth: educational sector, started working soon after a little pause and that pause was removed by e-learning practice. Students across the globe started learning back via online mode. But the few facts cannot be ignored that in few areas how negative impact was left out.

---

[1] K.Vengatesan,Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon
Email: vengicse2005@gmail.com.

It was like a dumbstruck especially for rural areas students. Those facts were indisputable that how a large number of students suffered with their studies and even dropout condition came in front of them.

Online learning is not the newest form of learning but it came in trend from the last few decades and now became more popular in this pandemic period. Positive and negative findings were surveyed for this practice, which are discussed in the below research.

Innovation will never put back a prominent instructor but innovation in the reach of great instructor is transformational. Because of the internet facility in today's world, we can take a step forward towards the advancement of technology. This bridge leads to the path of destination which is more expedient to travel on. Education is granted as a precinct which is immune to change; simultaneously it faces a predicament of efficiency. From the various researches, I got to know that whenever people think about education, still the frequent thought is of schools and universities. The veracity is that no one actually noticed the significance of online learning, till the time everything was going smooth. It is really true that virtual learning still lacks behind than offline classroom learning. Mind-set of orthodox people still believes that proper education can be received only through offline learning that is received through the medium of schools and universities. That is so true that medium of text book to give education is one of the best ways to gain knowledge but we can't ignore the digital pattern of learning. Learning can be found outside of the school also in the ways we approach towards the theoretical knowledge same way we can approach in practical way, like through museums, art galleries, exhibitions, etc. But is this sufficient or adequate? One thing we can easily notice in today's generation that interactive learning, fun learning can create more interest among children for leaning and even their remembering capability also increases.

2. Research Methodology

The above literature suggests that this topic of e-learning was always highly debatable but now it came into more lime light because of the pandemic time due to covid-19. We got to know both the positive and negative finding of this. After knowing the importance and of this type of learning, government also took few steps forward towards the betterment and we can see that in the new education policy issued by the government.

Some consequences are not immediately recognizable, but some are noticed and some of them will be noticed after a long period of time. Like, especially for those students who mostly require or need practical knowledge. After thinking of many strategies and after applying many hit and trial methods few steps and decisions were taken to apply online mode of learning for all students across the globe, but now if we specially talk about our India, we didn't get 100% of the positive results. But few things cannot be unseen like few private institutions or universities applied this mode of learning correctly and came out with the positive results. Even many students got the advancement in their education method. To carry on and discuss the above issues on learning methods the below findings can be considered.

Positive Finding: We can observe that large numbers of students were satisfied with the e-learning technique. This approach had left the positive impact for now and even for future. From the traditional way of teaching, students found e-learning more interesting in especially in contrast to interaction with teachers and other students, availability of study materials, building up the confidence to speak, improved viewpoint of studying. Participation of the students increased in webinars and class activities, leaving the fear of speaking behind in front of all. The deeper understanding of concepts also took place because of the recorded videos.

Negative Finding: The education of a country is theoretical and is carried in the classrooms itself. The treatise or notebooks are and have always been the root for the educational purposes. With the revolutionary idea that is adopted in the education system, books are being replaced by online notes or PowerPoint presentations or video lectures. India, a land where about 65% of people lives in rural areas, which are mal constructed and lack the basic amenities like Internet services or transportation. Generally, the people living in rural areas do not have access to technologies and believes in old school method of working. The children living in such areas face great difficulty in walking hand in hand with the new educational pattern. The online education, though, has coped up with the pace of the students and has not let our precious time go in vain but it also comes with a huge price of our health problems. Problems like stress, eyes weakness, migraine etc. The continuous use of mobiles and laptops Decreases the concentration of children and directly harms their health. Another major problem came out is of dropouts, because of not proper availability of resources students took this step and it cannot be measured in exact amount.

One of the most negative finding was found and noticed (at the time of covid-19 period) is about the non-teaching staff as they can't earn though online education. This department from educational institutions suffered a lot in financial aspect. At this time government and higher education department should build up the coordination to bring out the best possible outcome, keeping in mind about teaching, non-teaching staffs and students. All will take time to adapt the changes that occurred but it has become the necessity for today and future.
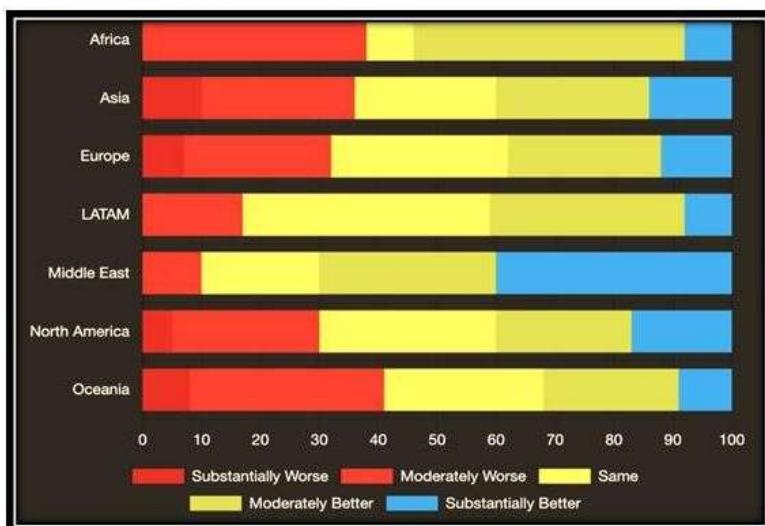
Figure 1. Graph Showing Expected Influence on Education Sector at The Time of Pandemic

This graph shows the estimated impact on the educational sector globally. It also shows how positive and negative effects were occurred. If we focus on Asia part because of our country then we will find collision between moderately worst and moderately better. We can observe both the situations around us. Once we have a look at the rural side then we will find number of difficulties facing by students and teachers and on the other if we have a look at the urban side, then we will find some institutions gave a very positive result via an e-learning approach. Even the above methodology suggests the same that how educational sector got impacted in both positive and negative manner.

3. Result and Analysis

The below findings can help us to know more about this practice. There are many reasons to have discussion over e-learning, but some of them are very important to be noticed.

Effectiveness in learning: The quality of education improves when we not only focus on theoretical knowledge but also on practical knowledge. The way of online learning has become surprisingly increased and widespread learning approach in India due to the technologies.
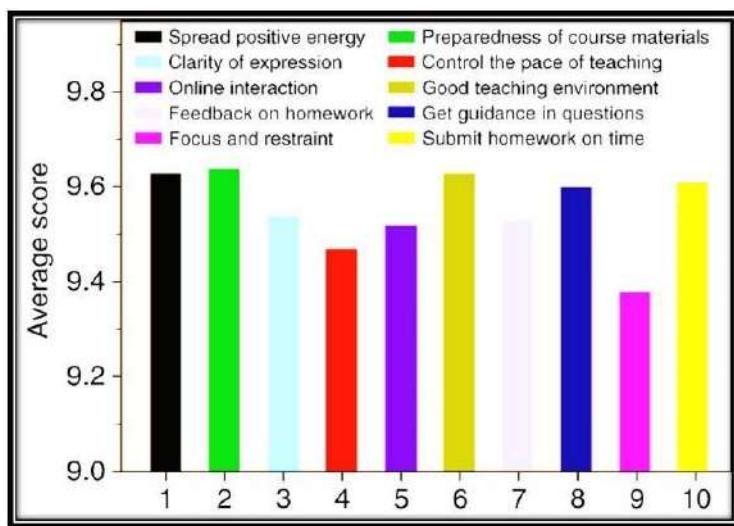
Figure 2. Coronavirus impact on e-learning

We can get the very clear view of how effective e-learning is, by seeing the growth of it at the time of pandemic (covid-19 period.) When there was no option left to maintain the proper decorum of studies then this process of learning was applied.

Networking moments: Because of e-learning students get the chance to develop their networks globally. They get the prospect to interact with different students across the nations and continents. This also helps them to become ethically strong and fit in the domain of others.

Access to Proficiency: Students get the chance to learn from more pundit (expert) teachers or faculties. As even sometimes teachers can't come to a particular institution to teach students but in online that problem is also solved. Mainly here barriers due to geographical boundaries are removed.

Study materials: In e-learning our database is always more secured as compared to offline learning because we don't need to maintain files in cupboards and desks and there is always risk of losing that data. But for online learning we can keep our data secured in E-mails or in the online classroom applications, this helps to save the time of students to find the data and keep it for long period of time (as required by them). Accessibility to the documents increase.

Reduction in student's fees: Reduction of cost takes place as students don't need to pay transport, maintenance fees, etc. This helps to nurture the economic stability of the learner's family.

Location is never hindrance: We can study and learn from anywhere we just need a stable internet connection. For learning we don't need to travel from place to place. Even we don't need to go to buy study materials.

Revise or Emend Lectures straight away: Lectures can be revised at any time as they can be recorded. According to California psychologist students lose their focus five times in an around of forty-five minutes of lectures. So, in online students get the facility to rewind the audios and videos as many times as they please.
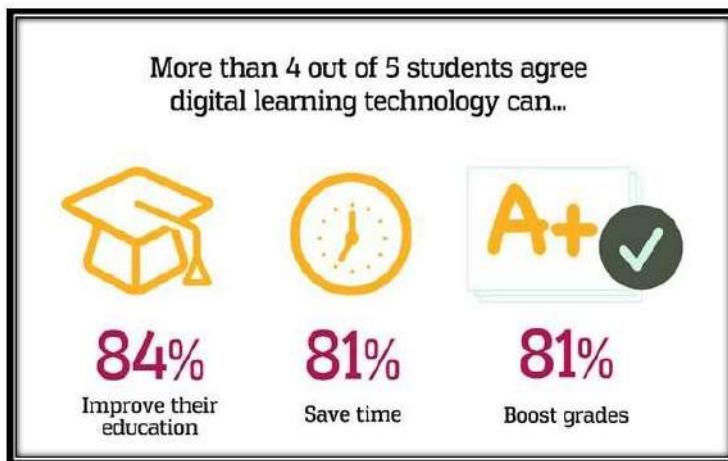
Figure 3. Net Survey analysis

This survey was mainly taken for college students but this is so true if we see around us. Mainly this image is shown as a reference over here just to show the positive impact on students because of online learning. These days only online learning is overcoming the offline method of teaching. The online method comes with lots of cons with it.

Problems in Rural Area: India is a country with about 65% of the population residing in rural areas where there is no source of electricity or Internet availability. A student of such areas faces a lot of daily problems to cope up with the studies to do online means. Many villages that do not have proper Internet or mean to access Internet hampers though education off the children. In many areas where the Internet is available the network connectivity issues become a hurdle in mode of learning. Due to

net network connectivity the voice breakdowns or the video exposed which makes it difficult for students to study and even to concentrate. Due to lack of education in such areas the technical issues that arises is difficult to overcome. The student of junior classes faces a lot of problems as online means of education is not easy to comprehend for them and each individual does not get the attention they need. For such small children if patients are also illiterate then it becomes almost negligible for a student to focus. Students below eighth standard face a lot of problems in understanding and comprehending the topics that are taught in the class.

Problems in clearing the doubts: As a student It has many other drawbacks like the assignments given in the classes are difficult to complete as the in the classes many students are not able to clear their doubts which ultimately leads to pending assignments and increases the work load and pressure on student's mind. Even for the teachers it is difficult to teach as many teachers lacks behind in use of technology and faces problem to operate gadgets. Due to technical issues the time of the class gets wasted and there is only limited time slot given for each class.

Increase in the Workload: The workload of teachers has also increased as they have to make power point presentations or prepare video lecture in advance of all the topics. The mischievous students often try to disturb the class by not muting themselves and making random noises, they often remove their classmates from the classrooms or try to create other technical problems. Students who are good in computers and other technicalities disturb the class by finding the loopholes of the application and perform inappropriate actions.

Students lacking for teacher's attention: Another drawback that occurs is that teachers cannot pay proper attention to each student and taking advantage of this situation students leave the class unattended and leave or engage themselves in other activities. The doubts clearing session is also difficult to keep with the packed schedule of the teachers and unlike offline method a student individually cannot reach to the teacher and ask for the help.

Lack of Practical Sessions: With the rising online mode of teaching the field surveys cannot be performed and also the practical work, which needs to be done in workshops or outdoor, is failed to do so. The practical based knowledge can never be attained in the online method and we can never be able to know how to mold things or how to carry other mechanical works.

Examination mode: The online examination pattern is also not honestly given. When there is no one to keep an eye on the students they cheat and conduct their examination, with this way they always remain vague on topic and it is unfair for students who give examinations honestly and score the marks on their own capabilities. The objective type question or the multiple-choice questions can be found in the internet and so the subjective or theoretical type of questions. Even if the examinations are conducted via video mode, then also the network connectivity makes it difficult to keep an eye on the students.
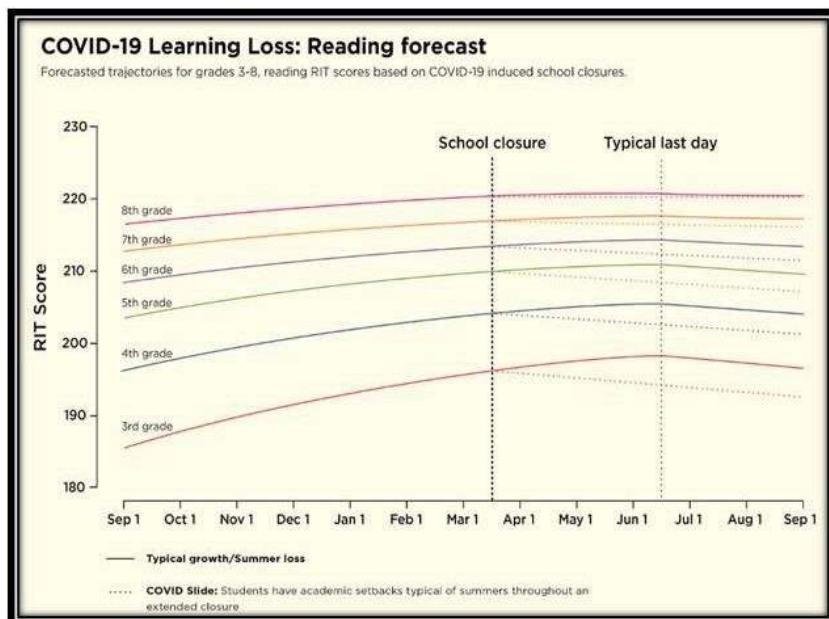
Figure 4: Impact of loss in student's education during pandemic

During the time of covid-19 time, students suffered lot because of hindrances which occurred in virtual learning. Especially for rural areas it was equal to classes going on or not.

4. Conclusion and Future discussion

Seeing the excitement and necessity for online learning in today's era it has become a vital part of educational zone. Furthermore, in offline learning, students interact with their faculties and peers to discuss the concepts or problems but they can't do same in online learning at the same level. But after seeing the both positive and negative sides of virtual learning and face to face learning, few steps are required to be taken. Uniqueness of both type of teachings should be found to bring out the finest results.

References

[1] S. Babić, S. K. Sučić and G. Sinković, "Understanding the Factors that Influence Secondary
[2] School Teachers' Intention to Use e-Learning Technologies for Teaching After the COVID19 Pandemic," 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2020, pp. 848-853.
[3] S. Dyubele, S. Soobramoney and D. Heukelman, "Using Smartphones as Constructivist Learning Tools," 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, 2020, pp. 1-8.
[4] S. Sharma and D. Verma, "Impact of COVID-19 and Nine-Minute Call on Indian Power Sector: Study of Blackout Prevention," 2020 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP), Bandung, Indonesia, 2020, pp. 301-305.

[5] M. Tropea and F. De Rango, "COVID-19 in Italy: current state, impact and ICT-based solutions," in IET Smart Cities, vol. 2, no. 2, pp. 74-81, 7 2020.

[6] Q. Pham, D. C. Nguyen, T. Huynh-The, W. Hwang and P. N. Pathirana, "Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-theArts," in IEEE Access, vol. 8, pp. 130820-130839, 2020.

[7] Kumar, A., Vengatesan, K., Rajesh, M., & Singhal, A. (2019). Teaching literacy through animation & multimedia. International Journal of Innovative Technology and Exploring Engineering, (5), 73–76.

[8] A. Kumar, K. Vengatesan, A. Singhal and D. K. Verma, "3D Lighting Courseware development for 3D Motion Picture Science," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), 2018, pp. 2321-2323, doi: 10.1109/ICRIEECE44171.2018.9009258.

[9] A. Kumar, A. Singhal, K. Vengatesan and D. K. Verma, "Study and Research of 3D Animation Courseware Development," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), 2018, pp. 2514-2516, doi: 10.1109/ICRIEECE44171.2018.9008670.

[10] J. Adam Jones, J. Edward Swan II, Gurjot Singh, Josh Franck, and Stephen R. Ellis. The Effects of Continued Exposure to Medium Field Augmented and Virtual Reality on the Perception of Egocentric Depth. In Poster Compendium, Proceedings of ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (APGV 2009), pp. 138, September 2009. DOI: 10.1145/1620993.1621032.

[11] M. N. Islam and A. K. M. N. Islam, "A Systematic Review of the Digital Interventions for Fighting COVID-19: The Bangladesh Perspective," in IEEE Access, vol. 8, pp. 114078114087, 2020.

[12] Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.

[13] D. D. N. Benty et al., "Use of Information and Communication Technology in Learning in the Covid-19 Pandemic Period to Improve Student Learning Outcomes," 2020 6th International Conference on Education and Technology (ICET), Malang, 2020, pp. 165-169.

[14] A. K. Bairagi et al., "Controlling the Outbreak of COVID-19: A Noncooperative Game Perspective," in IEEE Access, vol. 8, pp. 215570-215581, 2020.

[15] M. Pokorný, "Experience with e-learning in Teaching Combinatorics and Data Processing," 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2020, pp. 838-842.

637

Performance Measurement of Animation Design Pre-Production Artist During COVID-19 Pandemic in India

Arghya Kamal Roy [a,1], Abhishek Kumar [a], B. Muthu Kumaran [a], Sayyad Samee [b], Pranjal Singh [c], K. Vengatesan [d]

[a]*School of Computer Science and IT, JAIN (Deemed to Be University), Bangalore, India*
[b]*Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Kattankulathur, India*
[c]*Al Sadara. Al Ain, Abu Dhabi, AE, India*
[d]*Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon, MH, India*

Abstract. This research focuses on the development of the Indian animation industry. In Animation Pipeline pre-production is an important stage that determines the success of a film. Create and develop a story is the fast step and all other steps have to follow that storyline till final film realized. In this research paper present the survey base online questionnaire and the data has been collected of 300 artist who belong into Indian animation industry conducted in September 2020 by using google form. In general, Indian animation industry mostly run into production and post production (technical) base work so Indian animation industry has a smaller number of vacancies for pre-production (design and planning) artist and also having a very a smaller number of design and planning artist because of that they have highly demand. To evaluate and determine the factors that may affect the level of Indian animation pre-production industry. The study helpful to identify the animation industry current need and it focuses on planning stage to production of a movie. This research paper concludes 95% of artist working into the industry for production and post production if they are properly working into preproduction and then start into the movie work then there will be more vacancy for pre-production artist and end of the day production cost reduced up to 25 %.

Keywords. Physiological Learning, Animation, Pre-production, Indian animation industry, COVID-19, Performance Measurement

1. Introduction

In The time of 20th century media and entertainment industry take a vital role for human life style. The huge media and entertainment industry can't survive without animation, not only media and entertainment industry but also educational content, games, augmented reality, virtual reality all are very much depended on animation to full fill their goal. When it comes to animation there it's needs a proper planning for each and every frame and smallest element which is visible on screen, this proper

---

[1] Arghya Kamal Roy, School of Computer Science and IT, JAIN (Deemed to Be University), Bangalore

Email. arghyar4@gmail.com.

planning stage is call pre-production, and which department take care all this planning that department is call preproduction department, and which artist are belong in this pre-production department they are known as pre-production artist or pre-visualization artist.

In India has so many Oscar winning talent but it has really such a dearth of good quality content at home production. The Indian animation industry dominated by outsource project of television and film, which accounted that 85% of total animation project turnover in India only. Which outsource project come for Indian animation industry those are mainly for production and post production stage. So, each and every year Indian animation industry produced a lot of good quality animation project but those are not a home production content or those are not an IP (Intellectual Property) content of Indian animation industry, Indian animation industry just pushing there extremely talented artist as just their extended hand or as a technical artist to instant of money.

The Jungle Book is CGI (Computer Generated Imagery) film by Disney which awarded for best visual effect which created by only Indian animation industry so if Indian animation industry what to create a good detailed CGI (Computer Generated Imagery) movie or animation or live action instigated movie they can create in a very good manner. All the out-source project comes India after preproduction, only for production and post production so in outsource project has a smaller number of opportunities to participant for pre-production artist or pre-visualization artist.

On afraid of coronavirus Indian government announced lockdown that mean no public transport no office will run, everything will be stop to avoid spading of coronavirus, so in this time how animation industry can run smoothly or how animation industry can run as work from home. In the beginning of coronavirus crisis all animation industry was closed but animation industry use for entertainment and entertainment help human society to stay home and bring closer of a family member and also make happy and smiley and its helps to relive tension, mainly animation industry help to grow economy of country so animation studio cannot close for long time. Also, animation industry creates education content everything can be stop but education should not stop so animation industry has to run in as usual in won speed.

2. Research Methodology

In this pandemic situation of Covid-19 it's very difficult and dangerous to meet people personally for doing survey and gathering importance knowledge. So, keep in mind that covid-19 situation this research paper collects information by using virtual meeting and use of google form. This research paper is mainly depending on questioner base survey report of many artist who belong in Indian animation industry from several company and several place of country. This research paper also gone through many online portals such as website, geranial, block, article, previous research paper, Wikipedia to collect the valuable data about the current situation of Indian animation industry and its actual need for shining up in front of the world.

This research paper collects 306 artist's response, those artists represent from different variety of demographic profiles and several company's different departmental

work experience perspective. A holistic and pluralistic view has been provided by pre-production production and post production artist. This questioner consists two section, one work experience of artist and other is work experience of that studio where he or she work. This questioner may require approximate 5 to 10 minutes to complete an artist.

3. Result and Analysis

The To evaluate and determine the factors that may affect on tradition which is going on in Indian animation industry in the time of 20th century. In this time Indian animation studios are waiting for outsource project of production and post production level base work, because of that pre-production artist has very smaller number of opportunities to survive or get into in Indian animation studio.
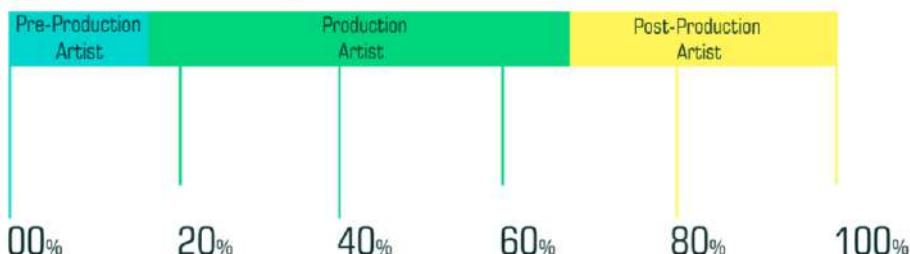
Figure 1. Different departmental number of artists

If the biggest Indian animation industry are also looking for outsource project as pre-production stage or pre-visualization stage then many artists who are interesting in creating IP (intellectual property) content they will get more opportunity to get into in Indian animation society.

The study will also helpful to identify the current situation of home production in Indian animation industry and its demands and needs. Yet you will almost not able to find a single number of original Indian home production animated movie which are highly visible or commercial hits or its really won national and international award, it's extremely stunning why it's happened, when parallelly this Indian animation society works for outsource project those are huge successfully growing all over the world.

This research paper will also try to look into how Indian animation industry, mainly pre-production industry survives in coronavirus pandemic situation. In the time of coronavirus how pre-production artist manage to do their duty, have any changes on there working process, are they getting any extra opportunities or threats because of this lockdown situation. Is this lockdown situation becoming golden time or tuff time for an artist this is main moto off this research? This questioner serves received overall 306 artists, in these 306 artists 33% are female artist and 77% are male artist that's mean in Indian animation industry mostly male artist presentence is more (from Figure 2). In these 306 artists 8.82% are pre-production artist, 67.32% are production artist, 23.85% are post-production artist that's mean in Indian animation industry has a greater number of production artist and then post-production artist then pre-production artist (from Figure 3). In that 306 artists 50% are 0 - 5 years' experience artist (fresher), 33.3% are 5 - 10 years' experience artist (mid), 16.66% are 10 - 15 years' experience

artist (senior) that means in Indian animation industry has a greater number of fresher artist and then mid artist and then senior artist (from Figure 4). In this questioner 36 studio record has been collected, in this data 25% studio are work for pre-production work and 75% studio are work for production and post-production work that mean most of the studio work for production and post-production work (from Figure 5). The main data of this questioner is in 306 artist 80.71% artist faced advantage and 19.28% faced disadvantage due to coronavirus lockdown work from home situation (from Figure 6).
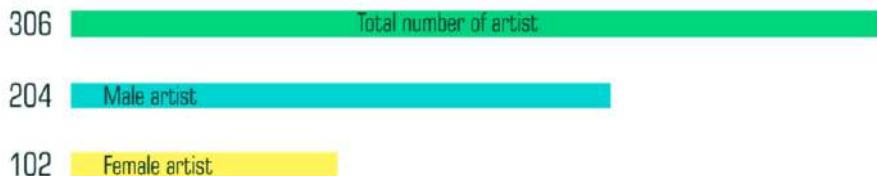
Figure 2. Male artist and female artist number of total questioners

Figure 3. Artist who response in questioner their work section
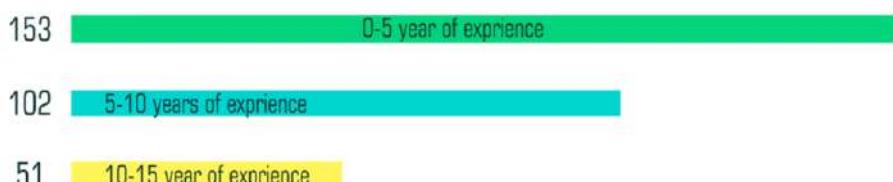
Figure 4. Experiences of artist who response in questioner
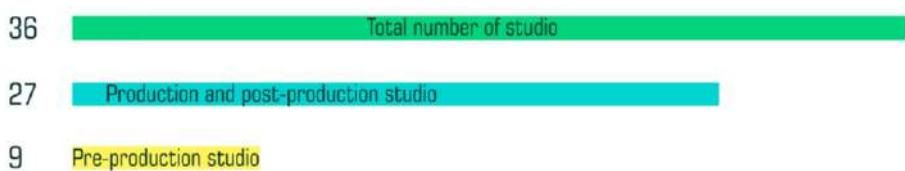
Figure 5. Artist who response in questioner their total studio and which section they work for animation film production
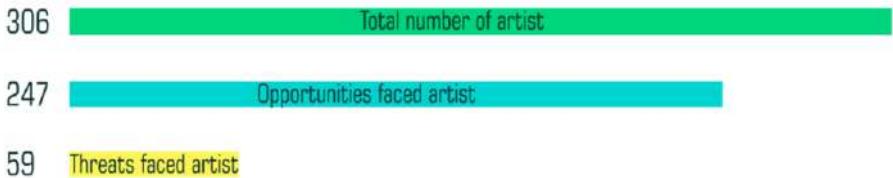
Figure 6. Different section artist questioner graph who face opportunities or threat due to 2020 (covid-19 lockdown)

In general, Indian animation industry has two face, one is production and post production work for international project those are high detail good quality work which are box office hits and also achieve a lot of international award, other one is domestic IP (intellectual property) oriented content project which are nots very much good quality work and as well as not so much box office hits.
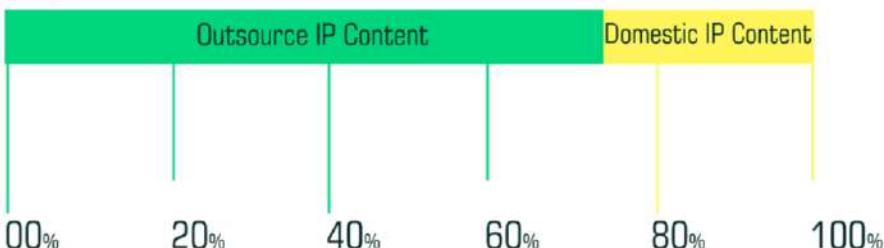
Figure 7. Indian animation studio work participant graph

Skyfall, how to train your dragon, Shrek those are international box office hit animation film those came in Indian animation industry as an outsource project but those are growing in a huge success and many international awards but Road side Romeo a Indian IP (intellectual property) content animation film this is not huge success in box office when both woks has done by the same Indian animation industry. Indian animation industry mainly focus in production and post production work of outsource project because of that last few years Indian animation industry pursuing a huge growth in production and post production section, and also mainly being dependent of outsource project industry is not focusing on create domestic IP oriented content, so if not focusing on creation of domestic IP oriented contend then don't need the pre-production team who mainly focus on planning stage and also pre visualization of a project.
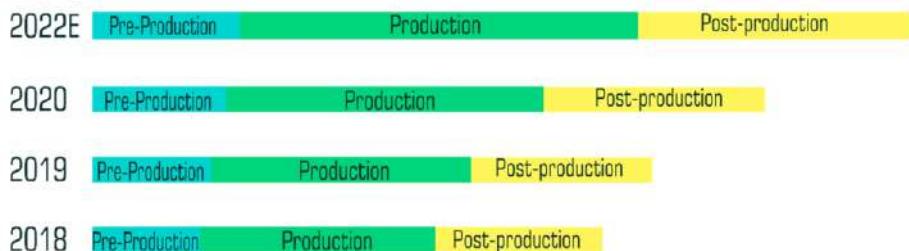
Figure 8. Overall animation studio graph since few year

The Indian animation industry mainly run under control of business people they want to made animation film with in half time and a quarter of the budged, but they expect as the same profit of Pixar or Disney film. A Disney or Pixar film take exactly three years while a Indian animation film hardly takes one year to one and half year. Within this three-year Disney or Pixar film spend almost one and half year for research, development and pre-production work and rest one and half year mainly concentrate on production, postproduction, marketing and publishing.

Figure 9. Time division of Disney, Pixar full length movie creating

In an Indian animation film pipeline, they maximum two to three month for preproduction work and less of the time concentrate for production and post production.
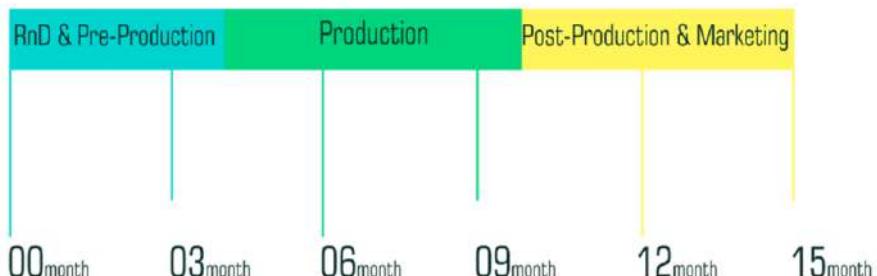
Figure 10. Time division of Indian full-length movie creating

Indian animation industry relatively needs to spend more time and more money because lack of concentration and focus on pre-production such as planning stage and pre-visualization stage. Indian animation movie also not able to shining in box office because of lacking in R & D (research and development) while Disney and Pixar studio use to do R & D a lot before a project start even, they use to reach to the child to show their design and small planning video to collect their reaction after doing that much research they get into a project finally. In outsource project in a hole day may artist have to maximum one or two second of animation some time less than that but in Indian project they have to complete six to eight second event some time client demand more than that, this timing issue became a big barrier for artist to create good quality detail output work in the case of Indian domestic production, also production and post production artist can speed up there working process if they get proper planning output as a reference.

In the fast stage of coronavirus crisis animation industry was slow down but that doesn't mean it had affected deeply, and then slowly it adapted this situation while they went to work from home condition. In this lockdown time majorly that project is going on which production are already funded. In that pandemic time studio utilize the

advantage that people don't need to share physical space to progress the project and studio also learned that how can be done much of the work even more work from home.
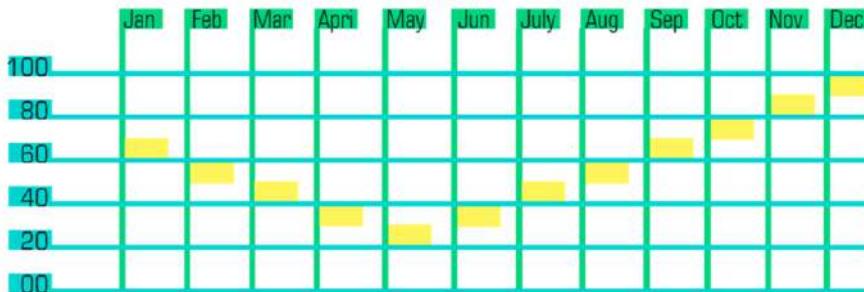
Figure 11. Overall animation studio graph at 2020 (Covid-19)

Some mid or large studio declared that 10 to 30 % of their productivity of technical and management got reduce due to lockdown situation because telework is not able to work always as effective for all type of workflow, some operation not going to control by remotely. Revision process and information flow got slow down due to increasing physical distance like group review takes longer time for a creative director to gives feedback and notes for every artist's work. In this situation to keep work flow as same speed studio had to board new member or supervisor so studio may have to spend some extra salary but parallelly some vacancy got increase.

Many actors are involving in animation film production pipeline those are also called mix media film or VFX film, in this lockdown all the shoot had to be postpone because of that so many studios suffered for this delayed of outsource captured raw footage file. Many studios directly miss opportunities or huge financials loss due to residual or postpones even cancelled the shoot of some project. In this situation launching a new project is more complicated but developing new project is less affected specially when international partners are involved. The studio who are work to develop project they are now warry about they will not get much work in future on that time there may have some certain chance to lose the job of production and preproduction artist for temporary time periods.
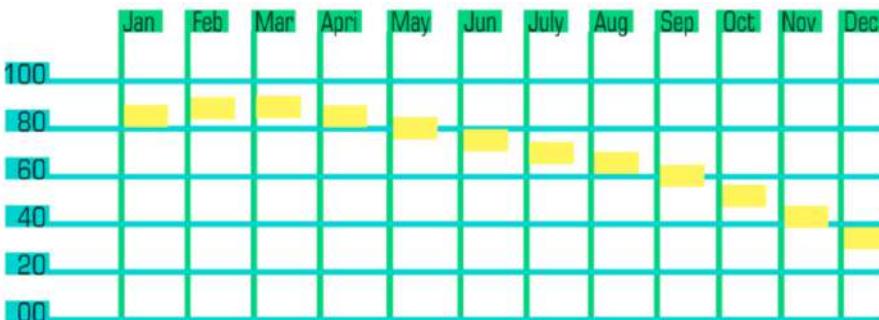
Figure 12. Production studio graph at 2020 (Covid-19)

Due to lock down many film festivals has postponed or close because of that many business and marketing opportunities generally get reduced. An animation films need

more time than a life action film so now animation developing studio are not facing any financial crises but, in the future, they have to suffer to pay all bills and expenses. But the only hope is the animation industry is growing very fast and also creating new job portal because of virtual landscape whose demand are so high now a days so thanks to their adaption ingenuity which really helps to being start layoffs.
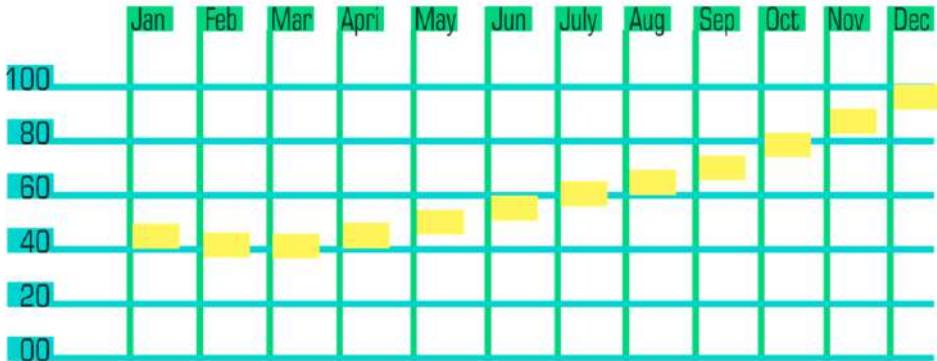
Figure 13. Developing studio graph at 2020 (Covid-19)

Unreal engine and unity are the two advance technology which help film making company to full fill their ambition of film making project due to coronavirus lockdown situation. No green screen will be required for movie production shooting par pass that will be ensure that by using of advance technology by this two-game development asset. These two new technologies have been taken new levels of creating and watching experiences by use of 3D readymade environment. This virtual landscape gaming technology already been use for two top successful move such as The Lion King and The Black Panther. By using this technique all shooting is continue due to lockdown situation because this work pipeline process can be control by remotely don't need much physically present of artist.

Figure 14. Kid content subscriber graph in OTT platform at 2020 (Covid-19)

The best thing of using virtual readymade landscape background is readymade background will be provided by game development company and the production people directly work on whole screen lighting adjustment in the shooting time only and actors need to act only while in traditional film screen has to shoot on the screen scene

and then background has to remove and place new background, tracking, light adjustment. This new technology of shoot and background placing is also help to reduce the production cost and production time while shooting is going on.

Due to lockdown for coronaviruses pandemic situation OTT (over the top) platforms subscriber number got increase. The broadcast Audience research council says that due to covid-19 consumption growth happened 39% in children content and also OTT platform Zee5 kids achieves 200% viewer ship growth other such as VOOT achieves 700% viewer ship growth. When the released content viewer ship growth is happened in lockdown time so new content has to release in market mean wile pre-production production post-production work should be on track.

4. Conclusion and Future discussion

This research paper concludes many of artist's experience and there thought and knowledge who belong in Indian animation industry in several department as story boarding, lighting and rendering, modelling and texturing, animation and character design etc. This big Indian animation society should start creating there domestic IP content as a Pixar or Disney pipeline where will be main focus on pre-production work then only production and post production work will be smoother and faster, it will also save some production cost in normal view it may increases the production cost but a good quality animation film will come which can double the profit at end of the day.

Pre-production who mainly very good at drawing they are pending their free time on creating Comic Books and Graphic Nobel digitally as well as traditional medium. By saving traveling time and energy they could also create their own IP content such as character design, illustration, concept art, background drawing, story, storyboard, asset design and those could be selling on many online portals such as shutter stacks, adobe stacks etc. As an artist their main achievement is peach of mind and happiness that can be achieve only through create their master pies art work which they create from there mind and also where will be no creative director, no team lead, no correction no, even no feedback what they want to create they can create, only this pandemic situation pre-production artists are getting their peace of mind.

In this Covid-19 pandemic situation new animation film project may not go to be lunched so pre-production artist have less work but animation pre-production artist is also connected many other industries such as gaming industry, education industry and virtual readymade landscape background creating industry, those industry has huge growth in this time. So, pre-production artist gets some new experience to work in different field project, also increase more scope for pre-production artist as well as production and post-production artist.

References

[1] S. Sharma and D. Verma, "Impact of COVID-19 and Nine-Minute Call on Indian Power Sector. Study of Blackout Prevention," 2020 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP), Bandung, Indonesia, 2020, pp. 301-305.

[2] M. Tropea and F. De Rango, "COVID-19 in Italy. current state, impact and ICT-based solutions," in IET Smart Cities, vol. 2, no. 2, pp. 74-81, 7 2020.

[3]   Q. Pham, D. C. Nguyen, T. Huynh-The, W. Hwang and P. N. Pathirana, "Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic. A Survey on the State-of-theArts," in IEEE Access, vol. 8, pp. 130820-130839, 2020.

[4]   M. N. Islam and A. K. M. N. Islam, "A Systematic Review of the Digital Interventions for Figurehting COVID-19. The Bangladesh Perspective," in IEEE Access, vol. 8, pp. 114078114087, 2020.

[5]   Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.

[6]   D. D. N. Benty et al., "Use of Information and Communication Technology in Learning in the Covid-19 Pandemic Period to Improve Student Learning Outcomes," 2020 6th International Conference on Education and Technology (ICET), Malang, 2020, pp. 165-169.

[7]   A. K. Bairagi et al., "Controlling the Outbreak of COVID-19. A Noncooperative Game Perspective," in IEEE Access, vol. 8, pp. 215570-215581, 2020.

[8]   M. Pokorný, "Experience with e-learning in Teaching Combinatorics and Data Processing," 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2020, pp. 838-842.

[9]   G. Molnár and D. Sik, "The virtual toolkit of digital instruction and its application in digital work forms," 2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Mariehamn, Finland, 2020, pp. 000597-000600.

[10] M. Pokorný, "Experience with e-learning in Teaching Combinatorics and Data Processing," 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2020, pp. 838-842.

Brain Inspired Visual Effects and Animation Psychological Computing Impact in Indian Television Advertisement Pre and Post 2000s

Gondi Surender Dhanunjay [a,1], Pranjal Singh [a], Sayyad Samee [b], K. Vengatesan [c], Abhishek Kumar [d] & Achintya Singhal [a]

[a] *School of Computer Science and IT, JAIN (Deemed to Be University), Bangalore*
[b] *Al Sadara: Al Ain, Abu Dhabi, AE*
[c] *Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon, MH*
[d] *Department of Computer Science, Banaras Hindu University, Varanasi*

Abstract. Technology in its immense boom in the last decade has made us aware of a lot of ways to increase consumer potential and engagement with different products in various spheres and aspects of production. Taking this idea forward, the main idea of this study is to identify the major visual effects facets being used and how they contributed towards consumer engagement. In this regard, a pilot study was done and then questionnaire has been prepared which was completed by 369 participants between the age group 18-60 years. Hence the main aim of this work is to use statistical data to understand how the last decade has proved beneficial for the Advertising industry through the use of visual effects Statistical analysis is used to interpret the data.

Keywords. Cognitive Science, Brain Inspired Animation, Television Advertisement, Performance Measurement

1. Introduction

The idea of graphic design is vast. Generally, people have an idea of graphic design, but people may have different ideas about the same thing which can be portrayed by different medium of art. Advertising is a form of communication used to sell products and services to the targeted audience. An animated character can become the face of a brand in India. A suitable space has been created for animation and VFX in the advertisement world because of its scope to evolve and adapt to creativity. VFX is very appealing to the younger audiences because it connects quite well with them. Due to easy availability of technology, VFX is frequently used as compared to 10 years ago. Visual effects domain is based on new technologies. new technologies offer a new way of re telling an old story. The rapid development of digitalization in the mid-90s brought about a new phenomenon by making use of new technologies in telling stories.

---

[1] K. Vengatesan, Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon, Email: vengicse2005@gmail.com.

It is evident that many advertisers have started using graphics and extensive use of visual effects in the recent past one decade.

According to the 2020 September report by KPMG on Media and Entertainment Sector it is seen that Animation, VFX and post-production segment revenue is confidently rising from financial year 2016. Overall Revenue increased to 62 Billion in FY17 from 53 Billion in FY16, it increased to 74 Billion in FY18 and 88 Billion in FY19. In the past four years Overall Revenue almost doubled from 53 Billion in Financial year 2016 to 101 Billion in Financial Year 2020.

Table 1. Animation, VFX & Post-Production Revenue in Billion (INR) from the FY 2016 to 2020.

Financial Year	FY-16	FY-17	FY-18	FY-19	FY-20
Animation, VFX & Post-Production Revenue in Billion (INR)	53	62	74	88	101

This research paper should not be mistaken as a software-specific guide, but should be viewed as a critical overview of how VFX has evolved the Advertisement industry.

2. Research Methodology

TV Audience can be classified into two categories 18-30 age group and 31-60 age group where 18-30 age group are young citizens born after 1990 and 31-60 age group could be slightly old mentality age group born before 1990. VFX is assumed as very appealing to young audiences because it has a lot of popular culture references. At present VFX in India is much easier to produce than 10 years ago with the Advancement in Computer technologies and availability of new software's which are comparatively easy to use and run on the modern-day super computers.

We have conducted survey of about 369 participants from four different cities which is geographically closely located in four different directions from central India, that is Hyderabad, Mumbai, Patna and Guwahati between the age group of 18 to 60 years. Pilot study was done by interacting with small sample of people from the four places which are selected for the study to understand if VFX in advertising plays a role in product persuasion of a particular brand? And to know what are the products which people of that place commonly use and if the advertisement is shown on TV? Majority of them responded saying advertising does play a role in brand persuasion and three products Colgate (toothpaste), Surf (Washing powder) and Lipton Tea (dip tea) were shortlisted to study the role of VFX and Animation in TV advertising.

Links of Old advertisement in which VFX was not much used and New Advertisement from the past few years which has significant use of VFX was selected for all three products Colgate, surf and Lipton tea. Survey was forwarded to Around 162 participants from the Mumbai city, 77 participants from Hyderabad, 80 participants from Patna and 50 participants from Guwahati from the age group of 18-60 years. People who use all the three products Colgate, surf and Lipton tea were given the

online google forms survey during which the participants had to watch the old video and new video embedded in the google forms and take the survey.

3. Result and Analysis

As per the Figure 1, we can see that of total 162 participants from Mumbai, 108 participants are Female and 54 participants are Male. Out of 77 participants from Hyderabad 35 participants are Female and 42 participants are Male.

PARTICIPANT FROM DIFFERENT CITY
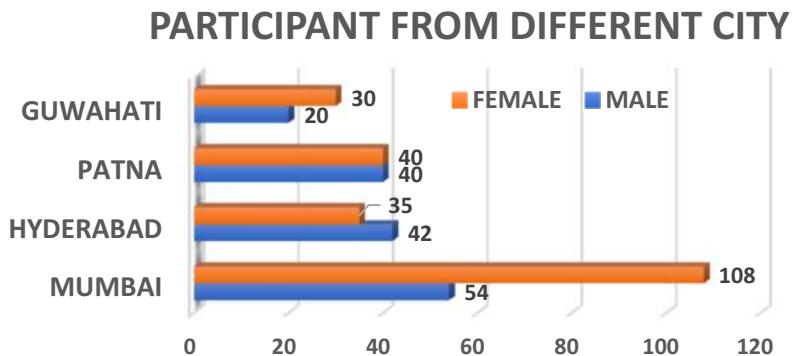
Figure 1. Graphical representation of number of participants from different place.

Out of 80 participants from Patna 40 participants are Female and 40 participants are Male. Out of 50 participants from Guwahati 30 are Female and 20 participants are Male.

COLGATE OLD   COLGATE NEW
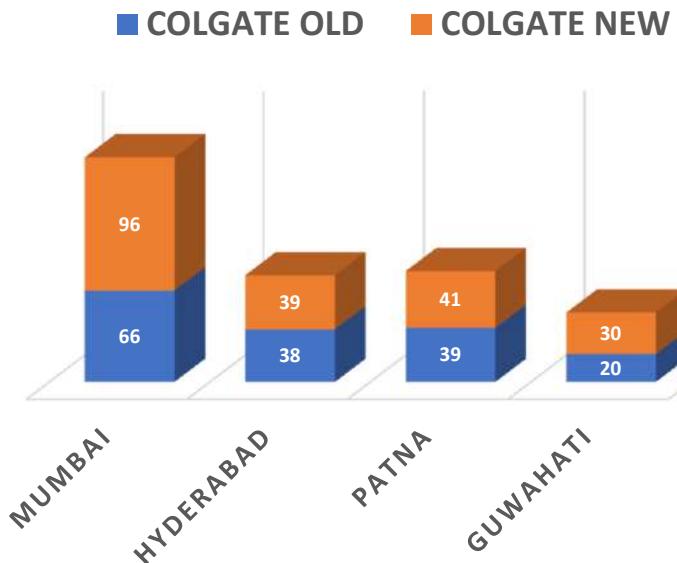
Figure 2. Graphical representation of number of participants from different cities and how many people prefer Colgate old advertisement and how many prefer Colgate New advertisement.

As per Figure 2 we can see that of total 162 participants from Mumbai, 96 participants prefer Colgate New advertisement and 66 participants prefer Colgate Old advertisement. Out of 77 participants from Hyderabad 39 participants prefer Colgate New advertisement and 38 participants prefer Colgate Old advertisement. Out of 80 participants from Patna 41 participants prefer Colgate New advertisement and 39 participants prefer Colgate Old advertisement. Out of 50 participants from Guwahati 30 participants prefer Colgate New advertisement and 20 participants prefer Colgate Old advertisement.
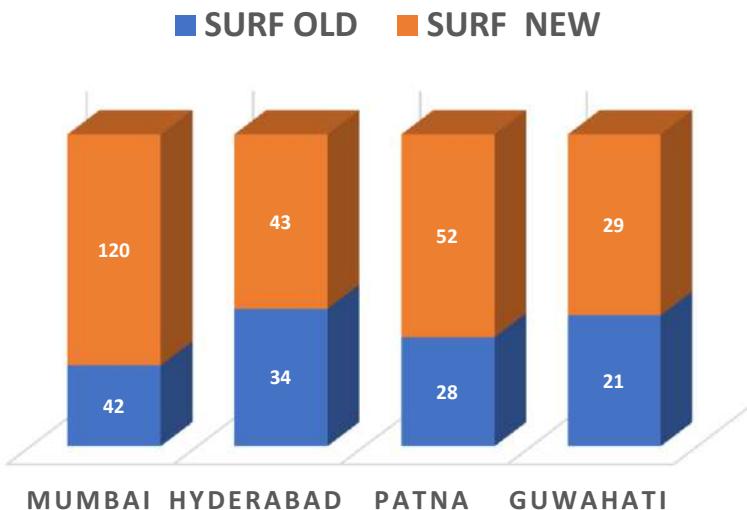
Figure 3. Graphical representation of number of participants from different cities and how many people prefer Surf old advertisement and how many prefer Surf New advertisement.

As per Figure 3, we can see that of total 162 participants from Mumbai, 120 participants prefer Surf New advertisement and 42 participants prefer Surf Old advertisement. Out of 77 participants from Hyderabad 43 participants prefer Surf advertisement and 34 participants prefer Surf Old advertisement. Out of 80 participants from Patna 52 participants prefer Surf New advertisement and 28 participants prefer Surf Old advertisement. Out of 50 participants from Guwahati 29 participants prefer Surf New advertisement and 21 participants prefer Surf Old advertisement.

Figure 4 shows total 162 participants from Mumbai, 90 participants prefer Lipton Tea New advertisement and 72 participants prefer Lipton Tea Old advertisement. Out of 77 participants from Hyderabad 51 participants prefer Lipton Tea advertisement and 26 participants prefer Lipton Tea Old advertisement.
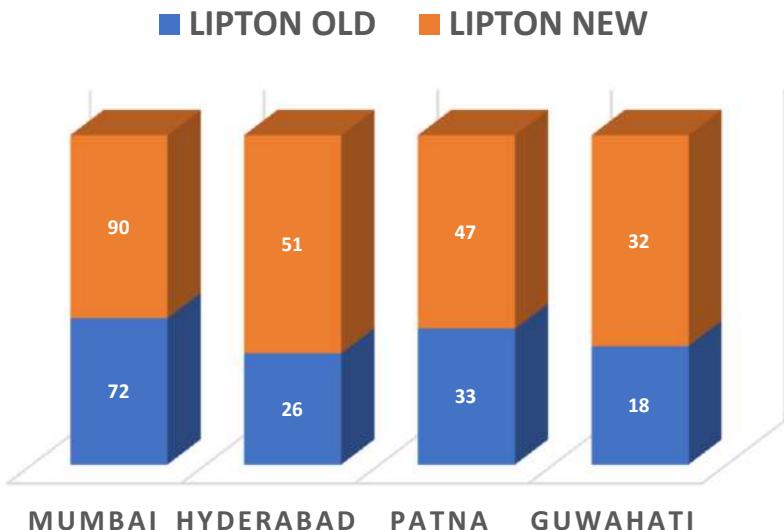
Figure 4. Graphical representation of number of participants from different cities and how many people prefer Lipton Tea old advertisement and how many prefer Lipton Tea New advertisement.
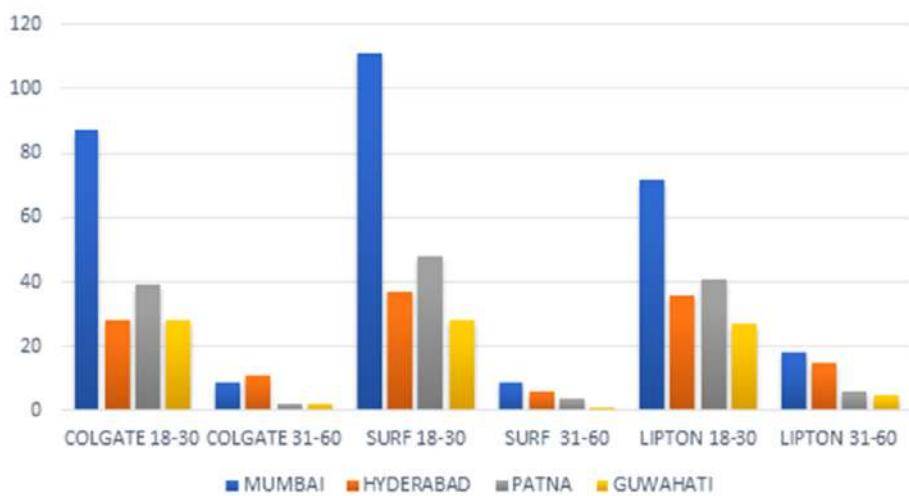
Figure 5. Age group of participants who agree that advertisement with VFX increase customer engagement towards the product

Out of 80 participants from Patna 47 participants prefer Lipton Tea New advertisement and 33 participants prefer Lipton Tea Old advertisement. Out of 50 participants from Guwahati 32 participants prefer Lipton Tea New advertisement and 18 participants prefer Lipton Tea Old advertisement. We can clearly see that age group of 18-30 are more interested in watching TV Advertising with VFX when compared to age group of 31-60.

4. Conclusion and Future discussion

The research examines the physiological impact of VFX advertisements on the viewers. The result suggests that the due to the improvement of VFX, the advertisements have now become more interactive. I divided the consumers in two groups and got the result, the younger age groups' engagement of product is higher than the older age groups. The older age groups were habituated to watch the advertisement of the last decade. In this type of advertisement, the effect of VFX was not much. But recently aired advertisement of the same products attracts consumer more because nowadays all the advertisements are mostly done with the help of Visual effects. VFX will have a great future as it associates the viewers with the product. So, they want to consume more. It will increase the demand of the product and thus the economy of the country will rise.

References

[1]   https://home.kpmg/in/en/home/insights/2020/09/media-and-entertainment-report-kpmg-india-2020-year-off-script.html
[2]   Barathi, C., Balaji C.D. and Meitei, C.I., (2014). Trends and Potential of the Indian Entertainment Industry- An In-depth Analysis, IRRJ, Vol.-II, Issue-2, pp. 286-297.
[3]   Dutta, D. (2009). Silver Screen at the doorstep of Legal Box Office, Indian Law Journal, Vol.2, No.3, pp.45-49.
[4]   Jin, C.H. (2006). Animated Commercials' Effects on Low-Effort Routes to Persuasion: Classical Conditioning Approach. University of Florida.
[5]   Nourani, S.M. (2013). Motion Graphics in Advertisement-A study. Unpublished PhD Thesis. University of Mysore.
[6]   Kumar, A. et al. "Teaching literacy through animation & multimedia". International Journal of Innovative Technology and Exploring Engineering 8. 5(2019): 73-76.
[7]   Kumar, A. et al. "3D Lighting Courseware development for 3D Motion Picture Science". 2018 International Conference on Recent Innovations in Electrical, Electronics and Communication Engineering, ICRIEECE 2018. (2018): 2621-2623.
[8]   Kumar, A. et al. "Study and Research of 3D Animation Courseware Development". 2018 International Conference on Recent Innovations in Electrical, Electronics and Communication Engineering, ICRIEECE 2018. (2018): 2514-2516.
[9]   Kumar, Dharmendra and Vats, Aman. (2016) Mapping Animation and Visual Effects in Indian Media and Entertainment Industry Opportunities and Threats, IJHAS, Vol. 1, Issue-2, pp. 19-29.
[10]   Kumar, Dharmendra and Vats, Aman. (2017) Usefullness of Animation and VFX in TV Advertisement, Communis, ISSN 2348-3601.
[11]   Potharaju, S. P., & Sreedevi, M. (2018). A novel cluster of quarter feature selection based on symmetrical uncertainty. Gazi University Journal of Science, 31(2), 456-470.

Data Confidentiality in Cloud Storage. A Survey

Shekhar S Kausalye [a,1], Dr. Sanjeev Kumar Sharma [a]

[a] *Dept. of Computer Science and Engineering, Oriental University, Indore, INDIA*

Abstract. In cloud computing security, privacy and data confidentiality plays important role when popularity in terms of cloud computing services is consider. Till now there are various schemes, protocols and architecture for cloud computing privacy and data protection are proposed which are based on data confidentiality, cryptographic solution, cipher text blocks, various transforms, symmetric encryption schemes, attribute-based encryption, trust and reputation, access control, etc., but they are scattered and lacking uniformity without proper security logic. This paper systematically reviews as well as analyze research done in this relevant area. First various shortcomings in cloud computing, architectures, framework and schemes proposed for data confidentiality will be discussed; then existing cryptographic schemes, encryption functions, linear transform, grid storage system, key exposure, secret sharing, AONT (All or Nothing Transform), dispersed storage, trust, block encryption mechanism, attribute-based encryption, access control will be discussed; thirdly propose future direction with research challenges for data confidentiality in cloud computing; finally focus is on concern data confidentiality scheme to overcome the technical deficiency and existing schemes.

Keywords. Cloud Computing, AONT, ciphertext, key exposure, dispersed storage, data confidentiality.

1. Introduction

The notion of utility computing, grid computing and distributed computing forms basis of Cloud Computing. In this concept extremely vast amount of computing, storage and networking resources along with software resources work together to form a group of virtually unlimited shared resources.

In Cloud Computing, owner of data is not aware about where their data is stored and that is why they are not having control over it which is being executed on the cloud platform. It can be said that the user does not know whether data is secured or not is there any kind of security is provided or not. To implement, use and deliver cloud computing technology owner of data must trust third party called CSP (Cloud Service Provider), now here comes the security, privacy, confidentiality, and trust problems despite of that SOC reports of CSP are reviewed by organizations in timely manner.

---

[1] Dept. of Computer Science and Engineering, Oriental University, Indore, INDIA
E- mail. shekhar_sk@hotmail.com.

There are many schemes available for privacy protection, based on encryption, block cipher, access control, secret sharing schemes, key exposure, trust but all of them are not scattered and not in symmetry [1]. It makes us to think on the recent results in various technologies for data confidentiality and privacy protection in cloud computing. As promising as it could be, cloud computing is additionally confronting numerous difficulties which, if not settled, may obstruct its quick and innovative development. Information security, it occurs in numerous different applications, is amongst these difficulties that would raise extraordinary worries from the client's side when client store their valuable or private data on cloud service providers server or called it as cloud server. These worries start from the way that cloud servers are normally controlled or handled by business suppliers or commercial cloud service provider which are probably going to be outside trust domain of the clients or users. In a few pragmatic application frameworks, information privacy is not just a security or protection issue but may be juristic concerns. For instance, in medical care application situations use and divulgence of ensured well- being data should meet the necessities of Health Insurance Portability, Responsibility Act, keeping client information secret against the cloud server is not only an alternative, however it is a necessity.

Re-appropriating information to cloud servers are beneficial because of economy, versatility, and availability, however critical specialized difficulties remain there alongside these benefits. Sensitive and delicate information stored in the cloud should be shielded from being perused by a cloud supplier that is straightforward honest yet inquisitive or curious. Step by step as the innovation expands the aggressors additionally be- come ground-breaking such assailants break the information classification by obtaining key by unapproved way. When the key is uncovered, information is lost.

Key openness is one genuine security issue for cloud storage framework or system. The review incorporates those security plans which settle the key openness issue in broadens. Existing encryption modes, methods, schemes and issues related to cloud are available which are discussed here.

2. Key Issues

Distributed computing organizations state that information is secure, however it is too soon to be totally certain about that. Cloud security concerns emerging because both client information and program are stored in cloud service provider's premises. While cost and convenience are two incredible benefits of cloud computing but has key security concerns that must be managed when thinking about shifting basic products and sensitive information to public and shared cloud conditions.

2.1 Cloud Data Security

To address drawbacks of cloud data security, the cloud service provider should create adequate controls and security policies to give the equivalent or a more noteworthy security. Security of information and confidence issue has constantly a vital and challenging issue in Cloud Computing [2]. It bases over improving security by utilizing

OTP confirmation framework, information confidentiality by employing hash calculations and mix data naturally with the most noteworthy solid or quick encryption calculation which guarantee the immediate recovery of information. To guarantee rightness of client information in cloud computing, first, client should be verified by means of various confirmation or authentication means. Authentication is one of the ways toward approving or affirming that credential given by a client are significant and valid.

2.2 Cryptography

Cryptography is the study of utilizing math to scramble and unscramble information. It is the specialty of securing data by changing the first message, called plaintext into an encoded message, called a code or cipher text. It empowers you to store sensitive data or on the other hand send it over insecure network so it cannot be perused by anybody aside from the proposed beneficiary. There are two unique sorts of cryptography which are private key cryptography and public key cryptography. In Private key cryptography a similar key is utilized for both encryption and decoding. Model for private key cryptography are AES, Blowfish, DES and Caesar Code. Out in the open key cryptography, two keys are required, one for encryption and one for decoding. Model for public key cryptography are RSA and YAK [3].

2.3 Advanced Encryption Standard

AES i.e., Advanced Encryption Standard works on block cipher having size of 128 bits for encryption and decryption as well. It is a block cipher. Input to AES is 128 bits of block of data and a key which then processes and at output ciphertext is generated. Size of key decides the various number of rounds that can be performed by AES. The larger the number of rounds more will be data secure. But one problem with a greater number of rounds to encrypt the data is it will increase the time but having one advantage that it is improved than the exhaustive key search attack.

2.4 CPA-Encryption and Secret-sharing

It is combination of the Chosen Plaintext Attack secure encryption with secret-sharing. When file is encrypted then it is shared with n-out-of-n secret-sharing scheme, then it is (n-1) CAKE secures as well as ind secure. Sharing of the encryption key then distributing it across different storage servers located at different geographical location with the ciphertext is not secure when we consider an ind-attacker. If the adversary is having access of all servers where data is stored and is able to download all ciphertext blocks, then adversary can also download all the key shares and stored along the ciphertext blocks to calculate final encryption key.

2.5 All-or-nothing Encryption

This is not encryption and does not required decryptor to have any mystery key. This shows that, All-or-nothing is not secure compared to an ind. One option is to combine

the use of All-or-nothing with standard encryption. Rivest proposes pre-measure of a message with an AONT and subsequently scramble its yield with an encryption mode. This perspective is suggested in the composition as All-or-nothing encryption and gives (n-1) CAKE security. Current AON encryption plans need in any occasion two rounds of

Table 1. Comparison of survey paper

Survey Work	Year	Technology Covered	Research Work
This Work	2020	Cloud Computing	Data Confidentiality, encryption
[4]	2013	Cloud Computing	Policy and encryption
[5]	2017	Cloud Computing	Side Channel
[7]	2015	Cloud Computing	Encryption
[14]	2019	Cloud Medical	Encryption
[15]	2018	Healthcare	Cloud Searchable Encryption
[16]	2019	E-health	Cloud Encryption

square code encryption with two keys. At any rate one round is required for the genuine All-or-nothing change that introduces the essential encryption key in pseudo-ciphertext. New encryption key is used in another round which guarantee CPA security. Regardless, two encryption changes set up a critical overhead while encoding and unscrambling gigantic reports. These game plans are either not pleasant as far as security is considered or achieve an enormous overhead when diverged from Bastion and might not be appropriate to store tremendous archives in a multi-dispersed capacity framework.

3. Related Work

Here the focus in on reviewing security and privacy in cloud computing, and rather in-volves related areas, like edge and fog computing, Blockchain and IoT. The comparison is shown in Table 1.

In academic Cloud computing security is the imp and hot topic for discussion. [4] has studied 5 security and privacy attributes which are confidentiality, integrity, avail-ability, accountability and privacy. Also, their relationship has been demonstrated, but the main missing point is lack of specific performance comparison and description. [5] stated various schemes for secret communication, some of which are side channel attack and secret channel, along with their advantages and disadvantages.

F. Cai et. al. [6] reviewed cloud computing key security and privacy challenges, they have classified the existing solutions, also compared their advantages and disadvantages. The only missing point was comparison with the other articles [7] which is based on cloud in medical field is a novel computing model for medical which focuses on challenges of electronic health reword abbreviated as HER.

R. Zang et. al. [8] reviewed four technologies. attribute-based encryption with key-word search, public key encryption with keyword search, searchable symmetric encryption and proxy re-encryption in terms of technical review only. K. Edemacu et. al. [9] discussed about various attribute-based electronic health encryption schemes. Security, efficiency and revocation ability was compared and analyzed. But privacy

protection technology discussed is relatively single.

M. Abd-el-Malek et. al. [10] proposed the lattice-based encryption algorithm for hardness of Ring Learning with Errors problem, to make it secure against the files which is stored in Cloud Storage. Lattice based encryption is cryptography techniques which is used to impede attack by both conventional and quantum computers. Using Lattice Based Cryptography technique files are not under attack when compare with another public- key techniques like RSA or Diffie Hellman. [11] are centering to endure the expanding number of fault-tolerant utilizing the deficiency, administrations without critical abatement in execution. This is accomplished by question/update protocol with the assistance of Byzantine fault-tolerant service.

Desai [12] proposed that when Blowfish algorithm is combined with AES it im- proves data security which is stored on the cloud. Symmetric cipher algorithm i.e., Blow- fish is used for the security of data and encryption. Block size of encryption is 64 bit and has a variable key length from 32 bits up to 448 bits. It has much better cost and security, time complexity compared to encryption based on Hierarchical Attribute Set. In the Advanced Encryption Standard used for encryption of electronic data, user can encrypt and decrypt the files and images with less amount of time. This ensures end to end secure communication of data without any unscrupulous data. Future scope may consist of to encrypt video and generate a stronger encryption algorithm which takes very less time and less memory.

Xiaojun Zhang et al. [13] introduced the public auditing scheme for identity based key exposure from lattices as key exposure is serious security issue in cloud auditing services. They studied few cloud auditing schemes addressing key exposure based on pairing and realized the common problem in all of those that these are not secure from quantum attacks also these relies on complex certificate management which is in Public Key Infrastructure (PKI). So, to overcome these problems they proposed a quantum resistant and independent of PKI scheme which depends on lattice assumption in cloud storage.

Chengyu Hu et al. [14] has discussed about auditing of cloud storage with key exposure resilience in continual key leakage. Forward security is provided by various re- searchers to address the key exposure problem. In this the lifetime of secret key is divided into various periods and then updating the key in these periods. This assures the validity of authenticators before the secret key is fully exposed. But the security of this protocol can be broken using side channel attack in which the secret key is not fully leaked instead it is leaked partially.

4. Research Gap

The level of the cloud storage security is enhanced by using various encryption algorithm which are combined with the integrity verification scheme. Storage selection phase is divided into three sections which are Hybrid, Private and Public. IDEA, AES, SAES and Blowfish technique are implemented. Data is encrypted using AES under Private section after which SHA-1 (Secure Has Algorithm) is applied on the encrypted data to generation 16-digit integrity verification code. This code is attached with the encrypted file or data before it has been stored in the cloud storage. Once the data is stored then SHA-1 is again applied to generate the unique token for the user having 16 digits, this token helps user

to access their data stored on the cloud. In Public section Blowfish encryption algorithm is used to encrypt the data with rest of the things are like that of private section. To adhere the data confidentiality encryption is used now a days. But main problem with the confidentiality and encryption is its secret key, which if get exposed or leaked it will create major risk to the data stored in cloud. To counter this the only way is to restrict the access to the encrypted data. If dispersed cloud storage is used the main   advantage is data is spread or stored on multiple servers located at different geographical locations with different admin domain. The novel method which was proposed is named as bastion which is claimed as efficient and fast compared to the existing schemes. The scheme uses only one bock cipher encryption round and then linear transform, which make it (n-2) cake. This ensures that plain text will not be recover on the condition that adversary is having all but two cipher-text blocks, this will be applicable when key is leaked. One of the major challenges in the data confidentiality in the cloud is cryptographic keys. Once the key is leaked or exposed, data confidentiality can be only preserved by limiting the attacker's access to the cloud data i.e., cipher text block. But all the existing work focus on the cryptographic key security and distributing the encrypted data over various servers in the network. But the attacker with proper keying material can compromise one of the servers and is able to decrypt the ciphertext block stored there. Also, current system requires pre-processing of block cipher encryption as well as another round of block cipher encryption which reduces the system performance.

5. Proposed Work

In addition to all above various schemes and technologies, the more secure and complete system which is standardized should be formed. In future, research challenge remains there like combining the various algorithms, increasing the strength of the key, increasing block cipher encryption, storage system and so on. So, considering these challenges in future a proposed system combines the bastion scheme with linear transform and more secure cipher algorithm which can overcome the key exposure problem and address the data confidentiality with less system overhead.

An All or Nothing Transform (AONT) is a capable of processing change which maps progressions of information or data to succession of blocks which are generated which has properties like. (i) given all generated blocks, change can be capably inverted, and (ii) everything except from one of the yield blocks, it's difficult to Figure any unique data. Proposed future system will ensures security of the plaintext even if attacker has the
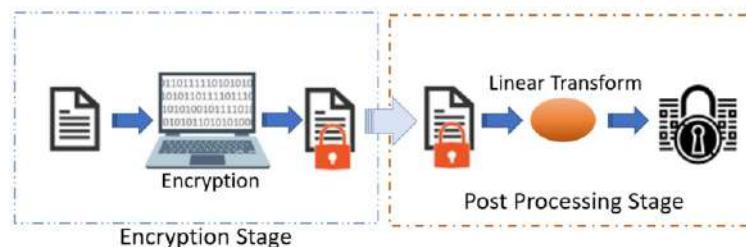
Figure 1. Proposed System uses Encryption along with Linear Transformation

encryption key and has few blocks. Current plan requires a pre-taking care of round of

block-cipher encryption for the AONT, trailed by next round of encryption. In an unforeseen manner, proposed system as shown in Figure 1., at first scrambles the data with first round of encryption, and by then applies a capable straight post processing to ciphertext. Subsequently, algorithm loosens up the thought of All-or-Nothing encryption. Algorithm uses block encryption in the CTR mode with irregular Key K. Proposed future system will ensures security of the plaintext even if attacker has the encryption key and has few blocks. Current plan requires a pre-taking care of round of block-cipher encryption for the AONT, trailed by next round of encryption.

In an unforeseen manner, proposed system as shown in Figure 1., at first scrambles the data with first round of encryption, and by then applies a capable straight post processing to ciphertext. Subsequently, algorithm loosens up the thought of All-or-Nothing encryption. Algorithm uses block encryption in the CTR mode with irregular Key K.

6. Conclusion

In the vast area of web and internet there are strong chances that your data may get exposed or leaked. So main concern in digital world is how to protect the security and confidentiality in the cloud computing. In this paper various cryptographic schemes, methods are addressed which tried to solve the problem of key leakage, data confidentiality and security. When all the schemes mentioned in this paper are analyzed and studies it has been observed that proposed system becomes more effective which can be achieved by using an efficient modified block cipher encryption followed by linear transform with very minimum system overhead when compared to all other schemes. Proposed system is suitable where data is stored on different cloud storage server which make it difficult for advisory to acquire all the blocks of data to decipher it in case of key leakage.

Acknowledgment

References

[1]   C. Basescu, C. Cachin, I. Eyal, R. Haas, and M. Vukolic, "Robust data sharing with key-value stores," in Proceedings of the 30th annual ACM SIGACT-SIGOPS symposium on Principles of distributed com- puting - PODC '11. ACM Press, 2011.

[2]   V. Padmapriya, J. Amudhavel, M. Thamizhselvi, K. Bakkiya, "A scalable service-oriented consistency model for cloud environment (SSOCM)," in Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015) - ICARCSET '15. ACM Press, 2015.

[3]   J. Amudhavel, S. Kumarakrishnan, B. Anantharaj, D. Padmashree, S. Harinee, and K. P. Kumar, "A novel bio-inspired krill herd optimization in wireless ad-hoc network (WANET) for effective routing," in Proceedings of the 2015 International Conference on Advanced Research in Computer Science

Engi- neering & Technology (ICARCSET 2015) - ICARCSET '15. ACM Press, 2015.

[4]   P. G. Shynu and K. J. Singh, "A comprehensive survey and analysis on access control schemes in cloud environment," Cybernetics and Information Technologies, vol. 16, no. 1, pp. 19–38, mar 2016.

[5]   M. S. Inamdar and A. Tekeoglu, "Security analysis of open-source network access control in virtual net- works," in 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA). IEEE, may 2018.

[6]   F. Cai, N. Zhu, J. He, P. Mu, W. Li, and Y. Yu, "Survey of access control models and technologies for cloud computing," Cluster Computing, vol. 22, no. S3, pp. 6111–6122, feb 2018.

[7]   J. Ni, K. Zhang, X. Lin, and X. Shen, "Securing fog computing for internet of things applications. Challenges and solutions," IEEE Communications Surveys & Tutorials, vol. 20, 601–628, 2018

[8]   R. Zhang, R. Xue, and L. Liu, "Searchable encryption for healthcare clouds. A survey," IEEE Transactions on Services Computing, vol. 11, no. 6, pp. 978–996, nov 2018

[9]   K. Edemacu, H. K. Park, B. Jang, and J. W. Kim, "Privacy provision in collaborative ehealth with attribute-based encryption. Survey, challenges and future directions," IEEE Access, vol. 7, 89 614–89, 2019.

[10]   M. Abd-El-Malek, G. R. Ganger, G. R. Goodson, M. K. Reiter, and J. J. Wylie, "Fault-scalable byzantine fault-tolerant services," in Proceedings of the twentieth ACM symposium on Operating systems principles - SOSP '05. ACM Press, 2005

[11]   N. Jayapandian, A. Z. Rahman, R. Sangavee, and R. Divya, "Improved cloud security trust on client-side data encryption using HASBE and blowfish," in 2016 Online International Conference on Green Engineering and Technologies (IC-GET). IEEE, nov 2016

[12]   A. Desai, "The security of all-or-nothing encryption. Protecting against exhaustive key search," in Advances in Cryptology — CRYPTO 2000. Springer Berlin Heidelberg, 2000, pp. 359–375

[13]   Xiaojun Zhang, Huaxiong Wang, Chunxiang Xu, "Identity-based key-exposure resilient cloud storage public auditing scheme from lattices", Information Sciences, Vol. 472, pp. 223-234 2018

[14]   Kumar, A., Vengatesan, K., Vincent, R., Rajesh, M., & Singhal, A. (2019). A novel Arp approach for cloud resource management. International Journal of Recent Technology and Engineering (IJRTE), 7(6).

[15]   Chengyu Hu, Yuqin Xu, Pengtao Liu, Jia Yu, Shanqing Guo, Minghao Zhao., "Enabling Cloud Storage Auditing with Key-Exposure Resilience under Continual Key-Leakage", Information Sciences, 2020

Implementation of IoT in Agriculture

B.Mohamed Arafath Rajack [a,1], N.Subramanian [a], N. Arun Pragadesh [a], R.Suvanesh [a],
S.Vignesh [a]

[a]*Department of Electrical and Electronics Engineering, SRM TRP Engineering*
College, Trichy

Abstract. In this modern world agriculture is one of the major booming sectors around the world. In India around 60 percent of GDP comes from agriculture sector alone. Also, around the world there are many technologies showing up in the field of agriculture. In this paper proposed a technology by means of which potential pest attack in the crops can be detected and the respective pesticide is also sprayed as well. Along with these there is a range of sensor employed in the field connected to the controller that will take the real time values from the field and can be displayed in the respective screen (monitor or mobile screen) by means of technology called IOT (Internet of Things). Raspberry-pi is used as the controller to perform IoT. system is linked with an application called "cain" Which allows us to display various values of sensors in the monitor or in mobile application.

Keywords. Internet of Things, Raspberry-pi, Sensors

1. Introduction

Agriculture is one of the most important and booming sector around the world lots of new and advanced technology is been introduced in this field. It is quite vulnerable field too because due to a single sign of disease can destroy the entire crops in matter of days before the farmer can do anything. So, it becomes really important to identify the kind of pest attack and take action immediately. Here, image processing technology is used to detect the pest in the plants [1]. This is achieved by preprocessing the image that are already stored in the storage system with the real time image captured by the camera that are installed in the fields over the crop .Here, image processing technology is used to detect the pest in the plants[2]. This is achieved by preprocessing the image that are already stored in the storage system with the real time image captured by the camera that are installed in the fields over the crop.

2. Implementation of IoT

The Internet of Things (IoT) refers to a system of interrelated, internet-connected objects that are able to collect and transfer data over a wireless network without human intervention. The personal or business possibilities are endless [3]. The interconnection of these multiple embedded devices will be resulting in automation in nearly all fields and also enabling advanced applications.

---
[1] B.Mohamed Arafath Rajack, Assistant Professor, Department of Electrical and Electronics Engineering, SRM TRP Engineering College, Email: mohamedarafathrajack.b@trp.srmtrichy.edu.in.

This is resulting in improved accuracy, efficiency, and economic benefit with reduced human intervention [4]. The IoT system is shown in Figure 1.
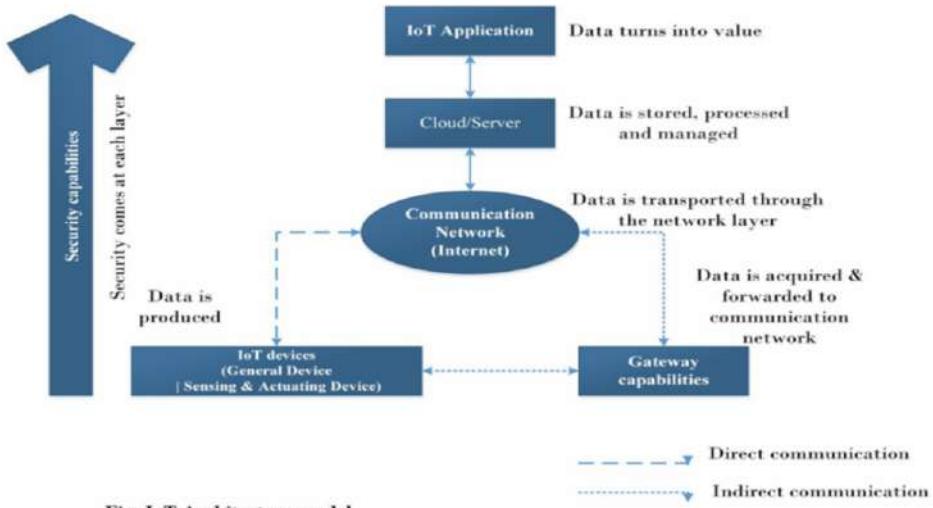
Figure 1. Convolutional Neural Network (CNN)

3. Convolution Neural Network (CNN)

This is the main part of the system in which image processing is done. CNN is a Deep Learning algorithm method which can intake a image from a user and assign various weights to it by converting it into the array to various aspects and object so that the algorithm can differentiate one image from the other [5]. It is required to pre-process Conv-Net is much lower in comparison to other algorithm's classification. This algorithm is primary need for the applications involving image processing and we have also used the same in it [6-8].

4. Block Diagram

As shown in Figure 2, the controller is the central part of the system and there are various sensors like PH sensor , Soil moisture sensor , humidity sensor and all other kind of sensors as well which need to be deployed We can see a SD card module at the bottom of the micro controller . This is been used as the controller can't hold all the programs. So for some additional memory the SD card is been used to store the programs
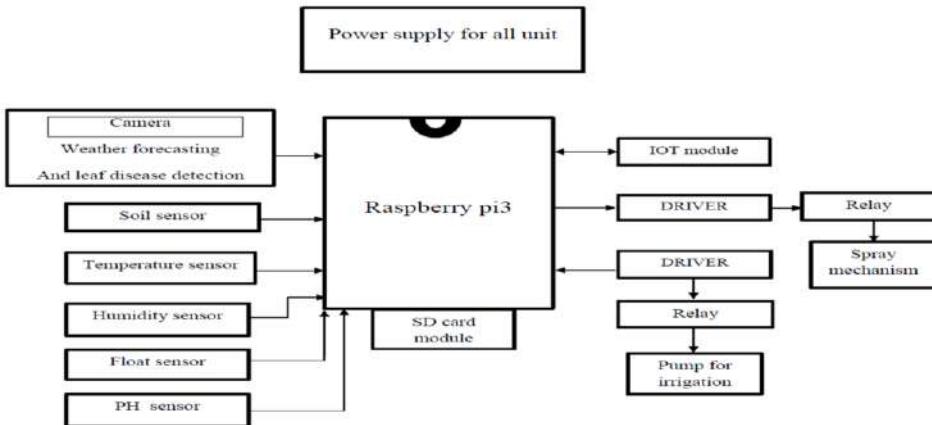
Figure 2. Hardware Block Diagram of Convolutional Neural Network (CNN)

5. Conclusion

A complete automatic crop disease detection system is proposed. The hardware components used to build the system with all their specification is discussed in the above section. Also, the software requirements for the system are also discussed in the above section. A prototype was designed with various type of sensors that will collect values from the soil sample collected from the fields and the result is also tested as well. Then the diseased plant is brought and tested with the algorithm we have in the system. This system can be further deployed in the practical field in the future and we can make it as a standalone system by installing a suitable solar panel.

References

[1]. S. D. Khirade and A. B. Patil, "Plant Disease Detection Using Image Processing," *2015 International Conference on Computing Communication Control and Automation*, 2015, pp. 768-771, doi: 10.1109/ ICCUBEA.2015.153.

[2]. Shandilya, U., & Khanduja, V. (2020). Intelligent farming system with weather forecast support and crop prediction. In *Proceedings of the 2020 International Conference on Computing, Communication and Security, ICCCS 2020*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICCCS49678.2020.9277437

[3]. Dahane, A., Benameur, R., Kechar, B., & Benyamina, A. (2020). An IoT based smart farming system using machine learning. In *2020 International Symposium on Networks, Computers and Communications, ISNCC 2020*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ISNCC49221.2020.9297341

[4]. Marcu, I. M., Suciu, G., Balaceanu, C. M., & Banaru, A. (2019). IoT based System for Smart Agriculture. In *Proceedings of the 11th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2019*.

Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ECAI46879.2019.9041952

[5]. Singh, R., Srivastava, S., & Mishra, R. (2020). AI and IoT Based Monitoring System for Increasing the Yield in Crop Production. In *International Conference on Electrical and Electronics Engineering, ICE3 2020* (pp. 301–305). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICE348803.2020.9122894

[6]. Ayaz, M., Ammad-Uddin, M., Sharif, Z., Mansour, A., & Aggoune, E. H. M. (2019). Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk. *IEEE Access*, *7*, 129551–129583. https://doi.org/10.1109/ACCESS.2019.2932609

[7]. Ganesh Babu R., Elangovan K., Maurya S., Karthika P. (2021) Multimedia Security and Privacy on Real-Time Behavioral Monitoring in Machine Learning IoT Application Using Big Data Analytics. In: Kumar R., Sharma R., Pattnaik P.K. (eds) Multimedia Technologies in the Internet of Things Environment. Studies in Big Data, vol 79. Springer, Singapore. https://doi.org/10.1007/978-981-15-7965-3_9

[8]. Vengatesan, K., Kumar, A., Parthibhan, M., Singhal, A., & Rajesh, R. (2018, December). Analysis of Mirai Botnet Malware Issues and Its Prediction Methods in Internet of Things. In International conference on Computer Networks, Big data and IoT (pp. 120-126).

[9]. Markkandan, S., Narayanan, L., Robert Theivadas, J., & Suresh, P. (2021). Edge based interpolation with refinement algorithm using EDGE strength filter for digital camera images. Turkish Journal of Physiotherapy and Rehabilitation, 32(2), 981–993

Threat Model for Secure Health Care Data Using EMR, EHR and Health Monitoring Devices

Ms. RA. Kamalaeswari [a,1], Dr. V. Ceronmani Sharmila [b]

[a]*PG Student, Department of Information Technology, Hindustan Institute of Technology and Science, Chennai, TN, India*
[b]*Professor and Head, Department of Information Technology, Chennai, TN, India*

Abstract. The main aim of this project is to propose a threat modeling framework that promotes the security of health care services. The threat model is used to analyze the cyber threats that makes the electronic health monitoring devices vulnerable to a cyber-attack. The model also helps in strengthening the security of the software-based web applications like EMR and EHR used in a health care organization. The information assets are identified and the threat agents are eliminated considering the software, web application and monitoring devices as attack surface. The major goal of this threat model is to analyze and establish the trust boundaries in the OpenEMR that render a secure data transmission. We use a STRIDE threat model and a DFD based approach using the OWASP threat modeling tool. The SIEM tools provide a continuous security methodology to document the process and result.

Keywords. Threat model, cyber-attack, EMR, EHR, information asset, threat agents, attack surface, OpenEMR, STRIDE, DFD, SIEM.

1. Introduction

Medicine may be defined as the most important part of the human life. Even though the field of medicine involves the interaction purely with the human body and organs, there are situation where we integrate the advancements of technology to interpret the working of human body. So, we use electronic health monitoring devices which provides the doctors with data regarding a particular organ or organs of the human body.

A Healthcare organization uses software and web application to collectively maintain the patient and hospital data. The commonly used software is EMR and EHR. The EHR run like any other software based on web application including both front-end and back-end.

---

[1]RA.Kamalaeswari, Department of Information Technology, Hindustan Institute of Technology and Science, TN, India; Email: kamaliraghu98@gmail.com.

Electronic medical records (EMRs) are an advanced form of the paper diagrams in the doctor's office. An EMR provides the hospital visit and treatment history of the patients in a single practice.[1] EMRs have benefits over paper storage. For instance, EMRs permit doctors to:

- Track information of all the patient history and conditions.
- Effectively recognize which patients are expected for screenings or tests
- Check how their patients are getting along on specific boundaries, for example, pulse readings or inoculations.

2. Working of the model

Professionals in medical field use EMRs for data storing, processing and retrieving due to their user-friendly nature and customization options. We in this project will use one of these widely used framework called as OpenEMR, which requires a web server and data base server to function in optimization. We store and manage clinic related sensitive information inside the EMRs [2]. Threat model is the security cycle by which we can recognize, sort, and investigate dangers. A threat model incorporates:

- A plan (graph) of the framework
- A procedure that incorporates a rundown of suspicions that can be checked. A portion of the mainstream systems are STRIDE, PASTA, TRIKE, and VAST.
- A rundown of threats and controls for alleviation.
- An approach to approve the model, vulnerabilities and check of accomplishment of the activities taken.

2.1. Methodology

STRIDE represents Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Escalation of Privilege. These are against security functionalities like Authentication, Integrity, Confidentiality, Non-Repudiation, Authorization and Availability of the data [8-10]. The three main goals of this threat model are to assess:

- Ability to understand the system proposed, which in our case is OpenEMR.
- Find the potential vulnerabilities.
- Prioritize these vulnerabilities and eliminate them respectively

The threat model is created by first defining a template workflow of the events in the proposed system. The threat stencils are determined and categorized. The threat is defined using properties as we identify and prioritize them. The workflow template as shown in Figure 1, is created which defines the critical information assets, data flow, user roles, user privileges, escalated privileges, attack surface and trust boundaries.

Figure 1. Open EMR login through local host and database.

2.2. Process and threat flow

The OpenEMR consists of various modules which help in executing each functionality according to the patient and clinic needs is shown in Figure 2. The addition of patients, billing, payments, electronic health reports, lab documents, procedure recordings, care coordination, patient validation, prescription of medicine and symptoms are all stored in the database. The clinical plans, rules and drug inventory are monitored using the interface. All the data specified are critical assets of the system and define the data flow within the system. The database and the user credentials are vulnerable attack surfaces.
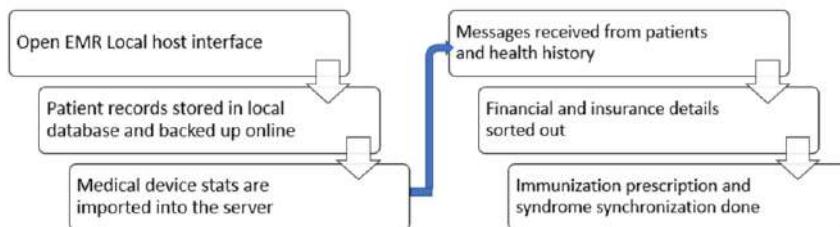
Figure 2. Open Emr process flow

    The critical information assets of the system proposed can be physical and logical. The hospital data contains most of the physical information about the patient which is highly confidential. The health data about a person can be used against them in various ways. In order to avoid all the potential threat factors from inside the organization we must identify the bad actors and the attack surface is given in Table 1. The bad actors contribute to most of the insider threats.

Table 1. Threat properties

Attack surfaces	Insider threats
Local host	Database flaws
Database	Pharmacist
Patient records	Doctor
Billing checkouts	Lab technician

Lab Documents	Patient
Electronic device	Billing accountant
Electronic Reports	External drug vendor

3. Related works

The medical care industry is highly considered by hackers more so than some other area at recent times. The internet that is being utilized in the medical services is one of the primary stores to acquire data from. Many researches have been made regarding the working and security of these EMRs. We have considered few of those studies as references in our project. The threat model is given in Figure 3.

a) The workflow of the data processing, mining inside an open source EMR is described in depth in this work. Various data algorithms are used in fetching the correct data regarding the patient medical treatment history. Conformance checking is the methodology used here. The building of user roles inside the software is determined and hence the understanding of the work done here can be used to prevent the escalation of privilege and access control issues.[3]

b) The second work showcased deploys more security in the EMR by deploying block chain technology. The authentication and integrity of the data is hence preserved. The security and transparency amongst the network nodes are also regulated for clear communications. The data stored is hashed and encrypted in order for more security. A reminder system is added in the set up to manage the time schedule of the patients.[4]

c) This thesis is part of the Arizona Center for Accelerated Biomedical Innovation's ongoing health analytics research and development programme. They use big data technology to improve the experience of both patient and doctor when it comes to EHR. The video and audio of patient visit along with the clinical measurements are recorded with the patient's consent. Even though this provides a better and clear methodology, this contains all the highly sensitive data but also most vulnerable to cyber threats.[5]

d) In this thesis the management of a secure hospital cyber space by the Dubai hospitals are analyzed and defined. The users present inside the hospital environment and their perception of the data security is questioned and researched. This was implied as a overall case study of data privacy inside the hospital and how the patient information is vulnerable to an unintentional insider attack.[6]

e) According to General data protection regulation all the highly sensitive and personal data of both patient and doctor must be secured. This work implies various cryptographic and data analysis algorithms to generate private keys and protecting the data. The cryptographic keys are also used for storage and transmission of data.[7]
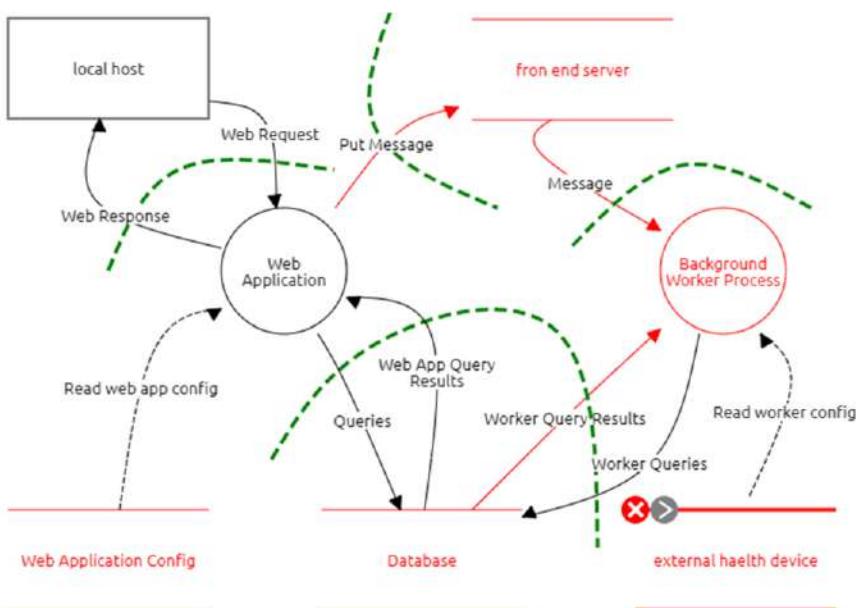
Figure 3. Threat model

4. Conclusion

The medical care industry is highly considered by hackers more so than some other area at recent times. The internet that is being utilized in the medical services is one of the primary stores to acquire data from. Many researches have been made regarding the working and security of these EMRs. We have considered few of those studies as references in our project.

References

[1] Akowuah, F., Lake, J., Yuan, X., Nuakoh, E., & Yu, H. (2015). Testing the security vulnerabilities of OpenEMR 4.1.1: A case study. *Journal of Computing Sciences in Colleges*, *30*(3), 26–35.

[2] Agbele, K. K., Oriogun, P. K., Seluwa, A. G., & Aruleba, K. D. (2016). Towards a model for enhancing ICT4 development and information security in healthcare system. In *International Symposium on Technology and Society, Proceedings* (Vol. 2016-March). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ISTAS.2015.7439404

[3] Asare, E., Wang, L., & Fang, X. (2020). Conformance Checking: Workflow of Hospitals and Workflow of Open-Source EMRs. *IEEE Access*, *8*, 139546–139566. https://doi.org/10.1109/ACCESS.2020.3012147

[4] Sheela, K., & Priya, C. (2020). Enabling the efficiency of Blockchain Technology in Tele-Healthcare with Enhanced EMR. In *2020 International Conference on Computer Science, Engineering and Applications, ICCSEA 2020*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICCSEA49143.2020.9132922

[5] Rahman, A., Mitra, A., Rahman, F., & Slepian, M. J. (2019). Smart EHR-A Big-Data Approach to Automated Collection and Processing of Multi-Modal Health Signals in a Doctor-patient Encounter. In *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019* (pp. 6198–6200). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/BigData47090.2019.9006574

[6] Bakkar, M., & Alazab, A. (2019). Information security: Definitions, threats and management in dubai hospitals context. In *Proceedings - 2019 Cybersecurity and Cyberforensics Conference, CCC 2019*

(pp. 152–159). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ CCC.2019.00010

[7]   Park, H. A. (2019). Secure electronic medical record (EMR) system. In *Proceedings - 6th Annual Conference on Computational Science and Computational Intelligence, CSCI 2019* (pp. 955–960). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/CSCI49370.2019.0018

[8]   R. G. Babu, A. Nedumaran and A. Sisay, "Machine Learning in IoT Security Performance Analysis of Outage Probability of link selection for Cognitive Networks," *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2019, pp. 15-19, https://doi.org/10.1109/I-SMAC47947.2019.9032669

[9]   Markkandan, S., Malarvizhi, C., Raja, L., Kalloor, J., Karthi, J., & Atla, R. (2021), "Highly compact sized circular microstrip patch antenna with partial ground for biomedical applications", Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2021.04.480

[10]   R. G. Babu, V. V. Nathan, J. Bino, C. Amali and S. Ganesh, "IoT Security Enhancement with Automated Identification Device using IOT SENTINEL," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 518-523, https://doi.org/10.1109/Confluence51648.2021.9377165

Smart Ambulance System with Remote Knowledge Communications Through Cloud

B. JAYASHREE [a,1], A. SHIVARANJANI [a], S. SUVETHA [a], M. JANSI RANI [b], P. SURESHA BARANI [b]

[a]UG Student, [b]Assistant professor, Department of ECE, Panimalar Institute of Technology, Chennai, India

Abstract. An ambulance is one of saving many lives by taking the people who need health emergencies. Saving the life of the person is one of the challenging and precious ones. Our key idea is to deliver a patient's health condition before the victim reaches the hospital in this project. Here we use some biomedical sensors like a heartbeat sensor, temperature sensor, and a respiratory sensor to check the patient health status. There will be a continuous update to the hospital about the patient's condition through the cloud with the help of the internet of things. The hospitals can also track the ambulance's live location through the GPS placed in the ambulance where it arrives, and they can know at what time the patient reaches the hospital. With this information, if the patient is in critical condition, the hospital staff can make all the earlier arrangements before the patient arrives at the hospital and saves their lives as soon as possible. Here we use the biometric sensor to know the patient's information by scanning the patient's fingerprint. The stored database obtains this information. In cases of accident situations, to avoid legal problems, the patient's information is sent to the cops through the GSM, and it is also intimated to the patient's relatives as soon as possible. The parameters which are measured by using biomedical sensors are viewed by doctors using the Blynk app.

Keywords. Biometric sensor, GPS, GSM, NODEMCU, Cloud server, Arduino UNO

1. Introduction

In India, an immense piece of metropolitan zone concocting brilliant Improving proficiency in the clinical advantages segment is a problematic undertaking which will require a portion of the authentic energy still, we can do it. We have made a thought out of bit-by-bit substance to make out a pointed recover vehicle with patient checking instrument concerning our endeavor [1]. We can see that dependably, and hour, somebody passes on considering not enduring proper treatment in an ideal manner. This problem can be eradicated by improving the efficiency of healthcare by transferring information from patient at rural health care unit to the doctor at other location through IoT for obtaining guidelines to undergo further treatment based on the patient data [5].

---

[1] B.Jayashree, Department of ECE, Panimalar Institute of Technology, Chennai, India; Email: jayashreebabu2000@gmail.com.

This paper notices the issue by stretching the odds of saving numerous people's existence by utilizing an arising headway in the Internet of Things, which is an accomplice for additional gadgets in essentially to the web. The idea of Internet of Things is a lot of accommodating to accomplish constant observing of sensor information [7]. IOT confirms our in general society more fruitful and helpful than we potentially envisioned whenever. Recently, clinical advantages structures have become a solid region and can be made rapidly by the clinical facility. The thriving status can be transported off to the emergency clinic. Android is an application that can be utilized by the clients by downloading android application [4]. So, we banded together with salvage vehicle but utilizing IOT to make it sharp crisis vehicle faraway data correspondence that can be gathered and give the persons flourishing status to the close by facility through the web which can view through Blynk application and the persons' information passes through SMS. GSM (SIM 800C), Arduino UNO (Atmega 328), GPS (NEO 6M) and NODE MCU (ESP8266) is used to achieve this, which accumulates essential information.

2. Related Work

A survey of this writing uncovers the progressing and expanded interest in IoT and IoT-based advances and arrangements [3]. An outline of the IoT is introduced. Despite the difficulties, like security, protection, Etc., that hinder IoT applications' advancement, the outline features the importance and advantages of the Internet of Things across different application spaces and regions [6].

Arita Baksi, Mayookh Bhattacharjee proposed a project in which each and every emergency vehicle, a GPS and GSM modem will be outfitted and at the emergency time GPS will send the fastest route from the vehicle's current location to the hospital through GSM [2].

3. Proposed System

The pictorial representation of proposed system is shown in Figure 1.

3.1. Working Principle
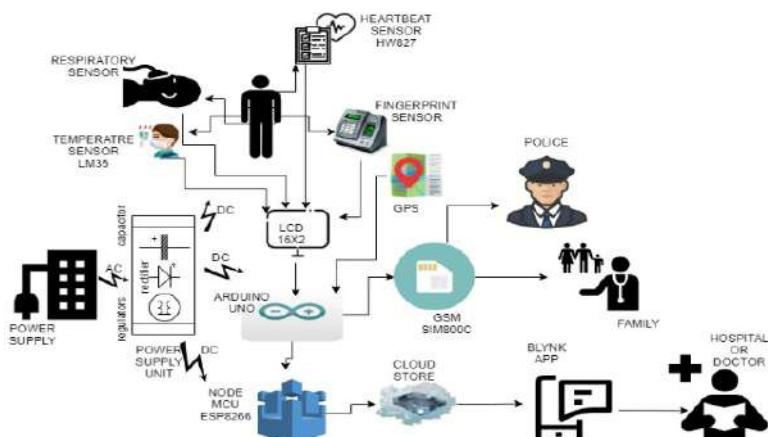
Figure 1. Pictorial representation of the project

3.2. Modules Description

- Fingerprint Authentication: Biometric sensor module used in this project is R307. This alludes to the robotized strategy for checking a match between two human fingerprint [8].
- Heartbeat Sensor: HW827 is the heartbeat sensor used in this project which is measured by just placing the tip of your index finger delicately ludicrous all over.
- Respiratory Sensor: Treatment pressures must be observed and controlled in clinical instruments respiratory device.
- Node MCU: Node MCU ESP8266 is getting extraordinarily famous and are nearly utilized in more than half of IoT-based tasks today.
- GPS: GPS provides users with accurate information about their position and velocity. GPS used in this project is Neo 6M.
- GSM: In an SMT module highlighting an industry-standard interface, the SIM800CS is a quadband GSM/GPRS module that works on frequencies GSM850MHz.
- Temperature Sensor: The temperature sensor senses the patient body temperature. The LM35 operates over a temperature range of -55°C to 150°C.
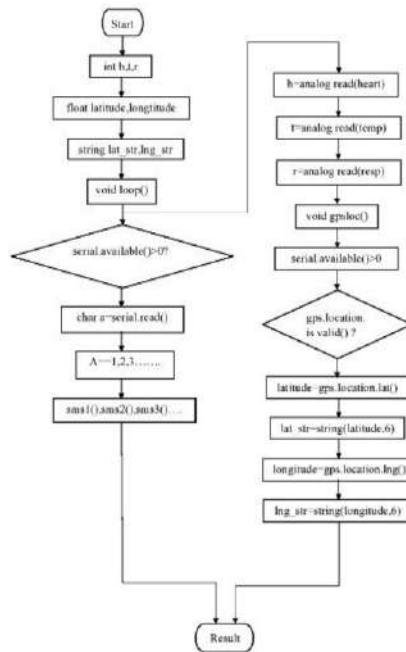
3.3. Projected Methodology

Assuming this is the case, the courses of action for the treatment utilize the sensors that distinguish the boundaries like heartbeat rate, temperature, respiration, and so forth, connected to the Arduino. These sensors measure the values and send the live deals to the IoT. Those values will be updated in the cloud and ship off to the Hospital .it is also personally viewed by the Blynk app by the doctors of the Hospital. Also, here we use GPS for live tracking of the Ambulance to monitor when the patient reaches the Hospital. Here our goal to execute an imaginative answer for this issue by building up a Person recognizable proof System utilizing biometric sensors. The result of all the sensors is displayed in LCD through the microcontroller and it sends the message to our mobile via GSM modem [6]. Ambulance offers assistance for an obscure individual who is confronting an accident or emergency and takes those individuals to the Hospital [9]. In this venture, I am adding some extra highlights inside the emergency vehicle to distinguish the individual confronting an accident or health emergency. After the recognizable proof is sent to Hospital and on the off chance that it is an accident case, the data is ship off close by the police station through the GSM module to stay away from lawful issues. It additionally diminishes the work weight to the police in distinguishing the subtleties of the individual. This framework has been created and carried out utilizing the unique biometric sensor-based embedded technology incorporated with the advancing Smart Device. It assists with offering the most astute support for the public smartly. The complete setup is projected in Figure 2.

Figure 2. Complete hardware setup

3.4. Flowchart

4. Results and Discussion

In this paper, we projected a patient checking framework inside the ambulance. We are utilizing some biomedical sensors. These sensors detect and send patients' conditions to the cloud through IoT, which can be constantly observed. Here, we associated the GPS

tracker with following the ambulance's live area. The aftereffect of these can be seen in a solitary application by the specialists by Blynk application, which is appeared in Figure 3. Here we utilize a biometric sensor to recognize the patient data. An SMS is shipped off to the police/relatives through GSM, as demonstrated in Figure 4.
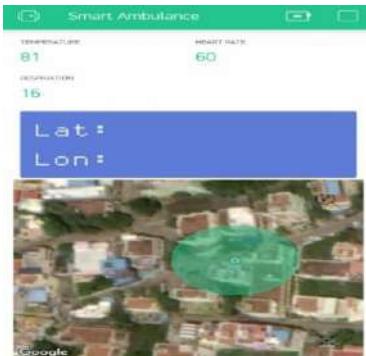
Figure 3. Result displayed in Blynk application   **Figure 4.** SMS received through GSM
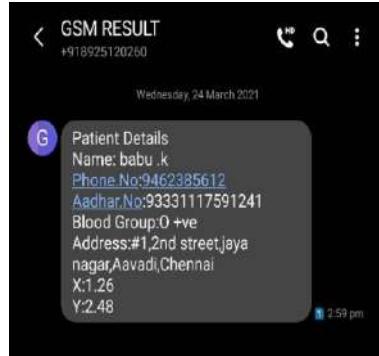
5. Conclusion

This project keeps up its adaptability by protecting all information in a solid IoT in CLOUD. The information moved through an IOT CLOUD is guaranteed to be confidential and secure. The graphical portrayal acquired in the approved web worker that speaks with IoT cloud is accurate and exact. Henceforth, directing the vehicle through a worker-decided way will save precious, meaningful time for the patient, bringing about diminished odds of mortality just as a considerable decline accordingly the paramedics' response time.

References

[1]   Divya D. Nanwani, Puja R. Kshirsagar, Bhavana P. Kawal   Kar, Pritish Deshmukh "Ambulance Tracking and Patient   Health Monitoring Using GPS and GSM" International Journal of Emerging Technologies in Engineering Research (IJETER) Volume 5, Issue 3, March (2017), ISSN: 2454-6410.

[2]   Baksi, A., Bhattacharjee, M., Ghosh, S., Bishnu, S. K., & Chakraborty, A. (2020). Internet of Things (IOT) Based Ambulance Tracking System Using GPS and GSM Modules. In *2020 4th International Conference on Electronics, Materials Engineering and Nano-Technology, IEMENTech 2020*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/IEMENTech51367.2020.9270120

[3]   Bhagchandani, K., & Peter Augustine, D. (2019). IoT based heart monitoring and alerting system with cloud computing and managing the traffic for an ambulance in India. *International Journal of Electrical and Computer Engineering*, *9*(6), 5068–5074. https://doi.org/10.11591/ijece.v9i6.pp5068-5074

[4]   V.Baarathi, T.Balaji, A.T.Navin Prasad, M.Jansi Rani "Offline SMS Internet" , International Research Journal of Engineering and Technology (IRJET), Vol.No.4, Issue No.3, PP. 2376-2378, March 2017, ISSN:2395-0056.

[5]   G.K.Srinandhala J.Navarajan, S.Manoj, R.Saravanan, "Remote Patient Monitoring Using Internet of Things" International Journal of Advanced Research Methodology in Engineering & Technology, Volume 1,Issue 3,Pages 14-16, May 2017.

[6]   M.Jansi Rani, Jahan.J, Anitha.D, Hennah Florida.V " Intelligent Farm House Automation Using Green House Method and Plant Image Processing Using MATLAB", International Journal of Advanced

Research in Electrical, Electronics and Instrumentation Engineering(An ISO 3297: 2007 Certified Organization)Vol. 5, Issue 3, March 2016.

[7]   J.Navarajan, B. Aswin kumar, S. venkatesh , T. Jayachandran "Detection of Water Pollution and Water Management Using Smart Sensors with IOT", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 04 Issue: 04 | Apr -2017.

[8]   C. Jain, G. V. S. Sashank, V. N and S. Markkandan, "Low-cost BLE based Indoor Localization using RSSI Fingerprinting and Machine Learning," Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2021, pp. 363-367, https://doi.org/10.1109/WiSPNET51692.2021.9419388

[9]   C. C., G. B. R., S. M., A. M., B. R. and M. Balaji, "Machine Learning Based Condition Recognition System for Bikers," 2020 7th International Conference on Smart Structures and Systems (ICSSS), 2020, pp. 1-6, 10.1109/ICSSS49621.2020.9202245

LORA (Long Range) Based Low Power Emergency Tracking System in Disaster Areas

S. Sweetline Shamini [a,1], A.P. Anuja [a], B. Archana [a], P. Premalakshmi [a]

[a]Department of Electronics and Communication Engineering, Sri Sai Ram Institute of Technology, Chennai, Tamil Nadu, India

Abstract. The communication network is one of the major assistances for disaster management purposes. At each disaster phase, the information flow between the disaster management departments and the population is critical, as it destroys the whole communication infrastructure where the base station to Mobile Switching Centre (MSC) connectivity is broken, power generators are exhausted, and the network may suffer from congestion. Even when radio equipment works, police officers, rescue teams, medical teams cannot communicate with each other because of incompatible radios. To analyze the characteristics of the geological environment in the disaster region and to alert the Disaster Management Department, this solution based on LoRa has been proposed. The penultimate goal of this paper is to fasten the rescue operations.

Keywords. LoRa, Spreading Factor, Communication, Disaster Management

1. Introduction

Often calamities such as earthquakes, hurricanes, floods affect communication services like mobile phone networks, commercial radio networks, and other communication services infrastructures, unavailable for any disaster management activities. LoRa is implemented using the non-cellular Low Power Wide Area Network (LPWAN), a wireless communication network protocol operating over a range of 2km - 5km without Line of Sight (LoS) and about 15km - 20km with Line of Sight. Since it works under LPWAN the power consumption varies between 10mW - 100mW which is minimum when compared to Wi-Fi, LTE, EDGE. LoRa uses a modulation technique which is similar to Chirp spread spectrum modulation, uses unlicensed radio spectrum in the ISM band that provides seamless interoperability between devices. Henceforth this idea is implemented using LoRa that focuses on data transmission which requires a minimum data rate over a wider area. LoRa wireless outcomes are the best alternative for eliminating the need for repeaters, minimizing cost consumption, improving battery life, and also enhancing network capacity. LoRa transceivers are Long Range wireless modem helps in preventing and improving the selectivity over traditional modulation approaches, to overcome the flaws of conventional methodology balancing between longer range of transmission as well as high interference immunity by minimizing current consumption.

---

[1] S. Sweetline Shamini, Department of Electronics and Communication Engineering, Sri Sai Ram Institute of Technology,   Chennai, Tamil Nadu, India, Email: sweetlineshamini.ece@sairamit.edu.in.

2.   Existing Methods

With respect to the paper, **Emergency communication in IoT scenarios by means of a transparent LoRaWAN enhancement** published by Emiliano Sisinni in 2020, concluded that their proposal is effective in reducing the loss of emergency messages without retries [1]. The LoRa (Long-Range) High-Density Sensors for Internet of Things, published by Alexandru Lavric in 2019, determined the maximum number of sensors that can be integrated to perform IoT service [2]

Post Disaster Relief Operations using Ad-hoc Network, Zigbee protocol evaluates whether peer to peer communication can reach out the long distance without any loss of information [3]. In reference to the proposed paper, Radio Data Infrastructure for Remote Monitoring System using LoRa Technology published by Ravi Kishore Kodali in 2017, concluded that The LoRa Wireless technology offers long range for communication for IoT services [4]. In disaster regions, Walkie Talkies and Citizen Band Radios are used which support communication for a shorter distance where walkie-talkies will perform only 1/10th of its performance and the efficiency of Ham or Amateur radios are also deteriorated at times of poor weather condition [5].

3.   Proposed Solution

3.1   Introduction

Communication is vital for providing the right support at the time of disaster. In this solution, we have used LoRa for long-distance transmission which allows us to make communication way easier and more cost-effective [6].

3.2   Working
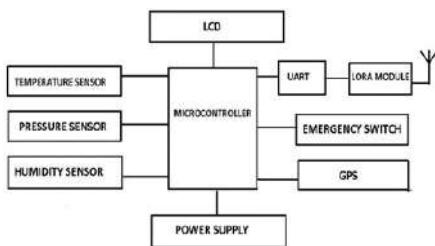
Figure 1. Block Diagram – Transmitter section   **Figure 2.** Block Diagram –Receiver section
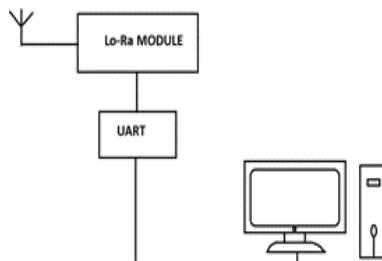
In Figure 1, Transmitter section and in Figure 2, Receiver section block diagrams are presented. At the time of any natural calamity, when there is no proper infrastructure-based communication, the proposed solution may be used. The emergency switch is pressed physically, the transmitting unit is triggered, that is the microcontroller continuously looks for the emergency impulse and this impulse is applied using a switch. With the press of the emergency switch, the external components are also

triggered and start reading the values from the surrounding environment. The GPS unit sends the current location and the sensor data such as temperature sensor, humidity, pressure sensor to the microcontroller. The power supply acts as a source for the microcontroller. The type of emergency message is prerecorded in different switches and could be transmitted easily [7].

At the receiver section that is placed at a significant distance, the transmitted data is received by the LoRa module and data are viewed in a tracker application.

Figure 3. Implementation of transmission

Figure 4. Implementation of receiver section

3.3 Output

For the implementations shown in Figure 3 and figure 4, the outputs are taken. An application was built using visual basic code to view the received data via the serial port from the LoRa module. This data can be logged with the help of an application called Tera Term into a CSV file from where we could visualize the data in the form of graphs in Figure 5 and Figure 6 for better understanding of the disaster hit zone.
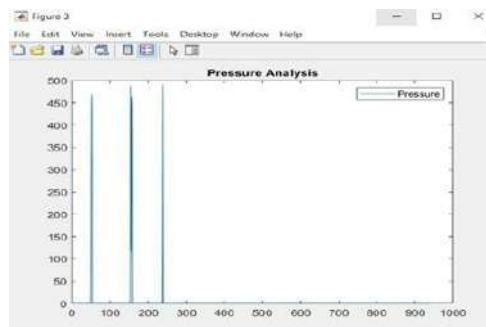
Figure 5. A Plot of received Pressure data.
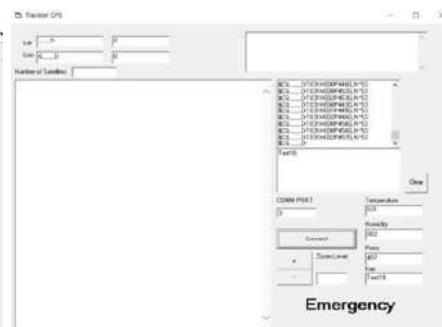
Figure 6. Tracker application showing the received data

LoRa modulation technique uses orthogonal spreading factors i.e. it can send messages concurrently with different spreading factors with the same frequency and can be received without interference as they appear noise to each other [8]. Thus, this technique offers high immunity to multipath fading and interference.

The capacity of LoRa is given by,

$$Capacity = 1*(\sum_{i=7}^{12} SF_i \ equivalent \ bit \ error \ rate \ )$$

From Semtech LoRa specification,

$$Capacity_{LoRa} = 12.156 \ kb/s$$

The Figure 7, shows the depiction of 500 random messages each of 25 bytes and the nodes/ devices are allowed to select a random spreading factor for concurrent use of the available spectrum. Here, Green depicts efficient transmissions while the Red depicts the failed transmissions with the use of the same spreading factor.
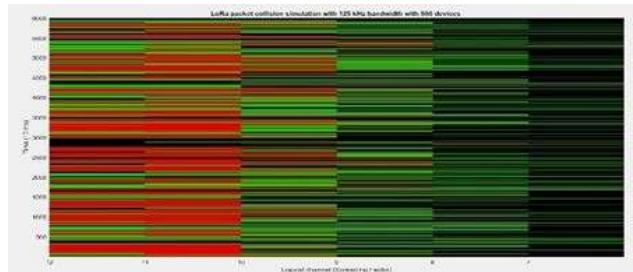
Figure 7. Simulation of collision of packets for different spreading factors
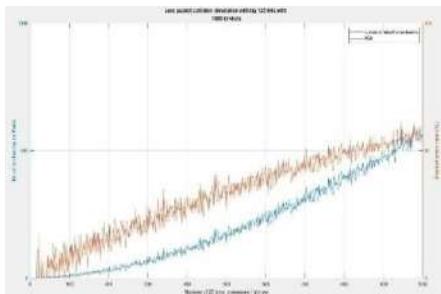
Figure 8. Simulation showing the Packet   **Figure 9.** Simulation showing the collision and collision
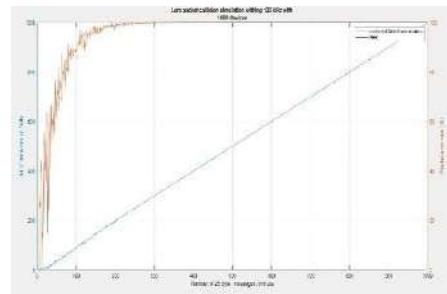
Figure 8 and Figure 9 shows the simulation outputs. Packet Error Rate with a spreading factor of 12 Error Rate with a spreading factor of 7. The above figures show a comparison between the ideal spreading factor (SF7) and the worst spreading factor (SF12).

If we consider the ideal case where we have only used a spreading factor of 7, it results in a short message of about 36 ms. Here the collisions are minimum and the bit rate is maximum while the range of transmission is around 2 km only.

In another scenario, the spreading factor is fixed as 12. This results in long messages (around 682 ms). Here the collisions are high and the bit rate is reduced but the transmission range is up to 10 km.

LoRa chooses the right spreading factor by considering the distance between the transmitter and the receiver. As this setup consists of several sensors, an optimal distance of 2 km is chosen as a range of communication to avoid packet loss.

4. Performance Analysis

To examine the performance of the proposed solution, as per the Figure 10, the following test was conducted. A random location was fixed as the transmitter and six random locations at different distances were chosen as receivers. The transmitting section was kept at a height of 5 m and the sensed data were transmitted to the receiver at each location and this experiment was repeated nine times. The Table 1, shows the success rate for each location.

From the table, we can conclude that up to a distance of 500 m, none of the trials failed. But as the distance gradually increases the success rate decreases. For a distance of 2 km, the success rate was about 70%. This could be improved by placing the transmitting module at a height of 30 m as it would improve the line of sight, enabling us to establish communication to a greater distance.

Figure 10. A Track of Performance Analysis

Table 1. Performance Analysis outcomes considering range

No. of Trials : 10		
Distance between the Transmitter and Receiver (in m)	No. of successful Trails	Success Rate (in %)
180	10	100
300	10	100
450	10	100
650	9	90
1000	8	80
1800	7	70

5. Conclusion

As discussed already, calamities are inevitable, and establishing communication during this situation is a key factor to bounce back. Thus, this solution based on LoRa

provides us a way to control the spreading factor thereby efficiently controlling the range of transmission which was not possible using other protocols like Wi-Fi and also consumes much lesser power. With such an increase in range of the transmission, the need for a proper centralized base station, all the necessary data could be obtained and this module can also be mounted on Unmanned Aerial Vehicles, thus monitoring places affected by disasters such as forest fire, where human being survival is impossible. Hence, this setup would act as an optimal solution with high reliability, extremely low cost, secure and high interoperability.

6.   Future Work

Alongside data transmission, this setup can be extended for voice and image transmission. Since, LoRa is suitable for communication involving low data rates, identifying proper compression techniques for voice and image is required. Doing so will enable us to analyze and understand the environment from a broader perspective and help us to calculate the remedial measures soon and with even more high accuracy.

References

[1] Kodali, R. K. (2017). Radio data infrastructure for remote monitoring system using LoRa technology. In *2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017* (Vol. 2017-January, pp. 467–472). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICACCI.2017.8125884

[2] Sisinni, E., Carvalho, D. F., & Ferrari, P. (2020). Emergency Communication in IoT Scenarios by Means of a Transparent LoRaWAN Enhancement. *IEEE Internet of Things Journal*, *7*(10), 10684–10694. https://doi.org/10.1109/JIOT.2020.3011262

[3] P.Lavanya, B.Somi, S.Dev Nandan, J.Mani Mohana Krishna, N.Veeraiah, Post Disaster Relief Operations using Ad-hoc Network, 2020

[4] Rene Tendeng, YoungDoo Lee,and Insoo Koo, Implementation and Measurement of Spectrum Sensing for Cognitive Radio Networks Based on LoRa and GNU Radio

[5] Alexandru Lavric, Valentin Popa, Internet of Things and LoRaTM Low-Power Wide Area Networks: A Survey, 2017.

[6] Ganesh Babu R., Karunakaran A., Manikandan G., Kalimuthu Kumar S., Selvameena R. (2021) IoT in Smart Automation and Robotics with Streaming Analytical Challenges. In: Singh K.K., Nayyar A., Tanwar S., Abouhawwash M. (eds) Emergence of Cyber Physical System and IoT in Smart Automation and Robotics. Advances in Science, Technology & Innovation (IEREK Interdisciplinary Series for Sustainable Development). Springer, Cham. https://doi.org/10.1007/978-3-030-66222-6_7

[7] Markkandan, S., Malarvizhi, C., Raja, L., Kalloor, J., Karthi, J., & Atla, R. (2021), "Highly compact sized circular microstrip patch antenna with partial ground for biomedical applications", Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2021.04.480

[8] C. C., G. B. R., S. M., A. M., B. R. and M. Balaji, "Machine Learning Based Condition Recognition System for Bikers," 2020 7th International Conference on Smart Structures and Systems (ICSSS), 2020, pp. 1-6, https://doi.org/10.1109/ICSSS49621.2020.9202245

Design and Analysis of 64 GHz Millimetre Wave Microstrip Patch Antenna

Swagata Sarkar[a,1], Sivakami Nagappan[a], Shafin Kadhir Badhusha S[b]

[a]*Assistant Professor ,*[b]*Student, Sri Sairam Engineering College, Chennai, TN, India*

Abstract. Millimetre Wave frequencies (30-300 GHz) can be used for different major applications of modern world like telecommunications, security screening, imaging, automotive radars, military applications, remote sensing, radio astronomy and many more. The internationally reserved frequency spectrum is used for Radio Frequency Energy. In this work 64 GHz antennas are compared with different design and a comparative study is taken. In this work Microstrip patch antenna with carpet architecture, and fractal island are designed and compared. The general comparative parameters for antenna are directivity, gain, return loss, bandwidth, specific absorption rate etc. After the comparison, it is found that return loss gave better result for carpet design at 64 GHz compare to fractal island design.

Keywords. Millimetre Wave Antenna, ISM band, Carpet Antenna, Fractal Island, Microstrip Patch Antenna

1. Introduction

In modern age, the design of Microstrip Patch antenna plays a major role in RF industry. After the major pandemic, no body of us can deny the relevance of communication. Antenna is the heart of the communication which is working as a transducer to convert electrical signal to electromagnetic signal in the transmitter side and vice versa in the receiver side. Different antennas have its own frequency of applications and designs. Microstrip Patch antenna got more popularity based on its compact size, low cost and other useful major parameters. The carpet and fractal island design were fabricated on basic Microstrip Patch Antenna. Fractal and Metamaterial can enhance the performance of the antenna as discussed in literature.

2. Literature Review

In literature different antenna shapes like pentagonal, octagonal, hexagonal [1] are furnished with different return loss. Metamaterial and Fractal structure [2] can also be used for enhancing the antenna performance.[3]. In modern age, the design of Microstrip Patch antenna plays a major role in RF industry. After the major pandemic, no body of us can deny the relevance of communication.

---

[1] Swagata Sarkar, Assistant Professor, Sri Sairam Engineering College, Chennai, TN, India; Email: swagata.b.sarkar@gmail.com

Antenna is the heart of the communication which is working as a transducer to convert electrical signal to electromagnetic signal in the transmitter side and vice versa in the receiver side. Different antennas have its own frequency of applications and designs. Microstrip Patch antenna got more popularity based on its compact size, low cost and other useful major parameters. The carpet and fractal island design were fabricated on basic Microstrip Patch Antenna. Fractal and Metamaterial can enhance the performance of the antenna as discussed in literature [4-6].

Design Specifications

The basic design for carpet design and fractal island design is the design of microstrip patch antenna [7]. The basic design specifications are explained below.

1.   Mathematical Modeling for Square Patch Antenna

Patch Width $(w) = \frac{c}{2} * \frac{1}{fr} * \sqrt{\frac{2}{\varepsilon r+1}}$

$\qquad$ (1)

Where c=free space velocity of light=3 x 10^{10} cm/sec

fr =Resonating Frequency=64 GHz

εr=permittivity=4.4

Patch Width is **1.43 mm**

Effective Permittivity

Eeff $=\frac{\varepsilon r+1}{2} + \frac{\varepsilon r-1}{2} * \sqrt[-1]{\left(1 + 12 * \frac{h}{w}\right)}$

$\qquad$ (2)

Effective permittivity is **3.148 mm**

Effective length of the patch

$L_{\text{eff}} = \frac{c}{2} * 1/fr\sqrt{Eeff}$

$\qquad$ (3)

Where c= velocity of light (Free Space) =3 x 10^{10} cm/sec

Resonating Frequency=64 GHz

Eeff=Effective permittivity =3.148 mm

Effective length of the patch is **1.32 mm**

Change of Length

Del L=$0.412 * h * \frac{Eeff+0.3}{Eeff-0.258} * \frac{\frac{w}{h}+0.264}{\frac{w}{h}+0.8}$

$\qquad$ (4)

Change of Length is **0.5375 mm**

Length of the patch L=L_{eff}-Del L $\qquad\qquad\qquad\qquad\qquad\qquad$ (5)

Length of the patch is **0.7825 mm**

Antenna Design parameters	Value
Resonance Frequency (Basic Square Patch)	64 GHz
Substrate	FR4 Epoxy
Height of the patch	1.6 mm
Permittivity	4.4
Length of one side of Square patch	**1.43 mm**&**0.7825** mm (for rectangular patch) Theoretical

3.   Discussions about Results

In this part the design aspects of different microstrip patch antenna are going to be explained elaboratively. Microstrip Patch antenna is the basic building block of all derived design shown in Figure 1. In this work carpet design and fractal island on top of microstrip patch antenna are designed with 64 GHz frequency. Both the designs are compared.

A. Sierpinski Carpet Antenna Design
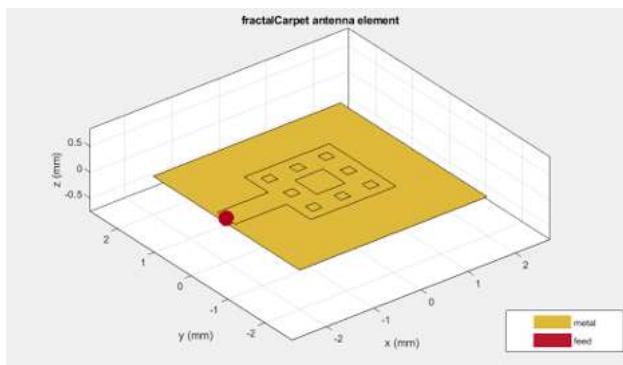
Figure 1. 64 GHz Carpet Design on Microstrip Patch Antenna

Fractal geometry with Electromagnetic theory is a very attractive antenna design. In this work Sierpinski Carpet Antenna Design is added with microstrip patch antenna to get better result is shown in Figure 1.

Antenna Impedance is the relationship between current and voltage at the input of the antenna. It has two parts. One is real and another one is imaginary. Real parts represent the power that is either absorbed or radiated, whereas imaginary part tells us the power that is reserved in the near field of the microstrip patch antenna.
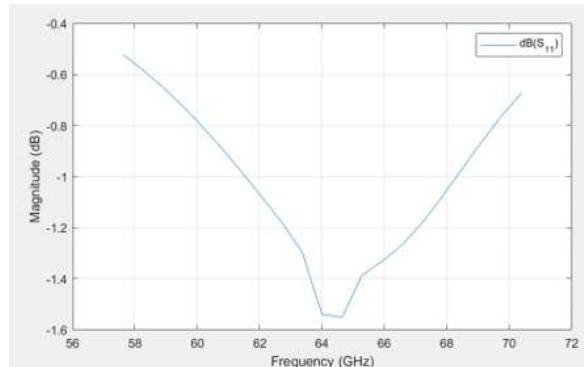
Figure   2. Return Loss of the recommended Microstrip Patch Antenna

The designed antenna Return Loss is 1.6 at 64 GHz. It can be further improved by changing the fractal design is shown in Figure 2.

B. Minkowski Fractal Island Antenna Design

The performance of the Microstrip Patch Antenna can be enhanced by introducing Minkowski Island design on the top of the existing antenna is shown in Figure 3.
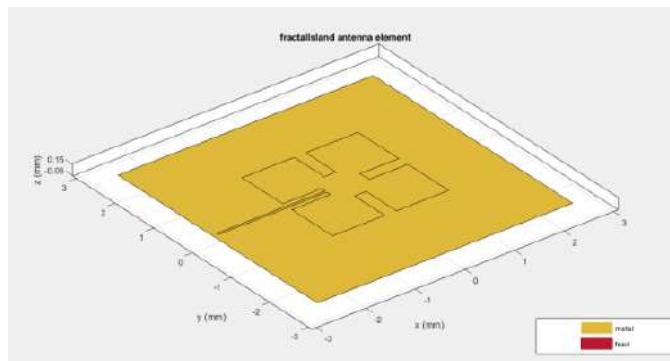
Figure 3. 64 GHz Minkowski Fractal Island Antenna

In this work four carpet structures are introduced in the microstrip patch antenna.

The Real and imaginary part of the impedance perfectly cross each other at 64 GHz .
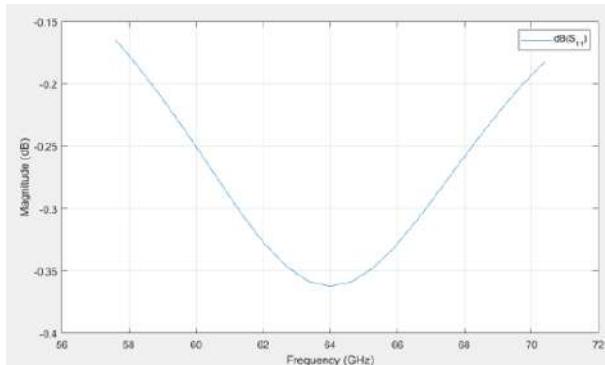
Figure 4. Return Loss of Fractal Island Antenna

Compare to Carpet antenna, return loss is reduced in Fractal Island antenna. But smoothness and perfection nature of the output make the design better than the carpet design is shown in Figure 4.

The current distribution of Fractal Island antenna has poor quality of distribution compared to carpet design. Radiation pattern shows the distribution of the field for fractal island antenna

5.   Conclusion

In this work, 64 GHz Carpet antenna and fractal Island antenna are designed and compared. It is shown that the performance of the carpet antenna is better than fractal island antenna in different aspects. The performance of both the antenna can be further enhanced by using the number of iterations. Fractal design is very much promising to add with antenna technology to enhance the performance of the antenna.

References

[1]   Rajesh, G. S., Kishore, K. V., & Kumar, V. (2015). Multiband microstrip patch antenna design using metamaterial for airborne SAR system. In *2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems, SPICES 2015*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/SPICES.2015.7091444

[2]   Werner, D. H., & Ganguly, S. (2003). An overview of fractal antenna engineering research. *IEEE Antennas and Propagation Magazine*, *45*(1), 38–57. https://doi.org/10.1109/MAP.2003.1189650

[3]   Joshi, J. G., Pattnaik, S. S., Devi, S., Lohokare, M. R., & Vidyasagar, C. (2009). Offset fed Diamond Shaped Split Ring (DSSR) planar metamaterial antenna. In *Applied Electromagnetics Conference, AEMC 2009 and URSI Commission B Meeting*. https://doi.org/10.1109/AEMC.2009.5430686

[4]   Ripin, N., Saidy, W. M. A. W., Sulaiman, A. A., Rashid, N. E. A., & Hussin, M. F. (2013). Miniaturization of microstrip patch antenna through metamaterial approach. In *Proceeding - 2013 IEEE Student Conference on Research and Development, SCOReD 2013* (pp. 365–369). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/SCOReD.2013.7002610

[5]   Patel, S. K., & Kosta, Y. P. (2014). S-Shape meandered microstrip patch antenna design using metamaterial. In *IEEE Antennas and Propagation Society, AP-S International Symposium (Digest)* (pp. 547–548). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/APS.2014.6904604

[6]   B. Sarkar, S. (2016). "Design and analysis of multiband octagonal microstrip patch antenna with different annular ring". *ICTACT Journal on Microelectronics*, *2*(2), 237–242. https://doi.org/10.21917/ijme.2016.0040

[7]   Markkandan, S., Malarvizhi, C., Raja, L., Kalloor, J., Karthi, J., & Atla, R. (2021), "Highly compact sized circular microstrip patch antenna with partial ground for biomedical applications", Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2021.04.480

Customer Analysis Using Machine Learning Algorithms: A Case Study Using Banking Consumer Dataset

R.Siva Subramanian [a,1], D.Prabha [b], B.Maheswari [c], J.Aswini [d]

[a]Research Scholar, Anna University, Chennai-600025,India
[b]Associate Professor, Sri Krishna College of Engineering and Technology, Coimbatore, India
[c]Assistant Professor, Rajalakshmi Institute of Technology, Chennai, India
[d]Assistant Professor, Sree Vidyanikethan Engineering College, Tirupati, India

Abstract. The aim of each enterprise is to achieve high revenue from the business and to stay in a high position from their competitors. To archive high revenue and high position from competitors the need of understanding the business consumers is a crucial one. However the firm business is completely dependent on the consumers the efficient analysis of consumers within the enterprises makes to achieve the business to high position. To perform effective consumer analysis, in this study different machine learning is studied and experimented. ML classifiers make to understand in-depth analysis about the consumer data and further enables to plan wise decision strategies to enhance the business revenue and consumer satisfaction intelligently. The use of different ML classifiers is to sort out how the customer prediction outcome changes accordingly to the ML classifier is applied. This makes to find the best ML classifier for the consumer dataset applied in this study. The experimental procedure is performed using different ML classifiers and the outcome achieved is captured and projected using various validity scores. This work applies different ML classifiers like K-NN, C4.5, Random Forest, Random Tree, LR, MLP and NB for customer analysis. The empirical results illustrate the C4.5 model achieves better accuracy prediction compare to other ML classifiers and also compared with the time complexity NB model works efficiently with running time.

Keywords. Artificial Intelligence, Business decision, Customer   Prediction, Machine Learning

1. Introduction

CRM is an efficient tool that helps to build, manage, and analyze a customer relationship with enterprises. The use of CRM makes the enterprises efficiently collect, store and assess the consumer-related data intelligently and make it available across the enterprise business people [1]. The assessment of consumer data efficiently makes the firms understand in-depth patterns of consumer behavior and further enables them to create wise decision strategies to increases the business revenue and customer retention & acquisition wisely. To perform consumer analysis in CRM, Analytical CRM is

---

[1] R.Siva Subramanian, Research Scholar, Anna University, Chennai-600025, India; E-mail: Email:sivasubramanian12@yahoo.com.

carried out. ACRM helps to analyze the stored consumer data in CRM to find interesting customer patterns which the help of machine learning approaches.

Unlike other CRM, ACRM works behind without having any interaction with the customers and predicts the customer analysis wisely. The use of ACRM helps to increases 1. Marketing technique, 2. Retaining potential consumers, 3. Understanding customer needs, 4. increasing consumer satisfaction, 5. improved consumer loyalty, and 6. Efficient customer experience [2]. In ACRM, ML approaches are applied to perform customer analysis.

This work applies different ML classifiers like K-NN, C4.5, Random Forest, Random Tree, LR, MLP and NB for customer analysis in banking enterprises [3]. The work aims to compute the best ML algorithm for the customer analysis problem and in turn, applies the further decision to improve the firm business. Here customer dataset applied is derived from the UCI repository. The experimental procedure is performed using different ML algorithms and the outcome achieved is indicated using the various validity scores. Further time complexity is measured to find out the running time of the ML classifier. The remainder of the work starts with a literature survey, discussion of ML classifier applied, experimental results, and lastly end with conclusions.

2. Literature Survey

The author addresses the problem of customer churn analysis and proposes LLM model to analyze customer churn [4]. The proposed approach is comprised of two phases: one is segmentation & another one is prediction. The experimental procedure is carried out using 14 customer churn datasets and the outcome captured is projected using AUC and TDL parameters. The outcome obtained is compared using LMT, LR, and RF, and results reveal proposed LLM is better. The author in this work assesses the problem of customer churn and to analyze the customer churn the author proposes four approaches based upon RST: one is Genetic algorithm, two is Exhaustive algorithm, third is Covering algorithm and the last one is LEM2 algorithm [5]. The experimental procedure is carried out using a telecom dataset and the results captured reveal GA-RST performs wisely compare to other approaches. [6] The author address the prediction of consumer patterns helps the firms to plan and allocate the resources efficiently and improve the business wisely [8]. To predict the customer analysis the author proposes some consumer-relevant attributes from previous purchases and use three different ML algorithm to analyzes these data by indoor localization and movement prediction [9]. Experimental results reveal gradient model performs well compared to the other two approaches. The experimental procedure is carried out using 10000 customer data

3. Methodology

The methodology first starts with the input customer dataset and then data processing is carried out to remove the missing and nosily data from the customer dataset. Then different ML classifiers are applied to perform customer analysis using NB, KNN, LR, MLP, C4.5, Random Forest, and Random Tree. Further, the outcome captured is recorded using different performance metrics like accuracy, recall, specificity, FNR,

FPR, F-Measure, and Precision are listed in Table 1. Lastly, the model which archives the best performance is projected. Different ML algorithm applied to perform customer analysis with advantages and disadvantages.
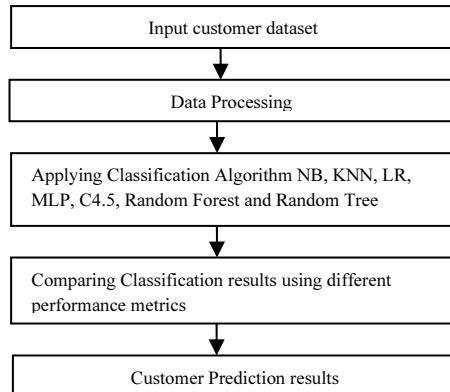
Figure 1. Overall methodology for customer analysis

4. Experimental Results

To analysis about the customer patterns in this study different Ml classifiers like NB, LR, MLP, C4.5, Random Forest, KNN, and Random Tree are studied and experimented.

4.1 Dataset and Validity Scores:

This study applies a Bank dataset derived from UCI for customer analysis purposes. The intent of the consumer data is to predict whether the consumer uses the banking service are not based on the analysis of the customer data. The database comprised of 45211 instances with 17 input predictors [7]**.** Performance is computed using various metrics like accuracy, recall, precision, f-measure, f-measure, TPR, FPR, and specificity are shown in Figure 2 and Figure 3, respectively.

4.2 Experimental Procedure:

1. The experimental procedure is carried out using the different ML classifiers like NB, LR, MLP, C4.5, Random Forest, KNN, and Random Tree.
2. The empirical outcome captured from the above approaches are represented and compared in terms of accuracy, recall, precision, f-measure, f-measure, TPR, FPR, and specificity.
3. Further the time complexity for the above approaches was also recorded and compared.

Table 1: Represents the outcome of different ML classifiers applied to the problem of customer analysis with a summary of accuracy, recall, precision, of F-measure, FNR, FPR, specificity and time complexity measure.

S.no	Methodology	Accuracy	Recall	Precision	F-measure	FNR	FPR	Specificity	Time Complexity in seconds
1	NB	88.3789	0.394	0.504	0.442	0.606	0.051	0.9486	0.11
2	LR	89.664	0.319	0.612	0.419	0.681	0.027	0.9732	12.35
3	MLP	89.622	0.427	0.576	0.490	0.573	0.042	0.9584	33.56
4	C4.5	89.8786	0.349	0.620	0.446	0.651	0.028	0.9716	0.64
5	Random Forest	89.1288	0.353	0.556	0.432	0.647	0.037	0.9625	11.6
6	KNN	88.5116	0.237	0.520	0.326	0.763	0.029	0.9709	0.40
7	Random Tree	87.3571	0.352	0.449	0.395	0.648	0.057	0.9426	0.45
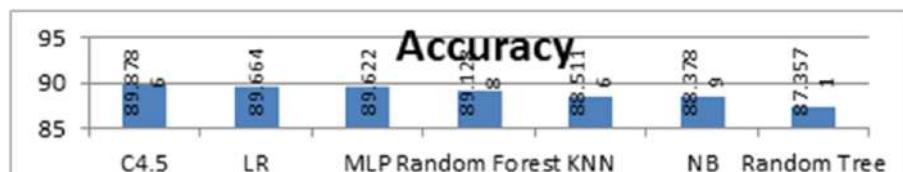
Figure 2. Represents the accuracy comparison obtained from the different ML classifiers
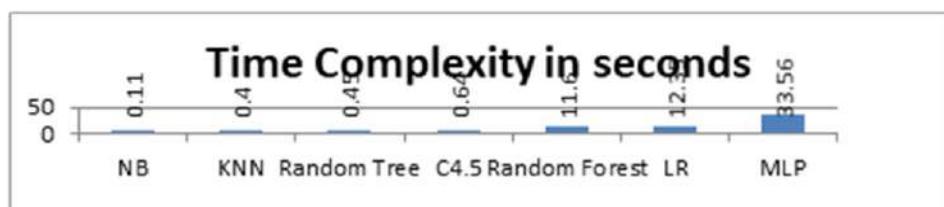
Figure 3. Represents the time complexity comparison obtained from the different ML classifiers

4.3   Result Discussion:

The experimental outcome captured to analyze customer patterns using different ML classifiers is presented in Table 1, it indicates the outcome of different ML classifiers in terms of accuracy, recall, precision, of F-measure, FNR, FPR, specificity and time complexity measure.From the empirical outcome, we can see C4.5 model gets 89.8786 higher accuracy compare to other approaches. The next higher accuracy 89.664, 89.662, and 89.1288 is derived from LR, MLP, and Random Forest approaches. The poor prediction is captured from Random Tree with an accuracy of 87.4588. Next in terms of recall parameter, we can see MLP achieves a high prediction of 0.427. The next higher recall prediction 0.394, 0.391, 0.353 is derived from NB, NBTREE, and Random Forest. Poor recall prediction is captured from KNN with a prediction of 0.237. Next in terms of precision parameter, we can see C4.5 achieves a high prediction of 0.62. The next higher precision prediction 0.612, 0.576, and 0.556 is derived from LR, MLP, and Random Forest. Poor precision prediction is captured from Random Tree with the prediction of 0.449. In terms of F-measure, we can see MLP model 0.49 higher prediction compare to other approaches. The next higher F-Measure prediction 0.446, 0.442, and 0.432 are derived from, C4.5, NB, and the Random Forest approach. The poor prediction is captured from KNN with a 0.326 prediction. Next in terms of the FNR parameter, we can see KNN achieves a high prediction of 0.763. The next higher FNR prediction 0.681, 0.651, 0.68 is derived from LR, C4.5, and Random Tree. Poor FNR prediction is captured from NB and MLP with the prediction of 0.606 and 0.573. Next in terms of the FPR parameter, we can see Random Tree achieves a high prediction of 0.057. The next higher FPR prediction 0.051, 0.042, and 0.037 is derived from NB, MLP, and Random Forest. Poor FPR prediction is captured from LR with the prediction of 0.027. Next in terms of the Specificity parameter, we can see LR achieves a high prediction of 0.9732. The next higher Specificity prediction 0.9716, 0.9709, 0.9625 is derived from C4.5, KNN, and Random Forest. Poor Specificity prediction is captured from Random Tree with the prediction of 0.9426. We can see efficient time complexity is got in NB of 0.11 seconds. The next efficient time complexity is got in KNN, Random Tree, and C4.5 of 0.40, 0.45, and 0.64 seconds. Poor time complexity is got in MLP of 33.56 seconds.

5. Conclusion

The intent of the work is to analyze the customer patterns using different ML classifiers and to sort the wise ML classifier in customer analysis. The analyzes of better customer patterns make to understand the consumer's behavior patterns intelligently and further based upon the analysis the customer satisfaction and business revenue can be increased. To understand customer patterns use of different ML classifier are applied and investigated. This research applies different ML classifiers like NB, LR, MLP, C4.5, Random Forest, KNN, and Random Tree are studied and experimented. The empirical outcomes indicate C4.5 model ends with 89.8786 higher accuracy compare to other approaches. Likewise, poor accuracy is generated in the Random Tree of 87.3571 compare to other approaches. In terms of recall, we can see MLP achieves a high prediction of 0.427 a likewise low recall prediction is captured from KNN with the prediction of 0.237. In terms of precision, we can see C4.5 achieves a high prediction of 0.62 and likewise, less precision prediction is captured from Random Tree with the

prediction of 0.449 precision. In terms of F-measure, we can see the MLP model gets 0.49 high F-measure prediction and likewise less prediction is captured from KNN with 0.326 F-measure predictions. In terms of time complexity, we can see the NB model gets 0.16 seconds best time complexity and likewise less time complexity is captured from MLP with 33.56 Seconds. This research applies only one customer data and in future uses of different customer dataset can be applied.

References

[1] Dalla Pozza, I., Goetz, O., & Sahut, J. M. (2018). Implementation effects in the relationship between CRM and its performance. *Journal of Business Research*, *89*, 391–403. https://doi.org/10.1016/j.jbusres.2018.02.004

[2] Subramanian, R. S., & Prabha, Dr. D. (2017). A Survey on Customer Relationship Management. *2017 4th International Conference on Advanced Computing and Communication Systems, ICACCS 2017*.

[3] Coussement, K., Lessmann, S., & Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, *95*, 27–36. https://doi.org/10.1016/j.dss.2016.11.007

[4] De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, *269*(2), 760–772. https://doi.org/10.1016/j.ejor.2018.02.009

[5] Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K. (2017). Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing*, *237*, 242–254. https://doi.org/10.1016/j.neucom.2016.12.009

[6] Martínez, A., Schmuck, C., Pereverzyev, S., Pirker, C., & Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*, *281*(3), 588–596. https://doi.org/10.1016/j.ejor.2018.04.034

[7] Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, *62*, 22–31. https://doi.org/10.1016/j.dss.2014.03.001

[8] Ganesh Babu R., Elangovan K., Maurya S., Karthika P. (2021) Multimedia Security and Privacy on Real-Time Behavioral Monitoring in Machine Learning IoT Application Using Big Data Analytics. In: Kumar R., Sharma R., Pattnaik P.K. (eds) Multimedia Technologies in the Internet of Things Environment. Studies in Big Data, vol 79. Springer, Singapore. https://doi.org/10.1007/978-981-15-7965-3_9

[9] C. Jain, G. V. S. Sashank, V. N and S. Markkandan, "Low-cost BLE based Indoor Localization using RSSI Fingerprinting and Machine Learning," Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2021, pp. 363-367, https://doi.org/10.1109/WiSPNET51692.2021.9419388

Covid Patient Health Monitoring Using IoT During Quarantine

Lavanya Dhanesh [a,1], Meena.T [b], Chrisntha.B [b], Gayathri.S [b], Devapriya.M.D [b]
[a]Associate Professor, [b]UG Student, Department of Electrical and Electronics Engineering, Panimalar Institute of Technology, Chennai, TN, India

Abstract. The term "COVID" is breaking the hearts of the entire human community. The Corona virus is more infectious and is exceptionally irresistible, it is vital to isolate the patients and yet the specialists need to screen Corona virus patients as well. With the expanding increase in the number of Corona cases, the doctors find it difficult to keep track on the medical issue of isolated patients. To address this issue, we designed a distant IOT based screen framework, that considers observing of numerous Corona virus patients over the web. The system uses temperature sensor, respiratory sensor and pulse oximeter to measure the health parameters of the patients. If any oddity is detected in patient's health, the patient presses the emergency help button which we installed in our system. This will alert the doctor and the care taker over IOT remotely. Our system thus provides a safe health monitoring design, in order to prevent the disease spreading through Corona virus and monitoring the individual health of each patient.

Keywords. Corona patient, Quarantine, Zigbee, Arduino, RTC, IOT, Sensors.

1. Introduction

Corona virus has made the requirement for populace level screening, and telemedicine is obviously situated to empower this. As telemedicine has developed in the course of the most recent decade, distant checking arose as another and incredible methodology. Corona virus requires scaled co-operations with populaces in close to continuous. Distant checking has explicit operational and configuration includes that are appropriate for the COVID-19, particularly the offbeat correspondence. Checking can be utilized specifically to accumulate pandemic information and acquire continuous clinical input. As telemedicine proceeds to develop and advance, distant checking is arising as an important instrument for payers, suppliers, and general wellbeing authorities the same [1]. Thus this system helps in monitoring the quarantine people's health remotely with the help of IOT and wireless sensor network. Each patient will be enabled with the test kit with them. The system will continuously monitor the patient health parameters like heartbeat, temperature, and blood pressure values [2-3]. Once the value reaches any abnormal value the system will send an intimation to the doctor or care taker with the help of WSN network. Then the care taker can take immediate action. Corona virus is compelling a prompt re-examination and encounter with how to deal with such episodes. Innovation and our advanced society have on a very basic level empowered telemedicine, and now it is show time for such new modalities of "computerized wellbeing" to perform and succeed.

---

[1] Lavanya Dhanesh, Department of Electrical and Electronics Engineering, Panimalar Institute of Technology, Chennai, TN, India, Email: lavanyadhanesh2007@gmail.com.

2. Existing System

a. Problem 1

In the existing system, health parameters of the patients cannot be monitoredremotely.

b. Problem 2

With the increase in number of patients, it is hard to screen them exclusively. And more specialists and medical attendants need to remain in the zone all the time to deal with the patients.

c. Problem 3

In the case that the patient under isolation leaves the zone, the information can't be passed.

d. Problem 4

This framework doesn't have any alert to the patient in isolation, as for any help, they cannot demand for it. This may prompt extreme issues.

3. Proposed Method

- Doctors get instant alert using an automatic health monitoring system.
- Reduces the time consumption of checking all the patients.
- Doctors can monitor the patients remotely without the risk of infection.
- In case if the patient under isolation leaves the conceded zone, the information will be passed to care taker and doctors using WSN.
- When the food is received in the right time a button is pressed by the patient for affirmation reason, where the time is known by RTC.

4. Block Diagram

Our Proposed system performs remote monitoring of patient health parameters using IOT and panic alert is sent to doctors as well as family members. To forestall the spreading of illness by giving programmed observing pack to every tolerant that will get the specialists far from the patient. In our Proposed system, Emergency help button is provided if any oddity detected which helps to send an alert through IOT remotely. The block diagram for the Patient section is depicted in Figure 1 and monitoring section is depicted in Figure 2.
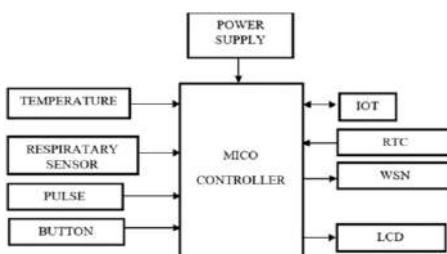
Figure 1. Patient Section   **Figure 2.** Monitoring Section
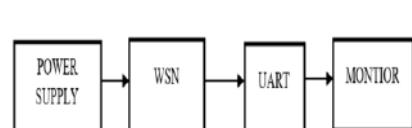
The Zigbee is utilized in our proposed system, for controlling and checking the patient conditions from outside. Arduino Uno is a microcontroller board subject to the ATmega328P is used in our prototype model [3-5]. The program coding is written for various operations to be done in our proposed system using Embedded C and is being burned in the Arduino chip. We employ LM35 temperature sensors in our project which helps to yield voltage comparing with the standard reference temperature [6-8]. The Respiration Sensor is used in our proposed system, to screen stomach or thoracical breathing in biofeedback applications.

The Respiration Sensor can be worn over dress which may be of 1 or 2 layers of pieces of clothing between the sensor and the skin. The Respiration Sensor is ordinarily situated in the stomach area, with the central piece of the sensor. The sensor should be set tight enough to prevent inadequacy of pressing factor. Pulse Sensor is a masterminded fitting for Arduino which we employ in our system.

The sensor cuts onto a fingertip or ear tendon and connects directly into Arduino with some jumper joins [9-11]. On the front there is a round opening, which is the place where the LED exudes from the back. The LED focuses light into the fingertip, ear tendon or other fine tissue, and sensor examines the light that weaves back. The Real Time Clock (RTC)gives specific time and date which will be useful for monitoring the patient and for giving food, medicines at proper time without any delay and thereby informed the doctors and care taker[12-15].The circuit diagram of our proposed system with the sensors is shown in the Figure 3.
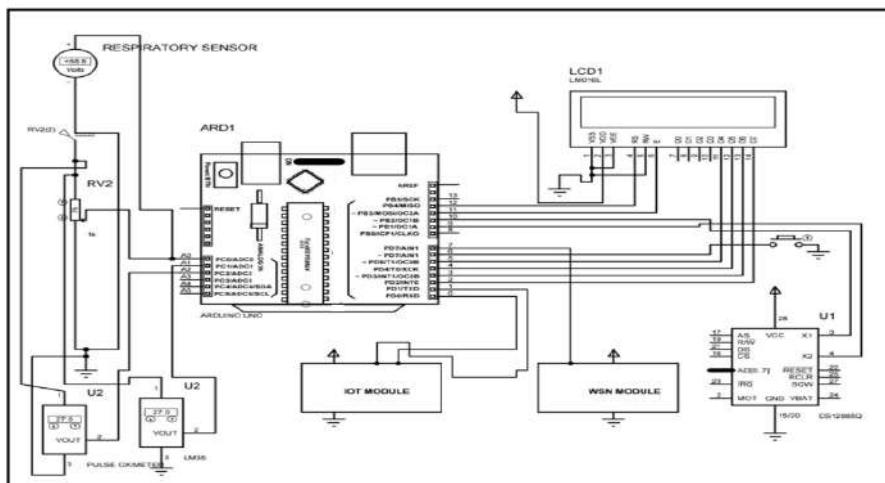
Figure 3. Circuit connections of the Proposed System

5. Results and Discussions

Our device monitors the breathing level, temperature and heart beat level of the COVID patientsand the controller checks for the abnormal value continuously.If any abnormally detected, alert is sent to the doctor or care taker with the help of WSN network for taking immediate action. The Prototype model of our proposed system is shown in the Figure 4 and the output programming results of the Patient conditions is shown in the Figure 5.
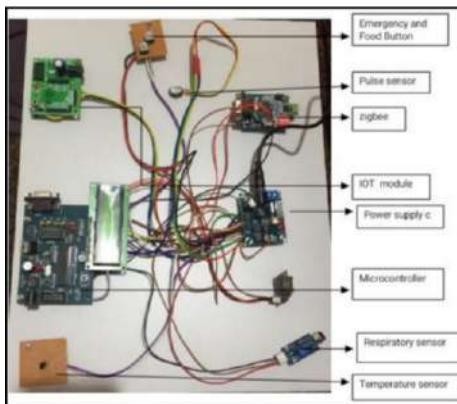
Figure 4. Prototype Model of Our Proposed System
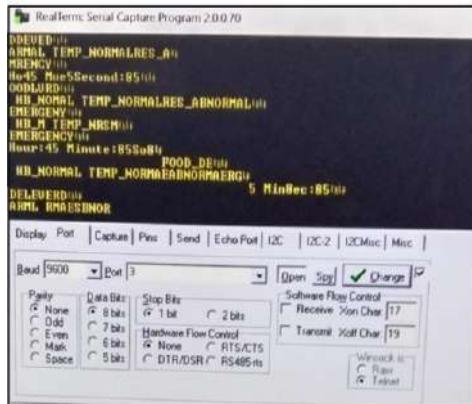
Figure 5. Output Programming Results

6. Conclusion

The current work was focused for vast major part around making life more worth for the people who affected by corona and subsequently kept in isolation. This new system has been made to diminish crafted by trained professionals and moreover the possibilities of ailment spreading. The structure has a twofold impact of both noticing and invigorating the checked data to the subject matter experts or to the regulators appropriated in the facilities in corona ward during the time of the pandemic situation.

References

[1]   S. Marathe, D. Zeeshan, T. Thomas, and S. Vidhya, "A Wireless Patient Monitoring System using Integrated ECG module, Pulse Oximeter, Blood Pressure and Temperature Sensor," Proc. - Int. Conf. Vis. Towar. Emerg. Trends Commun. Networking, ViTECoN 2019, pp. 1–4, 2019, doi: 10.1109/ViTECoN.2019.8899541.

[2]   S. Nubenthan and K. Ravichelvan, "A wireless continuous patient monitoring system for dengue; Wi-Mon," Proc. 2017 Int. Conf. Wirel. Commun. Signal Process. Networking, WiSPNET 2017, vol. 2018-Janua, pp. 2201–2205, 2018, doi:10.1109/WiSPNET.2017.8300150.

[3]   P. W. Digarse and S. L. Patil, "Arduino UNO and GSM based wireless health monitoring system for patients," Proc. 2017 Int. Conf. Intell. Comput. Control Syst. ICICCS 2017, vol. 2018-Janua, pp. 583–588, 2017, doi: 10.1109/ICCONS.2017.8250529.

[4]   Dr.Lavanya Dhanesh , "Automated Intimation to Ambulance for Emergency Medical Service and Traffic Clearance Using Internet of Things (IOT), in the International Journal for Research in Applied Science & Engineering Technology (IJRASET) , ISSN: 2321-9653, Volume: 6, Issue: 3, Mar-2018.

[5]   Dr.Lavanya Dhanesh, "An Smart Load Management System For Homes Using IOT", in the International Journal of Advance Research in Engineering, Science & Technology, e-ISSN: 2393-9877, Volume: 5, Issue: 3, Mar-2018.

[6]   Lavanya Dhanesh, Dr.P.Murugesan, "Smart Scheduling of the Real-Time Tasks Using the Cyclic Priority Preemptive Pipeline Scheduling Algorithm" in the International Journal named "Journal of Computational and Theoretical Nano Science" ISSN 1546-1955 Volume 14,Number 3, pp.1-8.

[7]   Dr.Lavanya Dhanesh, Dr.S.Deepa , " Enhanced and Energy-Efficient Program Scheduling for Heterogeneous Multi-Core Processors System", Lecture notes in Electrical Engineering ,(2020) 665, pp 737-747..

[8]   Lavanya Dhanesh, Dr.P.Murugesan (2018), "A Novel Approach in Scheduling Of the Real- Time Tasks In Heterogeneous Multicore Processor with Fuzzy Logic" in the International Journal named

"International Journal of Power Electronics and Drive System (IJPEDS)" ISSN: 2088-8694, Vol. 9, No. 1, pp. 80-88.

[9]   Dr.Lavanya Dhanesh , "Implementation of IOT Based Self Learning Home Energy Management System", in the International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, ISSN 2278-8875, Vol. 8, Issue 3, March 2019.

[10] Lavanya Dhanesh, Dr.P.Murugesan (2015), "Power Saving of The CPU By Improving The Performance of The Real - Time   System Kernal Using The PSCPPTS Algorithm" in the International Journal named "International Journal of Applied Engineering Research"ISSN 0973-4562 Volume 10, Number 5 (2015) pp. 12465-12473.

[11] Lavanya Dhanesh, Dr.P.Murugesan (2015) , "Analysing the WCET by implementing Cyclic Priority Pre- Emptive Task   Scheduling algorithm" in the International Journal named "Far East Journal Of Mathematical Sciences" ISSN 0972-0871 Volume 97,Number 6, 2015 pp. 667-688 .

[12] G. J. Bharat Kumar, "Internet of Things (IoT) and Cloud Computing based Persistent Vegetative State Patient Monitoring System: A remote Assessment andManagement," Proc. Int. Conf. Comput. Tech. Electron. Mech. Syst. CTEMS 2018, pp. 301–305, 2018, doi:10.1109/CTEMS.2018.8769175.

[13] T. K. Ramesh and C. V. Giriraja, "Wirelesss Sensor network protocol for patient monitoring system," 2017 Int. Conf. Comput. Commun. Informatics, ICCCI 2017, pp. 5–8, 2017, doi: 10.1109/ICCCI.2017.8117798.

[14] S. Shaikh, D. Waghole, P. Kumbhar, V. Kotkar, and P. Awaghade, "Patient monitoring system using IoT," 2017 Int. Conf. Big Data, IoT Data Sci. BID 2017, vol. 2018- January, pp. 177–181, 2018, doi:10.1109/BID.2017.8336594.

[15] Vengatesan, D. K., Naik, D. R., Ramkumar, M., & Bhaskar, T. Review On Cost Optimization And Dynamic Replication Methodologies In Cloud Data Centers. Journal of Advanced Research in Dynamical and Control Systems, 9.

[16] S. P. McGrath, I. M. Perreard, M. D. Garland,K. A. Converse, and T. A. Mackenzie,"Improving Patient Safety and Clinician Workflow in the General Care Setting With Enhanced Surveillance Monitoring," IEEE J. Biomed. Heal. Informatics, vol. 23, no. 2, pp. 857–866, 2019, doi: 10.1109/ JBHI.2018.2834863.

Regenerative and LoRa Based Trooper Monitoring System for Armed Forces

Hema R [a,1], Dathathreya P [a], Athitya V [a], Anumitha B [a]
aDepartment of Electronics and Communication Engineering, Easwari Engineering College, Chennai, TN, India

Abstract. Communication between soldier at border line is crucial. Existing system used for communication between soldiers at border line in military consumes a lot of power. The greatest difficulties in Indian armed forces operation is the Soldiers are not able to do transmission of messages with headquarters base station controller in case of emergency or when needed any help. Also, the current status and location of the soldiers cannot be detected with this system. The proposed methodology gives us Long Range (LoRa) based medical supervision and emplacement trailing and tracking system for soldiers. This type of advanced design can be mounted on the soldier's shoe to ensure their safety. In case of death of the soldier, the controller intimates to the camp office control along with soldier's location. The proposed system includes sensors, GPS, and transmission modules, as well as miniaturized wearable physiological equipment. Hence, it is possible to implement a low-cost mechanism to provide needed help in the battlefield.

Keywords: LoRaWAN, Piezoelectric Sensor, Wireless Sensor Network, Power generation

1. Introduction

The army administrator is not able to send any medical auxiliary to provide aid in case of poor health condition of soldiers. The existing method also does not monitor the health condition of the soldiers which is more important at war time. Thus, we move to some advanced technology which focuses on the safety of soldiers and also helps the army [1]. To overcome the disadvantages mentioned above we propose an advanced prototype design in order to provide safety for soldiers. Our proposed methodology uses piezoelectric sensor and communication using Lora. It is essential for the army headquarters controller station to know the emplacement as well as the medical condition of each individual armed force when they cross the allotted boundary. The portable remote soldier unit consists of micro controller, GPS, temperature sensor and heart beat sensor which monitor the health condition and location of soldiers [2]. If there present any abnormal condition the controller will intimate about the health parameters along with live location with the help of LORA communication network to the control room administrator along with a buzzer alert. This will be very useful for the control room to take needed action in critical situation [3].

---

[1]Hema. R, Department of Electronics and Communication Engineering, Easwari Engineering College, Chennai, TN, India; Email: hemaprabakar18@gmail.com.

Each soldier's shoe will have a unique ID which already contains all medical records of the soldiers. This helps the medical auxiliary to provide suitable medical treatment based on soldier's health condition [4]. It also has a regenerative charging unit to power the shoes by using piezoelectric, while walking

2. LoRaWAN

Long Rang is abbreviated as LoRa which is widely used in wireless communication where transmission of signal takes place at low power. LoRa is a least short power broad area network technology. The working principal of LoRa is a SSMA technique which is derived from from the CSMA technique. Here CSMA means chirp spread spectrum modulation and SSMA means spread spectrum modulation method. LoRa facilitates extensive coverage of transmission and communication of signals at economically least power consumption. Other technologies and protocols such as Long-Range Wide Area Network which is abbreviated as LoRa WAN comprise the upper layers whereas LoRa technology comprises physical layer of the overall LoRa architecture. During transmission or communication of signals, values between 0.3 kbps and 27kbps is the baud rates which can be accomplished by the usage of LoRa technology. The one of the factors which determines this baud rate is spreading factor. Its area of coverage is broad area and LoRa transmits data at   lower data rate. SemTech is a company that owns a proprietary radio modulation technology named LoRa.LoRa comprises of physical layer in LoRa protocol stack. The operation of LoRa takes place in the unlicensed ISM band. This provides an advantage which is the user does not require any license for transmission of data using the LoRa technology [5].
 In India, the original frequency range of LoRa is 868-867 Megahertz. This might vary from one national region to another national region; moreover, operation of LoRa takes place over ISM band. It is very economical to be used by armed forces as it consumes rudimentary power for transmission of signals and for communication purpose at a low data rate and at higher coverage area [6] .This helps the individual armed forces to avoid the usage of cellular network which not only consumes a lot of power for transmission but also the rate of drainage of stored charge in battery is more compared to LoRa network at the cost of sacrificing higher data rate which is absent in our proposed methodology as we used LoRa WAN. Its coverage area as mentioned is in range of kilometers [7]. Transmission is bidirectional in nature with transmission and reception of signals that are done simultaneously between armed force section and the base station section .Its extensible range of transmission makes it suitable to be used at armed forces and military land based industry .It uses gateway nodes which acts as a intermediary nodes for transmission at very long distance. Signal to Noise ratio for transmission of signal is very low compared to the mobile cellular network [8]. It enables transmission and communication of signals at very low data rate covering a very large coverage area. **Experimental setup:**
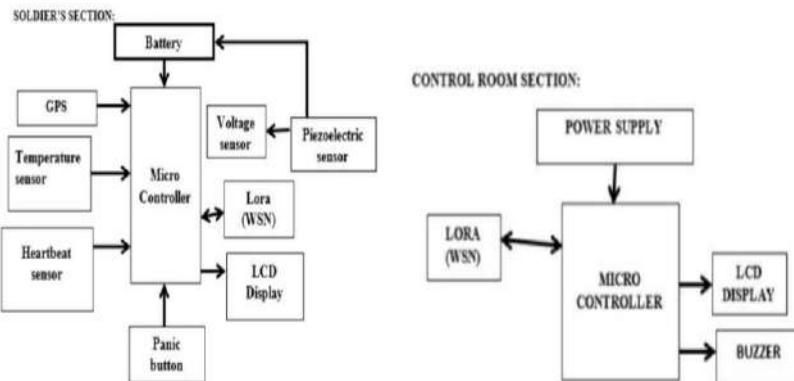
Figure 1. Block diagram of soldier section and control room section

As shown in Figure 1, Piezoelectric sensor unit is utilized by the rechargeable battery for charging purpose and this unit works on the principle of piezoelectric sensor. It generates electric voltage upon the movement of soldier during walking that is setup beneath the soldier shoe. The electric voltage generated varies upon two factors. First factor is duration for which the soldier is on movement that is total time period for which the soldier is walking. Second factor is the amount of pressure that the soldier is exerting on the setup beneath the shoe. The electric voltage generated is directly to the aforementioned two factors. Voltage sensor display that voltage generated by the piezoelectric setup. The voltage generated from piezoelectric is supplied to a rechargeable battery. The entire setup is supplied using rechargeable battery. Microcontrollers is used for processing, controlling all the signals from the different sensor of the setup. It also generates signal for enabling the LCD display and LoRa node for communication with the receiver station. Heartbeat sensor used to measure the pulse rate of the wearer and in case of any threatening abnormalities found in the values of pulse rate, it will enable the microcontroller to trigger the Lora WAN to send a alert message to headquarters situated at the base station section. Temperature sensor is utilized to find out the body temperature of the wearer, here the soldier or armed forces individual represents the wearer. PS unit of the soldier section is used by the controller room section which is also known as the headquarters situated at base station is used for tracking the live emplacement (location) of the wearer

At life threatening situation, the wearer at border line can press the panic button and enables the LoRaWAN to send a signal using microcontroller. Thereafter, the transmitted signal is received by the Base station and this will in turn trigger the buzzer located at the base station and alert the base station. Soldiers or the wearers' temperature is abnormal from the standard threshold then a alert message is sent to the receiver station. On the soldier section, a Liquid Crystal Display unit is used to show the wearer's heartbeat, battery voltage range, and temperature. LCDs perform the same functions as CRTs. Two separate LCD are used for this entire setup one on soldier section and another one is mounted on receiver section. In critical situation at border line, the wearer can trigger a panic button which activates or enable the buzzer at the receiver station. When the soldier station's panic button is pressed, the buzzer using LoRa is automatically activated.

3. Result and discussion

Our proposed methodology consists of health monitoring setup which monitors the health conditions of the wearer and the entire setup is power by a rechargeable battery which is connected to piezoelectric sensor. Voltage generated by piezoelectric sensor depends upon two factors duration for which the wearer is walking and depends upon the weight that the wearer (soldier) exerts on the piezoelectric sensor as this sensor is placed beneath the foot of the wearer so that the entire body weight is exerted. Voltage generated by piezoelectric sensor which is subjected to pressure is measured by using multimeter as shown in Figure 2 and the soldier section and receiver section as shown in Figure 3.

Figure 2. measurement of voltage generated at piezoelectric setup using multimeter.
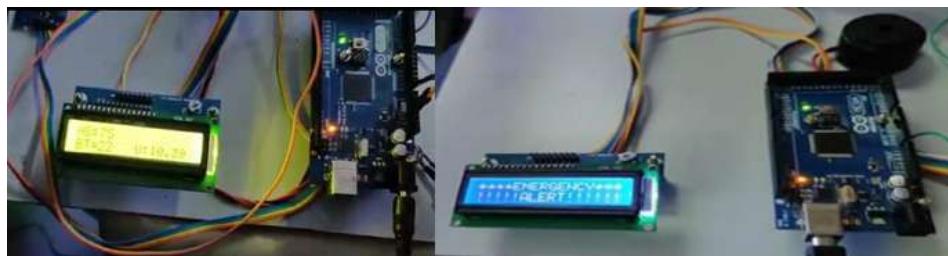
Figure 3. Soldier section and Receiver Section
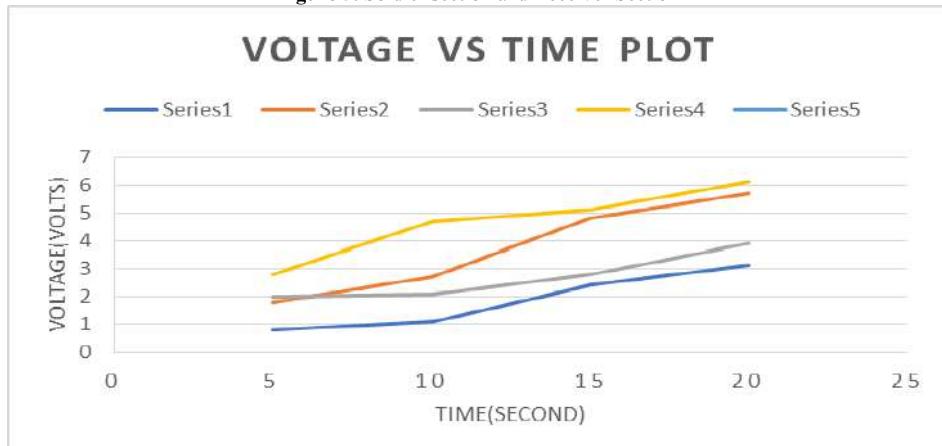
Figure 4. Voltage vs time plot (Note:series1-45kg,series2-50kg,series3-55kg,series4-60kg)
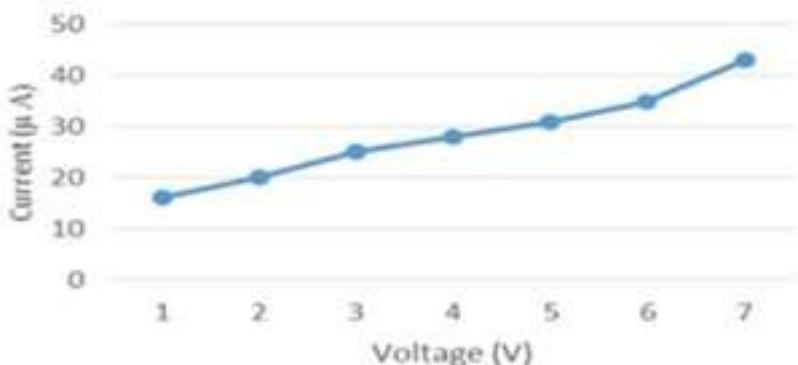
Figure 4 represents Voltage with respect to time plot and figure 5 represents VI characteristics measured.

4. Conclusion

Our proposed methodology consists of cheap, economical setup wearer by the soldier for medical supervising of that particular soldier health condition who is located at border line from a place is located at the base station headquarters which represents the controller room using the LoRa technology and wireless sensor networks. Our proposed methodology avoids the wastage of the power generated .Low Signal to noise ratio is thus obtained by using this module at the same time enabling a long range transmission and communication of signal   .This avoids the usage of the cellular network which transmits data at higher data rate .Our proposed methodology uses low power to transmits data signal or message at low data rate for large distance .Health monitoring consists of supervising body temperature ,heartbeat of the individual who wears this equipment .GPS unit is used to find out the emplacement   of the wearer .This makes a great change in the armed forces domain by making the setup economical and cost effective which can be used for a long run as the battery is rechargeable .This is used to reduce the fatalities that happen at the border line due to lack of communication at life threatening situation of soldier to the base station.

References

[1]   Zourmand, A., Sheng, N. W., Hing, A. L. K., & Rehman, M. A. (2019). Human counting and indoor positioning system using WiFi technology. In *Proceedings - 2018 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2018* (pp. 142–147). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/I2CACIS.2018.8603690
[2]   Choi, W., Chang, Y. S., Jung, Y., & Song, J. (2018). Low-power LORa signal-based outdoor positioning using fingerprint algorithm. *ISPRS International Journal of Geo-Information*, 7(11). https://doi.org/10.3390/ijgi7110440
[3]   Sinha, R. S., Wei, Y., & Hwang, S. H. (2017, March 1). A survey on LPWA technology: LoRa and NB-IoT. *ICT Express*. Korean Institute of Communications Information Sciences. https://doi.org/10.1016/j.icte.2017.03.004
[4]   LoRa Alliance. (2015). White paper: A technical overview of LoRa and LoRaWAN. *LoRaWAN- What Is It?*, (November).

[5]   N. Sornin M. Luis T. Eirich T. Kramp and O. Hersent, Lo-rawan specification, LoRa Alliance Technical Report, 2015

[6]   C. Jain, G. V. S. Sashank, V. N and S. Markkandan, "Low-cost BLE based Indoor Localization using RSSI Fingerprinting and Machine Learning," *Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET),* 2021, pp. 363-367, https://doi.org/10.1109/WiSPNET51692.2021.9419388

[7]   Manuel, M. P., & Daimi, K. (2021). Implementing cryptography in LoRa based communication devices for unmanned ground vehicle applications. *SN Applied Sciences*, *3*(4). https://doi.org/10.1007/s42452-021-04377-y

[8]   Ganesh Babu R., Elangovan K., Maurya S., Karthika P. (2021) Multimedia Security and Privacy on Real-Time Behavioral Monitoring in Machine Learning IoT Application Using Big Data Analytics. In: Kumar R., Sharma R., Pattnaik P.K. (eds) Multimedia Technologies in the Internet of Things Environment. Studies in Big Data, vol 79. Springer, Singapore. https://doi.org/10.1007/978-981-15-7965-3_9

Development and Evaluation of an Early Detection of COVID-19 Pneumonia Using CT Images

D. Eeswar Samhithan [a,1], C. Abarna [b], S. Anupama [b], S. Devi [b]

[a] *Assistant Professor,* [b] *UG Student, Department of Electronics and Communication Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, TN, India*

Abstract. In December 2019, an epidemic of a novel corona virus disease was indentified in Wuhan, China, which quickly spread across the world as a pandemic. This terrible virus had caused respiratory failure and alveolar damage in all, resulting in complete death. The use of a reverse transcription polymerase laboratory test to predict this virus results in a high rate of false positives and a significant time delay. As a result, chest CT images have become a valuable diagnostic method for the Covid-19. In this study, we propose using machine learning to detect and identify the corona, resulting in a timely and accurate report.

Keywords. Covid-19, CT images, Support vector machine, machine-learning model.

1. Introduction

Corona virus disease (COVID-19) is a recently discovered corona virus-related respiratory disease. A digital image is an image composed of picture elements that are output from Corona virus disease (COVID-19). Fever, dry hacking, and sleepiness are common side effects of this infection [1]. A throbbing headache, nasal clog, migraine, sore throat, and a loss of taste and smell called Anosmia are some of the symptoms. It also triggers a disorder known as Parosmia, in which one's favourite scents become revolting. The infection has spread to a large area, and cases and deaths seem to be increasing in a step-by-step manner [2]. With the help of an AI method named Prophet model, which was developed and presented by Facebook, this investigation aims to expose the anticipation of various boundaries associated with this infection, such as the increase in new events, recoveries, and deaths every day across the world. Validation of the diagnosis can be done with a reverse transcription polymerase chain reaction examination (RT PCR) [3]. While RT-PCR is the gold standard for diagnosis, verifying COVID-19 patients takes time, and it may be difficult to diagnose and treat presumptive patients early due to high false-negative rates and low sensitivities.

---

[1]D.Eeswar Samhithan, Assistant Professor, Department of Electronics and Communication Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai Email: deeswar@velhightech.com.

2. Methodology

We tend to use an SVM (Support Vector Machine) algorithm to detect COVID 19 using machine learning. In this method, Chest CT scan tool are used to diagnose and screen the affected area of chest. CT scan shares the imaginary characteristics between the Covid-19 affected lungs or with other lungs diseases (Pneumonia or non-pneumonia) [4]. By machine learning technology, it can be monitor, screen and predict the affected lung area of patients. Accuracy of the results are very high and time delay is very less compared to the existing system such as RT PCR [5]. The software and hardware section is shown in Figure 1.
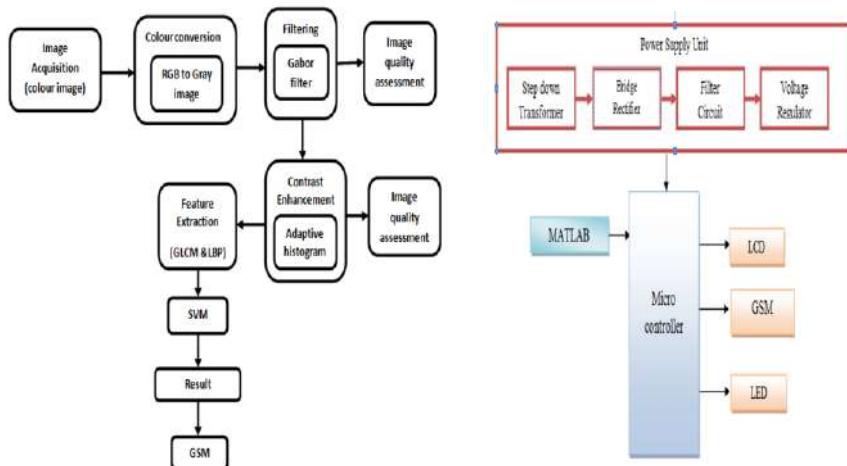
Figure 1. Software and Hardware section

3. Circuit Design

3.1 Software section

Image acquisition contains the input CT image of the patient which is in the form of color image. Using color conversion technology going to convert the color image (RGB which has three layer) into grey image(single layer image)[6]. Filtering is made for the texture analysis of the image by using the Gabor filter to extract the unnecessary frequency content. Image Quality Assessment is used check the accuracy of the image quality. After filtering, contrast enhancement is made for improving the local contrast and to over amplify the noise by using the adaptive histogram [7]. Feature extraction is the method used to extract the second order of statistical texture feature by GLCM (Grey Level Concurrence Matrix) and transforms the image into an array or integer by LBP (Local Binary Pattern) [8].

3.1.1 Support Vector Machine (SVM)

In machine learning techniques, SVM is one of the algorithms. It has a broad data set and outperforms other algorithms in terms of performance. SVM is required for small sample data by make use of training samples to observe a support vector and to obtain a good prediction and classifications. By making use of this algorithm, we can compare the input CT images with the data set which we already collected and stored from the medical sample (Chest CT scans) of COVID patients.

3.1.2 Global System for Mobile Communications

Global System for Mobile Communications (GSM) is a used as a mobile phone or modem to communicate over the network. In this necessary of Sim card is must to be operated with subscribed network. Here we are using GSM for SMS service. After completion of diagnosis and prediction of the COVID 19. The report can be send through as email or as SMS.

3.2 Hardware Section

Interfacing the Arduino uno and GSM modem is discussed in this section. The assembly of the modem and te Arduino. We test the GSM modem's functionality after the Arduino code has been burned into the Arduino uno module. The GSM modem sensor is interfaced on the Arduino Uno module, and the required codes are written.

3.2.1 Power supply unit

It contains such as bridge rectifier, filter circuit, step down transformer and voltage regulator. In this transformer is able to supply current to the circuit, bridge rectifier is to convert the AC to DC current, filters are used to rectifies the remaining AC pluses, and also voltage regulator are available for obtaining stable voltage and current.

3.2.2 Arduino uno

It is majorly used, because it is compatible with all Operating System like Windows, linux etc. It's an ATmega328-based microcontroller module. It will simply connect to the System through USB. And the UNO is the latest USB Arduino board.

3.2.3 UART (Universal Asynchronous Receiver/ Transmitter)

It's a microchip with software that controls a machine interface to the computer it's connected to. It converts bytes received from the device over a parallel circuit into single serial bit streams and vice versa.

3.2.4 LCD (Liquid Crystal Display)

It is a flat panel display used to display the result such as Positive or negative of that disease.

3.2.5 LED (Light Emitting Diode)

It is a process of emitting light in response of electric field. It is used to indicate the Positive/Negative by blowing the led with different colors such as green and red. GSM is used to send the report to the patient who had taken the test for COVID-19.

4. Result
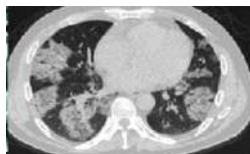
Figure 2. Load image
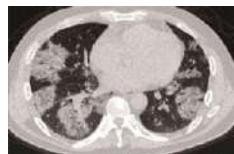
Figure 3. Grey image)

The load image is given to the data memory, which is RBG format CT image of three planes is shown in Figure 2 and Figure 3. The RGB has hue and saturation information. The three layer image is converted into a single layer gray image. Since there is a hue and saturation information in load image e are converting it into gray scale image. The gray scale image is useful for morphological operation and image segmentation.

Figure 4. Filtered image of Gabor filter
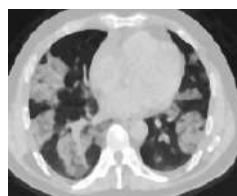
Figure 5. Enhancement image of Gabor filter

The above Figure 4 and Figure 5 are filtered and enhancement images by Gabor filtering technique. The filter gives localized spatial information of a given image. The filter is used for feature extraction, texture analysis, and stereo disparity estimation.
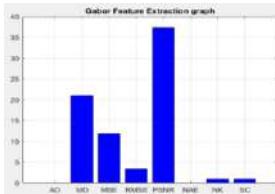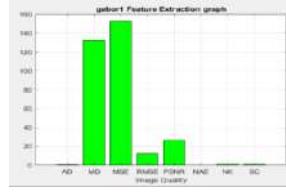
Figure 6. Extraction graph of filtered image Figure

Figure 7. Extraction graph enhancement image

The above two graphs in Figure 6 and figure 7 shows the Gabor feature extraction of the filtered image.
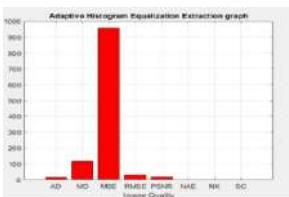
Figure 8. Equalization extraction graph

Figure 9. Positive output image

The enhancement of image filtration is given above in Figure 8. This graph shows the enhancement of the filtration of Gabor filter which make the analysis of image in better way. The positive output image is final output of the process is shown in Figure 10. The red color indication shows the affected area of lungs by corona virus. The information will be sent to patient via mail.

The result of affected patient mail and LCD display is shown in the Figure 11.

Figure 11. Result of affected patient mail and LCD display

5. Conclusion

Our weakly supervised machine learning algorithm achieved strong covid-19 detection efficiency without annotating the lesions in CT volumes for preparation. This algorithm has a lot of clinical potential in terms of accurate rapid covid-19 diagnosis, which is beneficial to frontline medical staff and also essential for global infection management. This analysis has several drawbacks. They could be enhanced in terms of network architecture and training. Cross-center validations were not conducted because the data in this analysis originated from a single hospital.

References

[1] X. Xie, X. (2020). Chest CT for Typical 2019-nCoV Pneumonia: Relationship to Negative RT-PCR Testing. *Journal of Clinical Microbiology*, *58*(April), 1–5. Retrieved from https://doi.org/10.1016/j.jcv.2020.104384%0Ahttp://dx.doi.org/10.1016/S2213-2600(20)30315%0Ahttps://doi.org/10.1016/j.rbmo.2020.06.001%0Ahttps://doi.org/10.1016/j.mayocp.2020.04.004%0Ahttps://doi.org/10.1016/j.cmi.2020.06.019%0Ahttps://doi.org/10.1080/1

[2] Wang, L., Lin, Z. Q., & Wong, A. (2020). COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-76550-z

[3] Apostolopoulos, I. D., & Mpesiana, T. A. (2020). Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, *43*(2), 635–640. https://doi.org/10.1007/s13246-020-00865-4

[4] Hemdan, E.E., Shouman, M.A., & Karar, M. (2020). COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images. *ArXiv, abs/2003.11055*.

[5] Farooq, M., & Hafeez, A. (2020). COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs. *ArXiv, abs/2003.14395*.

[6] Markkandan, S., Narayanan, L., Robert Theivadas, J., & Suresh, P. (2021). Edge based interpolation with refinement algorithm using EDGE strength filter for digital camera images. Turkish Journal of Physiotherapy and Rehabilitation, 32(2), 981–993.

[7] Ganesh Babu R., Elangovan K., Maurya S., Karthika P. (2021) Multimedia Security and Privacy on Real-Time Behavioral Monitoring in Machine Learning IoT Application Using Big Data Analytics. In: Kumar R., Sharma R., Pattnaik P.K. (eds) Multimedia Technologies in the Internet of Things Environment. Studies in Big Data, vol 79. Springer, Singapore. https://doi.org/10.1007/978-981-15-7965-3_9

[8] VENGATESAN, K. PULMONOLOGICAL DISORDERS CAUSED DUE TO COVID-19 PANDEMIC. Turkish Journal of Physiotherapy and Rehabilitation, 32, 2.

[9] R. G. Babu, V. V. Nathan, J. Bino, C. Amali and S. Ganesh, "IoT Security Enhancement with Automated Identification Device using IOT SENTINEL," *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2021, pp. 518-523, https://doi.org/10.1109/Confluence51648.2021.9377165

A GSM Based Assistive Device for Blind, Deaf and Dumb

Lavanya Dhanesh [a,1], Monica S [b], Revathi R [b], Suganya S [b], Swetha lakshmi K N [b]

[a]Associate Professor, [b]UG Student, Department of Electrical and Electronics Engineering, Panimalar Institute of Technology, Chennai, TN, India

Abstract. This paper tries to overcome the shortcomings of the recent technology that fails to enhance the communication between physically disabled people by designing an assistive device. This device uses a GSM modem with a SIM card and no smartphones are needed which makes the device affordable. Here, the sender sends the message to the phone number of the disabled person and he/she receives the SMS that gets converted to text and voice message and vibrations to Braille pad using a microcontroller which is easily readable by the disabled person.

Keywords. GSM modem, text and voice message, microcontroller

1. Introduction

In today's world, digital technologies play a major role in communication by bridging people anytime and anywhere across the world. Digital technology includes mobile phones, e-mails, the internet, social media, etc., which makes communication between normal people simpler [1]. But the physically impaired people particularly blind, deaf and dumb are lacking the ease of using modern technology. Around 285 million individuals are encountered to be visually disabled worldwide of which 39 million are visually impaired and 246 are said to have low vision and about 9.1 billion individuals have a hearing defect and are dumb. Despite the advancements in technology, a device that comprises all the necessities for better communication between the three categories of disabled people mentioned above is still lacking a proper structure of communication [2-4]. Hence, we have designed a device that makes communication easier and affordable for disabled people. This device mainly focuses on SMS technology using GSM [5-6]. Global System for Mobile communication is abbreviated as GSM and it is used with mobile phones as a digital mobile network.

The output of this device is in the form of braille language, text message, and voice message which is useful to the blind, deaf and dumb respectively. Braille is a tangible reading and writing system used by visually disabled people [7-9]. It is generally written on imposed paper. Braille was created by Louis Braille. Braille pad contains several cells and each cell has six elevated dots which is used to represent various characters. The arrangement and number of these dots represent different characters, numbers, and symbols. On the other hand, an LCD and the speaker convey the text and voice message to deaf and dumb people.

---

[1]Lavanya Dhanesh, Department of Electrical and Electronics Engineering, Panimalar Institute of Technology, Chennai, TN, India; Email: lavanyadhanesh2007@gmail.com.

2. Literature Review

Choudhary Tanay, Kulkarni Saurabh, Reddy Pradyumna; 16 April 2015; "A Braille-based mobile communication and translation glove for deaf-blind people", This paper is a novel methodology for the interaction between blind individuals independently. The paper defines the use of a Bluetooth device connected to a phone while a low-power microcontroller unit is used in it [1].

Banmare Amar, Chawhan S Ashika, Lanjewar R. Aditi, Nikhade R. Karishma, Siriya P. Meghna; March 2018; "Hardware Based Braille Pad on Mobile Phone",This paper proposes an aim to provide a Microcontroller-based Braille Pad system that provides information to blind people by using Global System for Mobile Communications (GSM), helping the visually disabled to communicate. The details of simulation in Proteus 8 professional software and embedded module have been furnished in this paper [2].

Shraddha R. Ghorpade, Surendra K. Waghamare; 2015; "Full Duplex Communication System for Deaf & Dumb People",This paper proposes a system that converts ordinary person's speech to text and displays the corresponding gesture on the display. In this way, the entire thought is to assemble a gadget that empowers two-route communication between a disabled individual and a typical individual [3].

Suharjito, Ricky Anderson, Fanny Wiryana, Meita Chandra Ariesta, Gede Putra Kusuma; October 2017; "Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output", This paper surveys different communication via gestures acknowledgment approaches and assists the researcher with tracking down the best methodology among them to develop better Sign Language Application Systems later on [4].

Brian Frey, Caleb Southern, Mario Romero; "Mobile Texting for the Visually Impaired" This paper focuses on IOS technology. The iPod model incorporates a plan that helps the user hold the device and position the fingers on the keys. Users hold Braille Touch by both hands and use their fingers in a layout with a one-to-one correspondence to Braille writer [5].

3. Need for the Study

Even though the advancements in the technologies invented many devices for the disabled people, they someway include the usage of smartphones which makes it less affordable. So, there is a need for a device that is more affordable and simpler [10].

4. Proposed System

The proposed system tries to overcome the difficulties by introducing a device that uses no smartphones and makes communication simple with SMS technology. He3re, the sender sends the message to the phone number of the disabled person and he/she receives the SMS that gets converted to text and voice message and vibrations in the Braille pad using a microcontroller. Signals will be sent to the device which will contain vibration motors attached to the Braille dots. The SIM card is kept in a slot through which the text input goes to the microcontroller and a program in

microcontroller then converts the text into Braille characters, voice message and also displayed in text format.

5. System Design

The system contains a GSM modem, Arduino board, ATmega microcontroller, relays, braille pad, speaker, and LCD Display. SMS is received by the GSM modem. Serial Communication Transmission Port (RS-232) is used to communicate between Microcontroller and GSM. SMS goes through a set of code where Microcontroller receives the SMS and vibrates the vibration motors based on text received which is attached to Braille Pad. Also, the microcontroller converts the SMS into a voice message and displays the text in LCD [11-12]. GSM means Global System for Mobile communication. It is a digital mobile telephone communication system. By the Advance GSM module interface in our proposed system, we send short instant messages to particular authorized users according to the application. GSM module is furnished with a SIM that utilizes the mobile service provider and sends SMS to the respective user according to programmed. This is a plug-and-play GSM Modem that is simpler to execute RS232 and TTL serial interface. It is used to send SMS, settle on and get decisions, and do other GSM activities by basic AT orders through a sequential interface from microcontrollers and PCs. The Arduinos are worked around an ATmega microcontroller, where the program of our proposed system is being uploaded and compiles the code written in Embedded "C" to the board. The relay is used to control different devices by turning on and off them whenever necessary. Here, six relays are used for six braille dots. Since the relays have jumpers no voltage regulators are needed and the direct power supply can be given to the relays [13-15]. The Braille pad is employed in our proposed system. It has tiny bumps called raised dots are present which is represented as a braille cell. The number and arrangement of these dots represent different characters, numbers, and symbols. A Braille cell contains six raised dots arranged in two parallel columns each having three dots. The dot positions are recognized by numbers from one to six. There is a Liquid Crystal Display interfaced with the microcontroller to verify the output of our proposed system.

6. Results and Discussions

The Figure 1 signifies the Prototype model of our proposed system. The Braille cell is shown in the Figure 2 and the Braille Reference Chart is shown in the Figure 3. The Braille Reference Chart, maps each cell of the Braille to an ASCII character The ASCII Braille refers to the codes which have to be sent to a Braille Embosser, to correctly print the cells corresponding to the text to be transcribed.
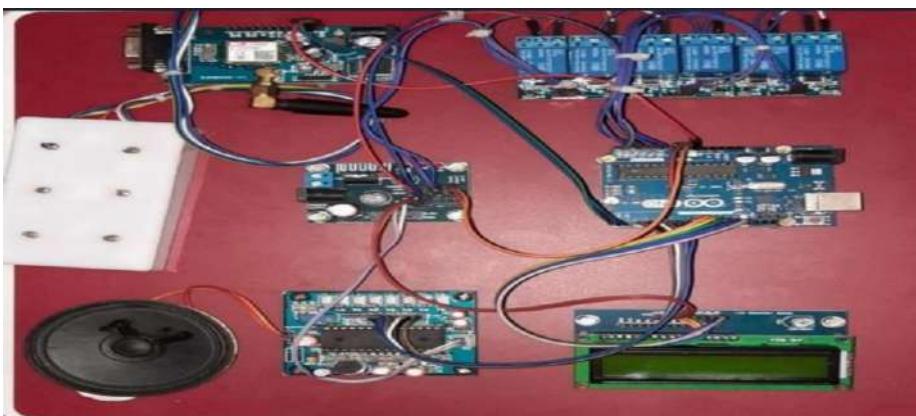
Figure 1. Prototype Model of Our Proposed System
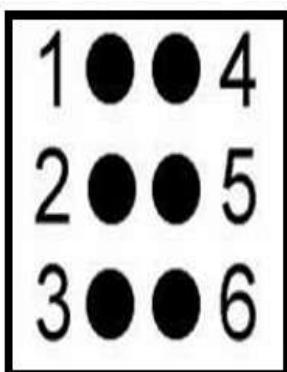
Figure 2. The Braille Cell   **Figure 3. The Braille Reference Chart**

7. Conclusion

In this paper, the proposed device aids disabled people to communicate more simply. The primary objective of this device is to provide an affordable and easily portable mode of communication. This model is a solitary smaller gadget exclusively for the blind, deaf, and dumb people. The primary factor of this model is to work with these individuals and to make them more certain to work autonomously.

References

[1] Miele, J.A., Weara Braille: development of a wireless, virtual, Braille keyboard, in 25th Annu. Int.Technology & Persons with Disabilities Conf. (2010)

[2] Caporusso, N. (2008). A wearable malossi alphabet interface for deafblind people]. In *Proceedings of the Workshop on Advanced Visual Interfaces AVI* (pp. 445–446). https://doi.org/10.1145/1385569.1385655

[3] *Assistive Technology for the Hearing-impaired, Deaf and Deafblind*. (2003). *Assistive Technology for the Hearing-impaired, Deaf and Deafblind*. Springer-Verlag. https://doi.org/10.1007/b97528

[4] Braille Touch: Mobile Texting for the Visually Impaired, Brian Frey, Caleb Southern, Mario Romero: Advance SMS System for Blind Peoples using Braille Pad, Harshal A Dahake and Sunil U Nyati Shraddha R. Ghorpade, Surendra K. Waghamare, "Full DuplexCommunication System for Deaf & Dumb People," International Journal of Emerging Technology and Advanced Engineering (IJETAE), Volume 5, Issue 5, May 2015, ISSN 2250-2459.

[5] Er. Aditi Kalsh, Dr. N.S. Garewal, "Sign Language Recognition System," International Journal of Computational Engineering Research (IJCER), Volume 03, Issue 6, June 2013.

[6] Bhame, V., Sreemathy, R., & Dhumal, H. (2014). Vision based hand gesture recognition using eccentric approach for human computer interaction. In *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014* (pp. 949–953). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICACCI.2014.6968545

[7] Schmid, C., Mohr, R., & Bauckhage, C. (2000). Evaluation of interest point detectors. *International Journal of Computer Vision*, *37*(2), 151–172. https://doi.org/10.1023/A:1008199403446

[8] Dhanesh, L. (2018). Automated Intimation to Ambulance for Emergency Medical Service and Traffic Clearance Using Internet of Things (Iot). *International Journal for Research in Applied Science and Engineering Technology*, *6*(3), 2829–2834. https://doi.org/10.22214/ijraset.2018.3455

[9] Dr.Lavanya Dhanesh , "An Smart Load Management System For Homes Using IOT", in the International Journal of Advance Research in Engineering, Science & Technology, e-ISSN: 2393-9877, Volume: 5, Issue: 3, Mar-2018.

[10] Dhanesh, L., & Murugesan, P. (2017). Smart scheduling of the real-time tasks using the cyclic priority preemptive pipeline scheduling algorithm. *Journal of Computational and Theoretical Nanoscience*, *14*(4), 1813–1820. https://doi.org/10.1166/jctn.2017.6510

[11] Dhanesh, L., Deepa, S., Elangovan, P., & Prabhu, S. (2020). Enhanced and Energy-Efficient Program Scheduling for Heterogeneous Multi-Core Processors System. In *Lecture Notes in Electrical Engineering* (Vol. 665, pp. 737–747). Springer. https://doi.org/10.1007/978-981-15-5262-5_55

[12] Dhanesh, L., & Murugesan, P. (2018). A novel approach in scheduling of the real- time tasks in heterogeneous multicore processor with fuzzy logic technique for micro-grid power management. *International Journal of Power Electronics and Drive Systems*, *9*(1), 80–88. https://doi.org/10.11591/ijpeds.v9n1.pp80-88

[13] Dr.Lavanya Dhanesh , "Implementation of IOT Based Self Learning Home Energy Management System", in the International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, ISSN 2278-8875, Vol. 8, Issue 3, March 2019.

[14] Dhanesh, L., & Murugesan, P. (2015). Power saving of the CPU by improving the performance of the real-time system kernal using the PSCPPTS algorithm. *International Journal of Applied Engineering Research*, *10*(5), 12465–12474.

[15] Lavanya Dhanesh, Dr.P.Murugesan (2015) , "Analysing the WCET by implementing Cyclic Priority Pre- Emptive Task   Scheduling algorithm" in the International Journal named "Far East Journal Of Mathematical Sciences" ISSN 0972-0871 Volume 97,Number 6, 2015 pp. 667-688 .

IoT Based Smart Cradle Using PI

KANNAN P [a,1], ALLWİN DEVARAJ [a], B.PRADHEEP T RAJAN [a], SWATHİRA P K [b], SUBHİKSHAA JAYARANİ M [b], SHEBNA V [b]

[a]*Assistant professor, Francis Xavier Engineering College, Tirunelveli, TN, India*
[b]*UG Student, Francis Xavier Engineering College, Tirunelveli, TN, India*

Abstract. In this modern era, parents are busy building their lifestyle, carrier etc. As we know it has now become hard because parents have to take care of their children simultaneously, which paves a lot of work pressure and family pressure especially for women. Health of the child is affected and better care has also reduced. So, in order to handle such situation, we use temperature, Humidity, ultra-sonic Sensor. The conditions of the external atmosphere help to detect increased body temperature, babies voice while crying and their movements while they are continuously moving and also indicates the time for the diaper to be changed. If there are any abnormal activities are observed in baby's atmosphere. An alert message is sent to the parents. In this system a video camera is attached and operated under the microcontroller's instruction and it records a video when the motion sensor detects any continuous movements. The recorded video is broadcasted in a display to the parents which helps them to monitor baby in live. And in addition, this system detects and displays the status of the infant and alert the respective guardian by collecting values from sensor like temperature sensor, ultrasonic sensor, and also the location value from raspberry pi.

Keywords. Temperature, Weight Detection, Ultrasonic, Location Value, Sensor, Raspberry Pi

1. Introduction

In common, to soothe and make the baby sleep cradle is being used. But even guardian must have a close monitoring until the baby falls asleep. The conventional cradle which is not equipped with battery or adapter that automates the cradle automatically [1]. Rural and non-developed area use conventional cradle because of their economic cost. Manpower is required continuously to take care of their child which serves as its major disadvantage and such conventional cradle may also cause discomfort to the baby. So, we are in need for an automated cradle which can take care of the child with a battery or electrical power source. Besides, this automated cradle carries extra features or function which is beneficial for parents [2]. Ample time is not given to take care of our children in this industrous world because in professional life people are not getting enough time for taking care of their children. It is also much expensive and uncomfortable to hire an baby sitter, who can take care of the child [3]. This may also increase the expenses more than a monthly income. Moreover, in this modern life it's even hard for house wives (mummy) for pampering their children when they are not feeling comfortable.

---

[1]Kannan P, Assistant professor, Francis Xavier Engineering College, Tirunelveli, TN, India; Email: kannamuthu1987@gmail.com.

Nurses in maternity units in the hospitals find it very useful automated [4]. The Main objectives of this project is to design and implement a baby care cradle, which monitors baby's movements when detected and wet condition of the baby as well as its body temperature.

- To design a comfortable cradle for baby with the help of PIR sensor which monitor's the baby's movements and its wet condition especially when away from the unhygienic environment.
- To provide an innovative and flexible cradle, this is economical in the market.
- User friendly- simple and with easy understandable user manual.

2. Automatic baby cradle

Metropolitan cities use more cradles. The benefits of this cradle include:

a) It serves as a user friendly; it also reduces the manpower. The cradle can be adjusted in custom as per the user's requirement and need.
b) User can set a particular time for cradle movement and while the user can complete their chores.
c) The most beneficial use of the product is, it uses neonatal and maternity units in hospital as bio-medical product. As this dwindles the labor of the nurses for taking care of infant and pamper them whenever they need.

The development of internet and mobile phones increased the popularity of IOT [5]. Working parents can watch their children sitting in live with the help of their mobile phones which serves as another important concept. The project Smart Cradle is enhanced with video monitoring. Also, it alerts by buzzing a buzzer on phone on following occurrence, first if baby cries beyond the noted time for example when the cradle does not handle the baby and it requires some personal care and also if the cloth of the cradle becomes wet [6]. Automated rotated toy is fixed in the cradle for entertaining the baby lively.

For those parents who are out of reach from the baby can still monitor their infant using a mobile phone. The smart cradle is enhanced with live video streaming feature that helps the parent to monitor their infants [7]. Even when the parent is in the next room, they can still monitor baby's movement so that the parent has an eye on their infant closely. This live video monitoring can also be set to stream automatically when the baby cries exceed beyond the set time with an alert notification. The most important and notable feature of the smart cradle is, it checks on the health condition of the baby [8]. The temperature sensor in the cradle senses for temperature change in the infant. Any abnormal heat change in infant body is alerted by the system using a alert notification. The temperature senses have a close monitoring of the temperature change in baby. It also indicates cradle temperature that helps parents to save them from heat in the cradle.

This is a project representing a baby monitoring system which helps busy parents to take care of their children safely. This system also notes the baby's motion and sound; especially crying and video output of baby's present position which can be displayed on a display monitor so that the mother or another responsible person can watch the baby while away from him or her [9]. Raspberry Pi B+ module controls the hardware of the system and baby's movement is monitored by incorporated ultrasonic

sensor and baby's motion is captured using USB camera. The video which has recorded the sleeping baby is displayed in the display monitor [10]. At last the final output which is the developed hardware is used for detecting the crying sound and the video of the baby. Busy parents can find it convenient and easier for taking care of their baby [11].

3. Proposed System

Figure 1, shows the representation of a system, which is organized and supports the reasoning and behaviour of the system is the architecture description which is also known as formal description. The system consists of temperature sensor, humidity sensor, sound sensor and the camera. All these are connected to the raspberry pi. The message is forwarded to the parent using GSM and the L298N motor driver is used to move the dummy shaft using the dc motor. In this proposed system the high accuracy about the children condition is recorded in the resulted data because it uses both sensors and forecasting clouds, also camera surveillance of the children is managed with the help of Wide Area Network (WAN) and the result can be seen from anywhere from the remote area in the web application.
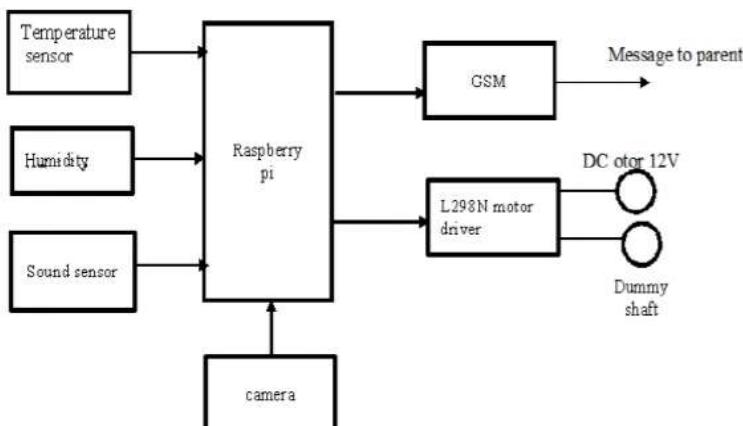
Figure 1. Represents the Architectural Design

4. Result and discussion

The system is based on raspberry pi in which any one of the saved values differs, it alerts the parents. This raspberry pi also gives instruction to the video camera attached with the system and the video will be recorded when the baby's movements is monitored continuously. The monitoring video will be displayed live on the monitor. At present, females have started working in industrialized sectors which in turn affects the child care in the families. Since nowadays managing the cost of living has become difficult, females started working which has affected their children's care.

This baby monitoring system is used to solve this problem. Video output is setup to monitor the sleeping baby. Finally, the output is analysed and the capability is

tested for detecting the motion and the sound of the crying baby is also displayed in the output. This proposed system helps in offering an easier and convenient way for the parents to take care of their children. A baby monitoring system which includes the video camera and microphone without any limitations of coverage is used for sending the data immediately and also to inform parents when there is any emergency. This helps in minimizing the time needed for handling the situation. Babies usually cry when they need food, tired, unwell or their diaper needs to be changed.

The circuit connection is properly completed, that is the ultrasonic sensor, buzzer, temperature sensor, motor driver and 12v dc motor. All these components are connected to the raspberry pi. The code is allowed to run in PC and hence the output is verified on the monitor. The value of ultrasonic sensor increases and decreases in accordance with echo and trigger so in order to monitor this, the buzzer alert is given. The video camera which is attached to the system is connected to the raspberry pi and is working under its instructions and is used for capturing the video when the baby's movements is detected continuously. The power supply is connected and the live video is monitored on PC. The website is created and the result is displayed in laptop. The person who has the IP address can watch the live video streaming.

Figure 2. Cradle setup

Thus, the final cradle setup as per Figure 2, is done and numerous sensors connected to this smart cradle will help in acquiring the accurate value.

5. Conclusion

This system which we have proposed aims at monitoring the vital signs of the baby which includes heartbeats and body temperature with the help of wireless technology and textile sensors that made it comfortable to wear and this is accurate and precise than other sensors. The information which can be transmitted over the internet which provides the remote access has been increased. The camera module which is attached to the cradle helps in finite monitoring of the baby in the finite area and their movements are monitored continuously. The drawbacks of the previous system which includes less user friendly, clumsy, and other factors are overcame in this system.

References

[1] Jabbar, W. A., Shang, H. K., Hamid, S. N. I. S., Almohammedi, A. A., Ramli, R. M., & Ali, M. A. H. (2019). IoT-BBMS: Internet of Things-Based Baby Monitoring System for Smart Cradle. *IEEE Access*, *7*, 93791–93805. https://doi.org/10.1109/ACCESS.2019.2928481

[2] Lazaro, A., Boada, M., Villarino, R., & Girbau, D. (2019). Battery-Less Smart Diaper Based on NFC Technology. *IEEE Sensors Journal*, *19*(22), 10848–10858. https://doi.org/10.1109/JSEN.2019.2933289

[3]   Berrezueta-Guzman, J., Pau, I., Martin-Ruiz, M. L., & Maximo-Bocanegra, N. (2020). Smart-Home Environment to Support Homework Activities for Children. *IEEE Access*, *8*, 160251–160267. https://doi.org/10.1109/ACCESS.2020.3020734

[4]   Gahlot, S., Reddy, S. R. N., & Kumar, D. (2019). Review of smart health monitoring approaches with survey analysis and proposed framework. *IEEE Internet of Things Journal*, *6*(2), 2116–2127. https://doi.org/10.1109/JIOT.2018.2872389

[5]   Hesse, N., Pujades, S., Black, M. J., Arens, M., Hofmann, U. G., & Schroeder, A. S. (2020). Learning and Tracking the 3D Body Shape of Freely Moving Infants from RGB-D sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(10), 2540–2551. https://doi.org/10.1109/TPAMI.2019.2917908

[6]   Gutiérrez, S. T., Fuentes, C. I., & Díaz, M. A. (2020). Introducing SOST: An ultra-low-cost star tracker concept based on a raspberry pi and open-source astronomy software. *IEEE Access*, *8*, 166320–166334. https://doi.org/10.1109/ACCESS.2020.3020048

[7]   Hu, D., Zhang, H., Wu, Z., Wang, F., Wang, L., Smith, J. K., … Shen, D. (2020). Disentangled-Multimodal Adversarial Autoencoder: Application to Infant Age Prediction with Incomplete Multimodal Neuroimages. *IEEE Transactions on Medical Imaging*, *39*(12), 4137–4149. https://doi.org/10.1109/TMI.2020.3013825

[8]   Liu, L., Li, W., Wu, X., & Zhou, B. X. (2019). Infant cry language analysis and recognition: An experimental approach. *IEEE/CAA Journal of Automatica Sinica*, *6*(3), 778–788. https://doi.org/10.1109/JAS.2019.1911435

[9]   Li, C., Pourtaherian, A., Van Onzenoort, L., Tjon, W. E., & De With, P. H. N. (2020). Infant Monitoring System for Real-Time and Remote Discomfort Detection. *IEEE Transactions on Consumer Electronics*, *66*(4), 336–345. https://doi.org/10.1109/TCE.2020.3031359

[10] Jiang, L., & Hassanipour, F. (2020). Bio-Inspired Breastfeeding Simulator (BIBS): A Tool for Studying the Infant Feeding Mechanism. *IEEE Transactions on Biomedical Engineering*, *67*(11), 3242–3252. https://doi.org/10.1109/TBME.2020.2980545

[11] Vengatesan, D. K., Naik, D. R., Ramkumar, M., & Bhaskar, T. Review On Cost Optimization And Dynamic Replication Methodologies In Cloud Data Centers. Journal of Advanced Research in Dynamical and Control Systems, 9.

[12] R. G. Babu, V. V. Nathan, J. Bino, C. Amali and S. Ganesh, "IoT Security Enhancement with Automated Identification Device using IOT SENTINEL," *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2021, pp. 518-523, https://doi.org/10.1109/Confluence51648.2021.9377165

Cascading of Air Quality Detector and Digital Data Transmission with Zero Error in Minimum Duration

Anitha R [a,1], Veerasundaram M [a], Rajarajan S [a], Jaya Prakash E [a], NandhaKishore J [a], Ajay Krishnan S R [a]

[a]*Department of EEE, Sri Sai Ram Institute of Technology, Chennai, TN, India*

Abstract. The paper proposes a system to monitor the environment air quality using an Arduino microcontroller, to enhance the quality IOT Technology is used. The utilization of IoT technology improves the method of monitoring various factors of the environment issues. The MQ-02 gas sensor is used to detect the various gases which are harmful to humans. A Wi-Fi module connects the whole system to the internet and an LCD is employed for the visual Output. This Automated Air management system may be a breakthrough to contribute an answer to the most important threat. The air quality detector overcomes the problems of the highly polluted areas which may be a major issue. This technique has features for the user to monitor the quantity of pollution on their smart devices using the appliance.

Keywords. IOT, Automatic Air management system, Electronic Image.

1. Introduction

Manual Scavenging is the most nauseating thing to try to do, for others, it's the sole thanks to making a living. Many manual scavengers die, asphyxiated by poisonous gases. A sewage system is a collection of underground pipes connected from every residential and commercial building to the treatment plant. from there the water is let into water bodies after treatment of sewage water. Due to improper care of more than thousand workers who works in manhole dies every year due to harmful gases like carbon monoxide, methane, hydrogen sulphide etc, [1] present inside the manhole. These gases need to keep up the track so that huge rise within the normal of effluents level should be known and corrective measures are often taken. On the contrary, the prevailing systems available aren't much portable and aren't affordable. Also, implementation is hard. The designed rover detects the presence of harmful gas in the monitoring system. The device consists of a processing section which takes input, processes and gives output [2]. This technique requires a base station with almost all the sensors.

In this project, an embedded system is meant with an Arduino Microcontroller and various gas sensors for the aim of monitoring, detecting and altering that helps in eliminating the lives of endangered humans [3]. The system is economical to implement and well-defined. within the existing system, a variety of jobs related to gas detection and ensuring security systems. It's been implemented among and a few were

---

[1] Anitha R, Department of EEE, Sri Sai Ram Institute of Technology, TN, India, Email: anitha.eee@sairamit.edu.in.

demonstrated in practice to detect the harmful gas but both approaches were effective gas sensing units have been implemented which is capable of detecting the harmful and toxic gases individually and display in the LCD. Harmful gases like carbon monoxide, Methane, Hydrogen Sulphide these gases are very toxic to the human, Heart Beat sensor this may be fixed on the worker's hand watch and the message is going to be sent to outside workers and Municipal Officers with the assistance of IoT [4].

2. Embedded System and IoT

2.1. Embedded System

An embedded system is designed to perform one or limited set of functions using a combination of software and hardware. It adds intelligence to operation. The characteristics of an embedded system are i)Should have small set of functions ii)Low power consumption since it uses battery iii)Limited memory & Limited peripherals iv)Application predefined v)Not accessible directly vi)Highly reliable & need to operate with time constraints The simplest model consists of a processor, sensors, actuators and memory [5].
This solves any real time problems. Sensor senses the physical change and send it to the processor which is connected with the memory and actuators. The process of translating the code that is written by humans to the code understandable by machine is called the build process [6]. A hardware designer must take into account following features 1. Operation at higher clock speed 2. High computing performance [Pipelining, Prefetch cache and RISC Architecture 3. Fast context switching in Multitasking system. 4. Power efficient embedded system 5. Burst mode access 6. Automatic operation for hardware solution to shared data problem [7-10].

2.2. IoT

IOT(Internet Of Things) is a highly distributed network of smart devices embedded with electronic sensors and software each capable of dynamically generating, analyzing and communicating intelligence that can be used to increase experimental efficiency and power and make life more easier and comfortable is shown in Figure 1 [11].
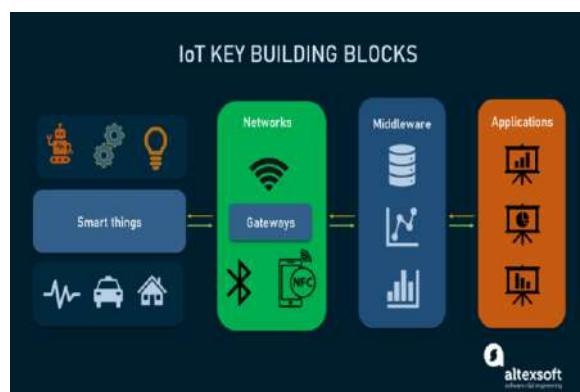
Figure 1. Structure of IOT

IOT devices can collect various types of data through the onboard sensors or attached sensors such as Temperature, Humidity and Gas sensor.

3. Existing System

In existing systems which were mostly used in industrial areas to detect the leakage of harmful gas around them and alert the user, which is large, firm and complex in circuit design and cannot be used in homes, manhole, tunnels to analyse the atmosphere of that working areas [12].

4. Flow Diagram
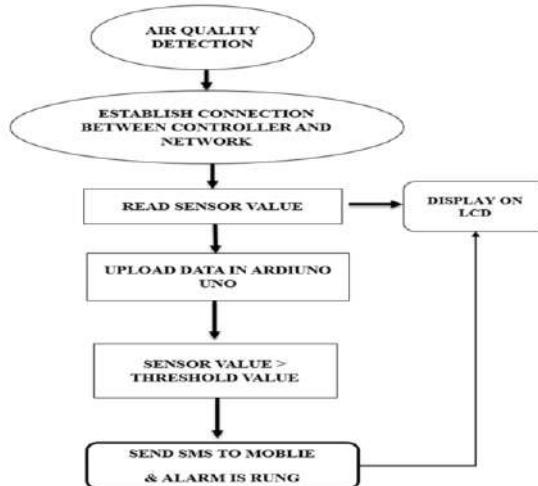
Figure 2. System Flow

As shown in Figure 2, Initially the gas sensor detects no gas. When a supply of 5v is passed to the arduino via a supply board. The mq-02 gas sensor is turned on. The gas sensor detects the surrounding polluted air in ppm (parts per milli). The value sensed by the sensor are analog signals when passed to arduino converts into digital values and gets displayed in the 16*2 LCD. The Arduino checks some threshold limits set by the user to indicate the safety level of air around them [13]. When the value goes beyond a certain threshold limit which is set as the most toxic level at this stage the arduino triggers some remedial measures and also sends the user the location, level toxicity where the device is set up temperature and humidity via sms [14]. An application is also set up so that the user can get an update of the surroundings temperature, humidity and gas values from any location as the values get uploaded to a web server [15].

5. Electronic Image File

This proposed system contains two sections. There are Transmission and shelter gas valve control sections are shown in Figure 3 and Figure 4 respectively. The temperature and gas sensor is interfaced with a microcontroller so as to form the smart system. Where the sensors connected with the microcontroller senses the quality of gas in ppm, temperature and humidity values and updates the IoT webserver. When the gas sensor reaches the edge level then a signal is sent to the gas valve system via WSN. At

the time it suddenly closed the solenoid valve and on the DC fan. Additionally, thereto we are distributing the robot mechanism in the drainage zone. a standing also will be displayed on the 16*2 LCD.

Figure 3. Aerial view of the robot

Figure 4. Aerial view of the remedial

6. Simulation

We used proteus to develop the simulation circuit which is used for electronic device designing in 2D or 3D format using ISIS software in the proteus. The Figure 5 and Figure 6 represents the simulated output.
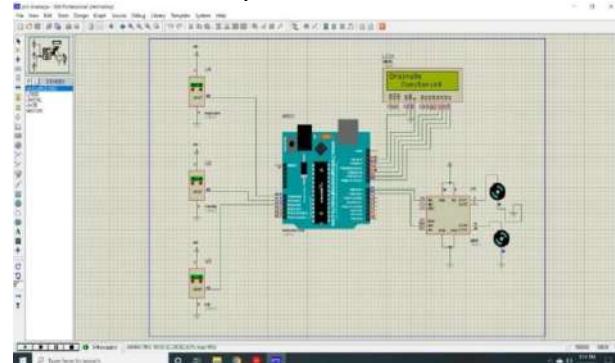
Figure 5. Connection of hardware using Proteus

In Figure 3, the initial values of the sensors are displayed in the LCD.
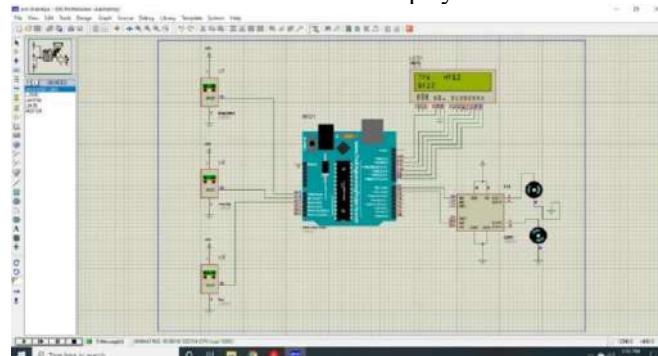
Figure 6. Simulation output

In Figure 6, while changing the sensor values it sends the analog signal to the microcontroller which converts the analog signal to digital signal and displays the changed values in the LCD module. And at the same time relay operates the DC fan and solenoid valve simultaneously.

7. Conclusion

In this paper, we have come up with a potent solution to the various problems such as life threat due to poisonous gas   inhalation by manhole workers, presence of toxic gases in any location of industries, no proper pollution level monitoring and indication and other numerous problems due to air pollution. Here we have proposed an IOT based smart monitoring and fast intimation of presence of toxic gases using Arduino controller. The hardware is designed not only to display the information using LCD display but also to transmit the real time data to the web at a very high baud rate efficiently. The hardware working is tested with the different injected toxic gases and the result is found to be very precise, accurate and fast.

References

[1] Jabbar, W. A., Shang, H. K., Hamid, S. N. I. S., Almohammedi, A. A., Ramli, R. M., & Ali, M. A. H. (2019). IoT-BBMS: Internet of Things-Based Baby Monitoring System for Smart Cradle. *IEEE Access*, *7*, 93791–93805. https://doi.org/10.1109/ACCESS.2019.2928481

[2] Lazaro, A., Boada, M., Villarino, R., & Girbau, D. (2019). Battery-Less Smart Diaper Based on NFC Technology. *IEEE Sensors Journal*, *19*(22), 10848–10858. https://doi.org/10.1109/JSEN.2019.2933289

[3] Berrezueta-Guzman, J., Pau, I., Martin-Ruiz, M. L., & Maximo-Bocanegra, N. (2020). Smart-Home Environment to Support Homework Activities for Children. *IEEE Access*, *8*, 160251–160267. https://doi.org/10.1109/ACCESS.2020.3020734

[4] Gahlot, S., Reddy, S. R. N., & Kumar, D. (2019). Review of smart health monitoring approaches with survey analysis and proposed framework. *IEEE Internet of Things Journal*, *6*(2), 2116–2127. https://doi.org/10.1109/JIOT.2018.2872389

[5] Hesse, N., Pujades, S., Black, M. J., Arens, M., Hofmann, U. G., & Schroeder, A. S. (2020). Learning and Tracking the 3D Body Shape of Freely Moving Infants from RGB-D sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(10), 2540–2551. https://doi.org/10.1109/TPAMI.2019.2917908

[6] Gutiérrez, S. T., Fuentes, C. I., & Díaz, M. A. (2020). Introducing SOST: An ultra-low-cost star tracker concept based on a raspberry pi and open-source astronomy software. *IEEE Access*, *8*, 166320–166334. https://doi.org/10.1109/ACCESS.2020.3020048

[7] Hu, D., Zhang, H., Wu, Z., Wang, F., Wang, L., Smith, J. K., … Shen, D. (2020). Disentangled-Multimodal Adversarial Autoencoder: Application to Infant Age Prediction with Incomplete Multimodal Neuroimages. *IEEE Transactions on Medical Imaging*, *39*(12), 4137–4149. https://doi.org/10.1109/TMI.2020.3013825

[8] Liu, L., Li, W., Wu, X., & Zhou, B. X. (2019). Infant cry language analysis and recognition: An experimental approach. *IEEE/CAA Journal of Automatica Sinica*, *6*(3), 778–788. https://doi.org/10.1109/JAS.2019.1911435

[9] Li, C., Pourtaherian, A., Van Onzenoort, L., Tjon, W. E., & De With, P. H. N. (2020). Infant Monitoring System for Real-Time and Remote Discomfort Detection. *IEEE Transactions on Consumer Electronics*, *66*(4), 336–345. https://doi.org/10.1109/TCE.2020.3031359

[10] Jiang, L., & Hassanipour, F. (2020). Bio-Inspired Breastfeeding Simulator (BIBS): A Tool for Studying the Infant Feeding Mechanism. *IEEE Transactions on Biomedical Engineering*, *67*(11), 3242–3252. https://doi.org/10.1109/TBME.2020.2980545

[11] R. G. Babu, V. V. Nathan, J. Bino, C. Amali and S. Ganesh, "IoT Security Enhancement with Automated Identification Device using IOT SENTINEL," *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2021, pp. 518-523, https://doi.org/10.1109/Confluence51648.2021.9377165

A New Tariff Based Energy Saving and Sharing Scheme from Renewable Energy Using Smart Grid

V.Annapoorani [a,1], J. Banupriya [a], P.Navaraja [a] V.Chinnammal [b]
[a]*Assistant Professor, Department of ECE, Mahendra Institute of Technology (Autonomous), Namakkal, Tamil Nadu, India*
[b]*Assistant Professor (SS), Department of ECE, Rajalakshmi Institute of Technology, Chennai, Tamil Nadu, India*

Abstract. Abstract of our paper's major intent is to manage power and to share the solar load power to grid system by using smart grid technologies that is called as Demand Side Management (DSM). This paper gives the idea of modernized delivery system of electricity in which it observes, safeguards and adjusts accordingly with the energy that is used in home. The objective of the work is when the renewable resources are plentiful and electricity becomes affordable, time-of-use pricing, which allows customers to move some of their energy use to consistent and convenient moment of the day.

Keywords: Load power, grid system, DSM, renewable energy sources, Electricity, Arduino Uno Microcontroller.

1.   Introduction

The entire load which is attached to the power grid varies outstandingly with respect to time. Even though the entire load is added up with the various separate options of the consumers, the whole load is not required to be either compulsorily stable or not moving faster. Considering the instance that a widespread TV schedule begins then million numbers of TV would start taking the current directly [1]. But conventionally a few free generators are pulled to be in a backup manner in order to acknowledge the usage of shoot up power. Thus, the smart grid might alert the TV sets or the customer to decrease the power for a shorter period or at regular intervals [2]. With the support of prediction algorithms found in the mathematics, it is easy to anticipate the number of backup generators utilized and also to obtain a definite failure rate. In the standard grid, the failure rate could be decreased only by the cost of the backup generators. In smart grid the decrease in load occurs even a little part of the customer is removed.

Though conventional load balancing techniques aimed to adapt consumers' spending habits in addition to making demand more uniform, advances in energy storage and individual renewable energy generation have made it possible to create sustainable power grids without impacting consumers' actions. Typically, saving energy during off-peak hours helps with peak-hour availability [3].

---

[1] V.Annapoorani, Assistant Professor, Department of ECE, Mahendra Institute of Technology (Autonomous), Namakkal, TN, India, Email: annapooranisathya@gmail.com.

Condition 1: Hospital
The most important is to take care of human life, so 24 hours no limits of solar power, in demand time period.

Condition 2: Industrial
Generally, EB board provides a power of KVA based on equipment if the solar power demand occurs of about 50 %rapidly the total solar power will be stopped automatically. When the generating solar power increases or reaches 75% automatically total power enters into the industries [4]. Every time total solar power is calculated and monitors the wastage of power. Any one of the industries that involves power wastage, the CT is monitoring and intimating the power wastage period of few times and thus minimizes the load. Since the loads are not minimized automatically total solar powers are shutdown [5].

Condition 3: Domestic
Generally, EB board providing a single phase or three phase power supply based on appliance and if the solar power demand occurs 75 percentages suddenly the total solar power is stopped automatically and at the same time solar power engaging the demand period when generating the solar power increasing or reaching 95 percentages, automatically total solar power enters into the domestic. Each and every time total solar power is calculated and monitored the power wastage and any one domestic involved power wastage the CT are monitoring and intimating the power wastage periods of few times, suddenly minimizing the load. Because the loads are not minimized automatically total solar powers are shutdown [7].

Demand Switches
In this system, the two demand switches are used. Here, the toggle type of demand switch is used. These switches are interfaced with the Arduino Uno microcontroller such that when demand occurs the toggle switch starts to operate automatically [8].

Buzzer
To indicate the wastage of power a buzzer is used. When the solar power consumption exceeds the prescribed level, the power wastage buzzer provides an indication in order to trip the circuit.

Node MCU Unit
A node MCU unit provides the connection between the IOT and the working model is shown in Figure 1.
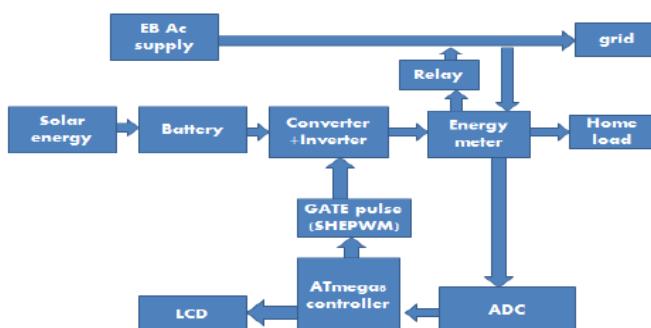
Figure 1. Block Diagram of Proposed System

2.   Description

The ATmega16 is an AVR RISC-based energy saving Complementary Metal Oxide Semiconductor eight-digit microcomputer. For every time period, the outputs of the microcomputer reach one MIPS/MHz with the help of strong commands of solar energy. Thus, this makes the designers to have an equal match of energy usage and time periods. A thirty-two universal storage elements and huge instruction set is present in automatic voltage regulator. A part of central processing unit that performs calculations and logical operations is in straightaway contact with every thirty-two storage elements. This makes the 2 storage elements to be approached every clock cycle with a single instruction [9]. As a result, the architecture is more code effective than traditional CISC microcontrollers. The Atmel ATmega16 is a solar-powered microcontroller that can handle a wide range of embedded control applications is shown in Figure 2.

Smart Grid and Micro Grid

This is related with the upcoming peer groups of electrical system which has an enhancing power saving structure. With the addition of electrical network and IT technology, it interchanges the actual power demanded/ power system data two-way joining the vendor and the buyer [10]. But the buyer part has a smaller sized grid and different factors like energy control, networking etc. which makes others to make up the micro grid. Smart meters, solar panels for the home, slick machines, motor car charging are examples of electric products. Slight distributed style energy storage systems are deployed in the micro grid, and they help to improve smart grid performance by reacting quickly to real-time electricity demand/supply data.

Energy Storage System (ESS)

Renewable energy generated electrically is stored in a storage system for further use. Based on ESS Capacity, ESS products are used. They are categorized as three groups: ESS has given as megawatt (MW) level in case of electricity frequency adjustment. The next level of ESS is large capacity (MW) used for coping with demand at peak times of electricity, and final level of ESS is medium capacity kilowatt (kW) used for domestic use.

Relay
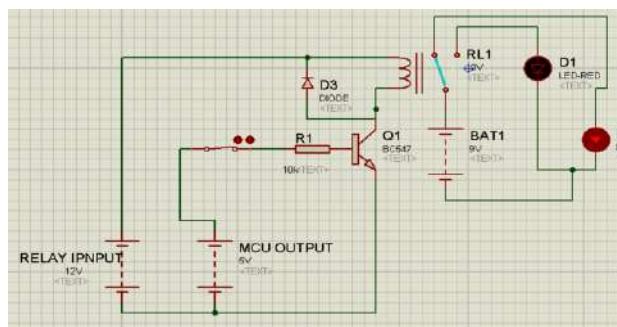
Figure 2. Circuit diagram of Relay

Working of Relay

A relay is a switch which operates electrically so that it attracts lever and changes the position of switch. Magnetic field is generated by current flowing through the coil which is attracted by lever and changes the switch contacts. So current in the coil is in on or off position. This can be double thrown (change over) switches.

The operation of Relays shows that one circuit is turned on whereas the second circuit is disconnected. For example,230 V AC main used as relay to switch a 230V AC mains circuit. Due to maximum output of 555 timer IC which has current of 200mA, so these devices can act as relay coils directly without amplification. Relays are usually SPDT or DPDT

IOT WIFI ESP8266
WIFI module Wi-Fi ESP8266 is a low-cost chip with TCP/IP stack and microcontroller is shown in Figure 3. In our paper, we use a Web server to display information such as the amount of energy consumed in units, the bill, and if any theft occurs. As a result, any user can access the information from anywhere in the world. The paper's information is displayed on the Thing speak web page.

Figure 3. Diagram of WIFI

Block Diagram of Solar Power Supply
The safe operating area of a transistor is defined by minimizing available output current for large input to output voltage differentials. These regulators are economical and easy to use, which is suitable to design a system with bountiful printed-circuit boards by which unregulated dc is brought to each board and regulation is performed is shown in Figure 4.

Figure 4. Power Supply Block Diagram

Bridge Rectifier Circuit
The circuit shows as a bridge rectifier which contains four diodes are connected as shown in the Figure 5. The input to the circuit is given diagonally opposite sides and the output is taken between other two corners. It produces twice that of output voltage as compared to full wave rectifier. Main advantage of bridge rectifier is it produces two times of output voltage over a conventional full-wave rectifier.

LCD Display
LCD stands for Liquid Crystal Display which has limited viewing angle is shown in Figure 5. LCD used as economical one in order to display 16 characters and symbols. D0-D7 will represent as commands or as data on logic states. RS: RS = 1 for this status Bits D0 - D7 used for representing addresses of characters to be displayed. Some Built in processor addresses has built in "map of characters" and to show appropriate

symbols. RS = 0 for this status Bits D0 - D7 are commands which is used to find the display mode is shown in Figure 6.
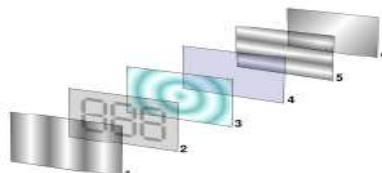
Figure 5. Structure of LCD Display

Features of LCD Display

- Cursor has 5 x 8 dots
- Built in integrated controller (ks 0066 or equivalent)
- Operates power + 5v

Schematic Diagram of LCD Display
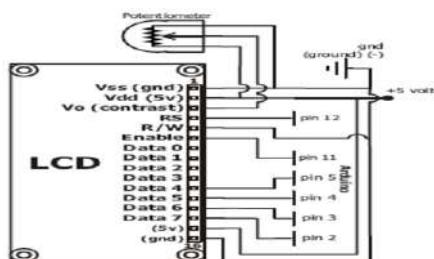
Figure 6. Schematic diagram of LCD Display

Operation

When demand occurs on the grid side the demand switch 1(toggle switch) will operates. The priority is given to the hospital and industrial concern. So that the load from the domestic side is shared to the industrial and hospital facilities. At the same time the minimum amount of supply is given to the domestic side at that situation if they consume too much of supply then the power wastage buzzer will give intimation and then trips the circuit. If the demand on the grid side is further increased then the demand switch 2 will operate, so that the load from the industrial side is shared to hospital facilities. At that situation if the industry consumes power wastage then the power wastage buzzer will provide intimation and then trips the circuit.

 Here we are using toggle type demand switch. The demand switch operation depends on the fixed rating. The current transformer continuously senses the current rating and if there is any excess current then the current transformer will sense the signal and gives the signal to the relay circuit which trips the circuit accordingly. The IOT module continuously monitors the overall operation of the system.

3.   Result and implementation
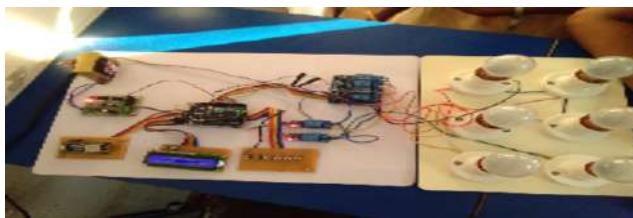
Figure 7. Solar based power sharing

Apart from the regular monitoring of the solar power line, it also helps to monitor the weather and its effect on the solar power system is shown in Figure 7. With this idea about the weather patterns [7] near the solar power transmission line is obtained and also which event is going to occur based on the data collection is predicted shown in Figure 8. Also, in India the problem of solar power theft is common, thus this problem is avoided with the help of IOT and prevents the solar power theft.

Figure 8. Diagram for Smart mode

4.   Conclusion

Thus, we can conclude that this work using smart grid solves the problem of unidirectional information flow, wastage of energy and manages the growing energy demand and provides intelligent and advanced power control using AVR Studio software and can be applied in three different modes as home mode, Grid mode and smart mode. Hence the tariff can be reduced.

References

[1]   Yaacoub, E., & Abu-Dayya, A. (2014). Automatic meter reading in the smart grid using contention based random access over the free cellular spectrum. *Computer Networks*, *59*, 171–183. https://doi.org/10.1016/j.bjp.2013.10.009
[2]   Sun, H., Nallanathan, A., Tan, B., Thompson, J., Jiang, J., & Poor, H. (2012). Relaying technologies for smart grid communications. *IEEE Wireless Communications*, *19*(6), 52–59. https://doi.org/10.1109/MWC.2012.6393518
[3]   Yigit, M., Gungor, V. C., & Baktir, S. (2014, September 9). Cloud Computing for Smart Grid applications. *Computer Networks*. Elsevier B.V. https://doi.org/10.1016/j.comnet.2014.06.007
[4]   Deng, R., Yang, Z., Chow, M. Y., & Chen, J. (2015). A survey on demand response in smart grids: Mathematical models and approaches. IEEE Transactions on Industrial Informatics, 11(3), 570–582. https://doi.org/10.1109/TII.2015.2414719

[5]   Ma, R., Chen, H. H., Huang, Y. R., & Meng, W. (2013). Smart grid communication: Its challenges and opportunities. IEEE Transactions on Smart Grid, 4(1), 36–46. https://doi.org/10.1109/TSG.2012.2225851

[6]   Bush, S. F. (2014). Network theory and smart grid distribution automation. IEEE Journal on Selected Areas in Communications, 32(7), 1451–1459. https://doi.org/10.1109/JSAC.2014.2332132

[7]   Collier, S. E. (2017, March 1). The Emerging Enernet: Convergence of the Smart Grid with the Internet of Things. IEEE Industry Applications Magazine. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/MIAS.2016.2600737

[8]   Temel, S., Gungor, V. C., & Koçak, T. (2014). Routing protocol design guidelines for smart grid environments. Computer Networks, 60, 160–170. https://doi.org/10.1016/j.bjp.2013.11.009

[9]   Meng, W., Ma, R., & Chen, H. H. (2014). Smart grid neighborhood area networks: A survey. IEEE Network, 28(1), 24–32. https://doi.org/10.1109/MNET.2014.6724103

[10] Wang, W., Xu, Y., & Khanna, M. (2011, October 27). A survey on the communication architectures in smart grid. Computer Networks. Elsevier B.V. https://doi.org/10.1016/j.comnet.2011.07.010

[11] Hannoon, N., Vijayakumar, V., Vengatesan, K., & Hidayat, N. (2019). Small Signal Fault Analysis for Renewable Energy (Wind) Power System Distributed Generation by Using MATLAB Software (Simulink). Journal of Computational and Theoretical Nanoscience, 16(2), 537-543.

Automobile Authentication and Tracking System

N.Duraichi [a,1], K.Arun Kumar [b], N.Lokesh Sathya [b], S.Lokesh [b]

ªProfessor, Department of Electronics and Communication Engineering,
ᵇUG Student, Department of Electronics and Communication Engineering,
Vel Tech High Tech Dr Rangarajan Dr Sakunthala Engineering College, Chennai, TN,
India

Abstract. Vehicle robbery and unknown car thefts has become a intense issue around the nation. Many culprits use unapproved vehicles to perform numerous illegal activities and leave the vehicles. The utmost reason for accidents is due to the vehicles driven by unknown users, who perform reckless and inexperienced driving without the speed limit will cause many accidents that increases the death rate. Our goal is to make a system which will allow the person who have authorized license. For this purpose, we plan to install an automated system in the vehicle to introduce smart license verification technology. Various techniques and technologies are being explained to detect the details of the driver, and also Various vehicle thefts are being done in spite of various surveillance cameras are set down to keep an eye on the activities and various technologies are being implemented to diminish the vehicle robbery. So, we proposed the system with the concept of deep learning. As compared to normal detection techniques deep learning collects N number of input samples and compares it with the database details. After the authentication process the engine mechanism starts, if not authorized it gives a buzzer sound and vehicle doesn't start until the details of registered person is authenticated.

Keywords. Deep Learning Techniques, Image Processing, Arduino ATmega328 microcontroller

1.   Introduction

Nowadays, we can see that many people who doesn't have license (children below age 18) are driving vehicles without thinking about the accident or the danger it can create for them and as well as others [1]. Our aim is to make a system that which allows only the person who has the authorized driving license. For this purpose, we need to set up a system in vehicle. For input we use the RFID tags and for output digital screen will be used [2]. Now the person who wants to drive has to place his RFID tag near the Reader and then the system will verify it with the central data. If the input details are wrong then the system will not start and the person cannot drive the vehicle. If the detail in the RFID tag matches with the details in the central data, then the vehicle will start and the person can drive the vehicle now [3].

---

[1] Dr.N.Duraichi, Department of Electronics and Communication Engineering, Vel Tech High Tech Dr Rangarajan Dr Sakunthala Engineering College, TN, India; Email: duraichi@velhightech.com.

2. Existing System

In the existing system, there were no advanced techniques used. The only way to verify the drivers was to stop the vehicle in the traffic and verify the driver's license and this causes the traffic jam in the peak hours. The checking police or RTO'S cannot check each and every vehicle in the traffic section and the rate of accident have not been reduced [4].

Disadvantages:
- Manual method is used and rate of accident is high
- Time consumption for verification is high.

3. Proposed System

As shown in Figure1 and Figure 2, In our project we are using the Infrared Sensor which will detect the whether the person is sitting in the driver seat and the details will be sent the microcontroller. If the Infrared Red sensor detects the person, then he has inserted the key. The user has to place the RFID tag in the reader and in that the driver's license details is fed in the micro controller and it reads the details of the driver [5]. From the microcontroller the instructions will be sent to the camera and it will take the picture of the driver and then the captured picture will be compared with the picture stored in the system. If both the details of the person match with the stored details then the car mechanism will turn ON the car and if the details doesn't match then the buzzer which will give the alarm. By this setup we can check the driver license Addition to this setup we are adding checking device in which we need setup a checking lane at the tollgates and check posts to receive the details of the driver using the ZIGBEE module and it will be uploaded in the server through the IOT module [6]. And if the Infrared sensor didn't sense the presence of person, then the details will be sent to the microcontroller and from there the car will be immediately turned off with the help of motor mechanism. From that we are able to trace the car theft and the driver without license which will reduce the rate of accidents [7].
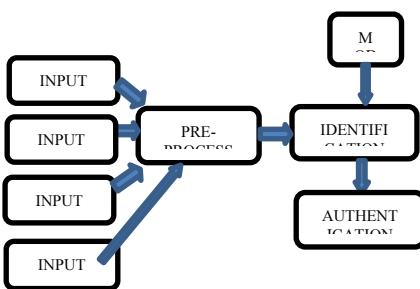
Figure 1. Block Diagram for Testing Section   **Figure 2**. Block Diagram for Training Section
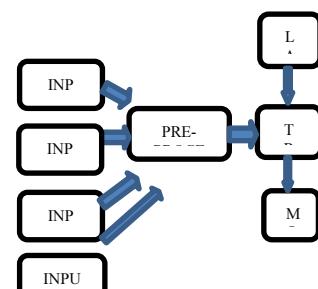
3.1 Operating principle

When looking for somebody, the machine will take a photo of them to check their identity. This is referred to as a 1:1 verification process; if we search for a specific person's photo in a database, it is referred to as a 1: N search. In both cases, the device

will generate a corresponding remuneration result, either success or annoyance. The block diagram depicts details about face enrolment testing and preparation [8]. After attaching the camera to the PC, we must first open the computer and repeat the enrolment of the same person image several times. Once the face enrolment is complete, we can add the approved license ID number and the authorized person's Aadhar card number, and then that information will be saved in Centralized database [9].

3.2 ARDUINO ATmega328

Arduino is a software and hardware platform for projects and users that designs and produces microcontroller kits for creating digital devices and interactive objects that sense and manage objects in both the physical and digital worlds. A variety of microprocessors and controllers are used on the Arduino board. The boards have digital and analogue input/output (I/O) pins that can be connected to a variety of expansion boards (shields) or breadboards (other circuits on them). Arduino's architecture, or more precisely, Arduino's integrated circuits (ATmega328p). The ATmega328/P is a low-power CMOS 8-bit microcontroller with an AVR instruction set that reduces risk. The AVR uses Harvard design to optimize efficiency and similarity, with separate reminiscences and buses for program and details. With one stage of pipelining, instruction inside the program memory region is dead. Associate degree external 16MHz oscillator controls the clock [10].

3.3 USB Camera

The USB camera is a gadget that is used to capture a picture of the person. Image of a person will give us a scope for the quick identification of a particular person. Face capturing system are widely used for security purpose due to its acceptability, accuracy, Feasibility, and reliability.

3.4 Image Processing

Image processing functions are based on uniformity measurements are used to obtain better performance person authentication, creating an image of the particular person. The image processing technique is used to extract high resolution images from the low-resolution images by removing the noise in it. A CNN application, perfectly trained, provides suitable for particular person face recognition when used in coexisting with a previously developed face recognition module. It is also shown that image processing functions based on SIFT measurements can be used to obtain performance object features location, creating a key point face reorganization for every feature. A grid of Scale Invariant Feature Transform (SIFT) key points was drawn and a neural network was used also for vehicle model recognition.

3.5   Deep Learning

Deep learning is a machine learning subcategory that employs multiple layers to derive progressively such higher-level options from raw inputs. Lower layers in an image processing technique, for example, could decide edges, while higher layers could

determine concepts related to human identification, such as numbers, letters, ears, or fingerprints. Every stage learns to rework its input file into a variety of abstract, linear, and composite representations using this technique. In a picture recognition program, the raw input could be a matrix of pixels; the primary realistic layer could abstract the pixels and write edges; and the secondary realistic layer could abstract the pixels and write edges. The second layer could write and compose edge arrangements; the third layer could write a nose and eyes, and thus the fourth layer could recognize that the image contained a face. Significantly, a deep learning system can learn on its own which options to best position within which stage.

4.   Results

Here in the proposed system, we will place an IR Sensor, when the IR Sensor detects the person, it will display the "P DETECT" on the screen as shown in the Figure 3.

Figure 3. Displaying P DETECT

Next after the person detected by the IR Sensor, we need to place the RFID Tag near the RFID Reader. Here in my proposed model, we assigned the name of the RFID Tag as ARUN. So, it displays "P DETECTARUN" on the Screen as shown in the Figure 4.

Figure 4. Displaying P DETECT ARUN   **Figure 5.** Displays IMG MATCHED V:ON

Once after the RFID Tag reading all the details will be stored in the device will be checked with the driver sitting in the driver seat. Then it displays "IMG MATCHED V:ON" on the screen as shown in the Figure 5 and Figure 6 and the vehicles will starts. Whenever an unauthorized person tries to drive the vehicle the buzzer goes ON and it will displays the "IMG NOT MATCV:OF" on the screen as shown Figure 7, Figure 8 and Figure 9.
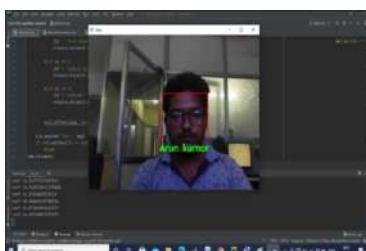
Figure 6. Image Verification Process   **Figure 7.** Displays IMG NOT MATCHED V:OFF
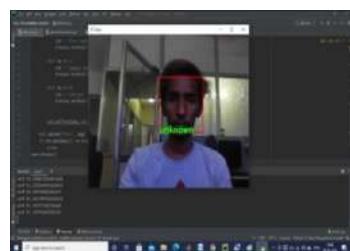
Figure 8. Experimental Setup of the Car Section   **Figure 9.** Experimental Setup of the Toll Gate Section
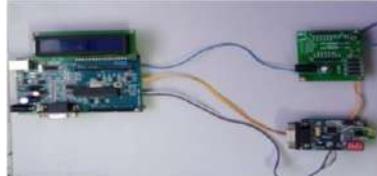
5. Conclusion

Proposed method is sketched for verifying the driving license basing on Arduino ATmega328 and face authentication are designed to achieve the gadgets by coding which are essential for the face authentication. This system integrates both the intersection and coding using PyCharm and Arduino ATmega328 microcontroller. This system can be future enhanced by using different techniques like Deep Neural Network (DNN) etc. This system can be utilized for the authentication of the driving license, detection of fake license for decreasing the crime and accidents rate.

References

[1]   Deep Das, P., & Sengupta, S. (2017). "Implementing the next-generation system to protect vehicles from thefts and accidents". 2017 International Conference on Innovations in Green Energy and Healthcare Technologies(IGEHT).doi:10.1109/igeht.2017.8094045

[2]   Rajalakshmi, G. (2015). Real time smart car lock security system using face detection and recognition. *International Journal of Applied Engineering Research*, *10*(4), 11307–11312

[3]   Luo, D., Lu, X., Hu, C., Peng, X., & Wang, W. (2017). A robust traffic driver face detection method. In *PIC 2016 - Proceedings of the 2016 IEEE International Conference on Progress in Informatics and Computing* (pp. 285–288). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/PIC.2016.7949511

[4]   Saifullah, Khawaja, A., Hamza, Arsalan, Maryam, & Anum. (2010). Keyless car entry through face recognition using FPGA. In *2010 International Conference on Future Information Technology and Management Engineering, FITME 2010* (Vol. 1, pp. 224–227). https://doi.org/10.1109/FITME.2010.5654862

[5]   Kolli, A., Fasih, A., Al Machot, F., & Kyamakya, K. (2011). Non-intrusive car driver's emotion recognition using thermal camera. In *Proceedings of the Joint 3rd International Workshop on Nonlinear Dynamics and Synchronization, INDS'11 and 16th International Symposium on Theoretical Electrical Engineering, ISTET'11* (pp. 156–160). https://doi.org/10.1109/INDS.2011.6024802

[6]   Faisal, M., & Thakur, A. (2016). Autonomous car system using facial recognition and geo location services. In *Proceedings of the 2016 6th International Conference - Cloud System and Big Data Engineering, Confluence 2016* (pp. 417–420). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/CONFLUENCE.2016.7508155

[7]   Zhixiong, L. (2005). A new embedded car theft detection system. In *ICESS 2005 - Second International Conference on Embedded Software and Systems* (Vol. 2005, pp. 89–94). https://doi.org/10.1109/ICESS.2005.6

[8]   Al-Shemarry, M. S., Li, Y., & Abdulla, S. (2020). An Efficient Texture Descriptor for the Detection of License Plates from Vehicle Images in Difficult Conditions. *IEEE Transactions on Intelligent Transportation Systems*, *21*(2), 553–564. https://doi.org/10.1109/TITS.2019.2897990

[9]   Du, S., Ibrahim, M., Shehata, M., & Badawy, W. (2013). Automatic license plate recognition (ALPR): A state-of-the-art review. *IEEE Transactions on Circuits and Systems for Video Technology*. https://doi.org/10.1109/TCSVT.2012.2203741

[10] Huang, R., Tawfik, H., & Nagar, A. K. (2009). Licence plate character recognition based on support vector machines with clonal selection and fish swarm algorithms. In *11th International Conference on Computer Modelling and Simulation, UKSim 2009* (pp. 101–106). https://doi.org/10.1109/UKSIM.2009.64

Deep Learning Based Indian Sign Language Words Identification System

P. GOLDA JEYASHEELI [a,1], N. INDUMATHI [b]

[a]*Professor*, *Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, TN, India*
[b]*Assistant Professor, Department of Computer Science and Engineering, P.S.N.A College of Engineering and Technology, Dindigul, TN, India*

Abstract. In Indian Population there is about 1 percent of the people are deaf and dumb. Deaf and dumb people use gestures to interact with each other. Ordinary humans fail to grasp the significance of gestures, which makes interaction between deaf and mute people hard. In attempt for ordinary citizens to understand the signs, an automated sign language identification system is proposed. A smart wearable hand device is designed by attaching different sensors to the gloves to perform the gestures. Each gesture has unique sensor values and those values are collected as an excel data. The characteristics of movements are extracted and categorized with the aid of a convolutional neural network (CNN). The data from the test set is identified by the CNN according to the classification. The objective of this system is to bridge the interaction gap between people who are deaf or hard of hearing and the rest of society.

Keywords. Convolutional neural network, wearable device, sign language, sensors, gestures

1. Introduction

Deaf and dumb people communicate using physical gestures in Sign language. Hand motions and body expressions are used to express the message. Indians choose a sign language which is known as Indo-Pak sign language, that is a combination of Indian and Pakistani sign languages [1]. Based on the expressed signals, the movements are divided into two categories: The term "manual signals" refers to finger movement, hand orientation, and posture. Mouth action, face expression, and body posture are all important factors to consider as non-manual signals [2]. A sign language identification system is required for non-deaf and mute people to understand deaf and mute people's hand gestures. A smart framework for gesture recognition is being designed for this purpose. The Indian sign language is effectively identified by a smart sign language recognition device in this article. Wearable sensor-based gloves will be used in the smart sign recognition system to decode Indian sign language gestures. On each hand, there are five flex sensors, a pressure sensor, and a Bluetooth module connected to microcontrollers on both sides [3]. The sensor data is gathered in order to identify gestures. The microcontroller is connected to the Bluetooth modules in both hands. After that, the sensor information is sent to the CNN, which categorizes and displays the identified sign term [4].

---

[1]N. INDUMATHI, Assistant Professor, Department of Computer Science and Engineering, P.S.N.A College of Engineering and Technology, Dindigul, TN, India; Email: indhujegan1996@gmail.com.

2. Related Work

There are two techniques for determining the manual signals which are produced by the deaf and dumb people. The first is a direct approach, while the second is an indirect approach. The gesture is classified using neural networks in a direct approach. Changshui Zhang et.al. [11] suggested a method for converting sign language from videos into a sentence format. For feature extraction, the system uses a deep CNN, and a bidirectional RNN is used to understand the sequence of motions. Ying Xie et al [12] suggested Deaf and dumb people will get benefit from a computer vision-based approach to interpret sign language. This method converts sign language into text, making it easier for the people to interact. The CNN is being used in the model to derive the sign image's spatial features from a series of video clips. And, to train the temporal features, RNN is applied to the image's extracted features [5]. After building, this model illustrates and justifies the actual model to operate on the dataset by using the image dataset to train the network for categorizing the images. Until passing on to the next surface, Inception performs many of the parallel convolutions and combines the resulting feature maps [6].

The gesture is determined by a sensor wearable is being used in an indirect approach, with the neural network obtaining image as input. To determine the letters of American Sign language, the wearable interface uses flex, pressure, and three-axis inertial motion sensors. Sensor data is obtained and evaluated to use an embedded SVM classifier. The acknowledged alphabet is then sent through Bluetooth low-power wireless networking is used to connect to a mobile device [7]. A text-to-speech function has been added to an Android app, which transforms obtained text into loud audio files. A SVM tool used to classify the signs in several different categories [8-10]

3. Prototype Design

The structure of the system is composed of two different modules: one is a wearables interface module and another is sign identification module that identifies deaf and mute people's sign words. These modules will make deaf and mute people's lives easier people to reach out the general public. Figure 1 depicts the proposed system's overall architecture. The smart wearable system is the hardware part in this paper, and it is made up of flex, pressure and Gyroscope sensors that are all linked to the Arduino MEGA microcontroller [13]. They're stitched to the cloth glove to make it more accessible. The master and slave wearables communicate in serial fashion is developed using the HC-05 Bluetooth module. We used flex sensor with 4.5 inch and flex sensor 2.2 inch [14]. On both the hands, the 4.5-inch flexible sensors are used for all fingers except thumb finger. The thumb finger on both the hand uses the 2.2-inch flexible sensors.

Above the palms of both hands, the pressure sensor is mounted. A resistor is linked to the pressure sensor. To ensure that the current flow is stabilized, the Flexi sensor is mounted to a resistor. The above three sensors are interconnected to the microcontroller in order to obtain precise examinations for each sign generated by the smart wearables. Figure 2 depicts the equipment configuration for the wearable's module [15].
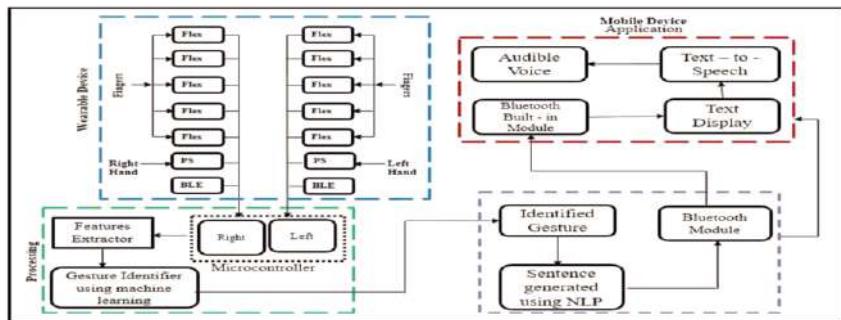
Figure 1. Proposed System Architecture

Figure 2. Hardware setup for the wearable device module

Sensor information is gathered from Fifty different individuals. The data collection was generated with all of the sensor data obtained from people who used the system. The sensor examinations are saved in a spreadsheet. The captured sensor information is used in the gesture identification module to classify and identify the gesture. For the categorization of the sign generated by the wearables module, the sensor information is analyzed using a CNN. Figure 3 represents a diagram of the CNN's architecture.
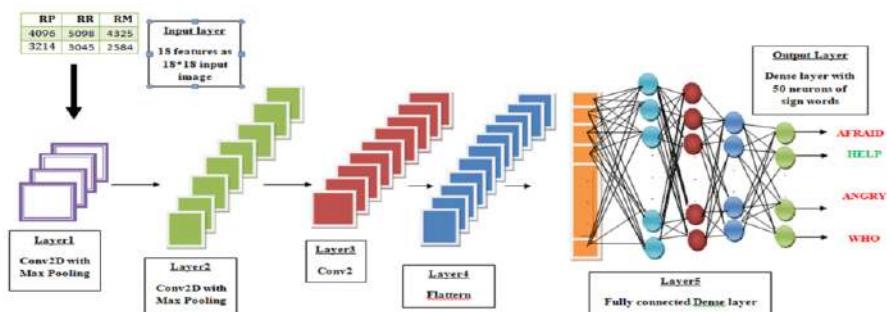
Figure 3. CNN Architecture

In the input layer, eighteen input attributes are passed into the CNN, including the value of all the 3 types of sensors used for all Fifty signs. The output layer would have

the same number of neurons as the network's classes. It represents the estimated word as a value of one other word with the value of 0. As a result, the output layer generates Fifty neurons for each identified sign term.

4. Execution

The measured sensor data serve as bend categories, indicating how each term bends depending on its sensor data. To prevent differences between sensor readings, the wearable system generates normalized values. The flex bend values determine how movements are categorized. CNN are used to assess the system's identification. Although this system contains 50 signs with ordinal attributes, classification of the signs centred on the data from their sensors is necessary. To be able to accurately accommodate changes in hand movement, the gesture is identified with a duration interval of 10 seconds. CNN is loaded with random weights and implements numerical data conversion into image format for processing. After the image conversion, the images are stored as the image data store and perform convolution operations on the images. The validation of the training and testing set will be evaluated and the recognized words are displayed as the class labels based on the prediction of the test data. The image gets reshaped in the size of 18*18 for making all the images in the same size for validation. The class labels are displayed based on the prediction.

5. Results and Discussions

The network's accuracy is determined by analyzing the samples of both training and testing are shown in Figure 4. The data samples are collected at random, so further training would be required to boost accuracy. For better accuracy, a certain ratio of the dataset is retained. In this project, the training and testing ratio is defined as 60:40 to achieve higher accuracy. The 50 signs will be identified by the machine.
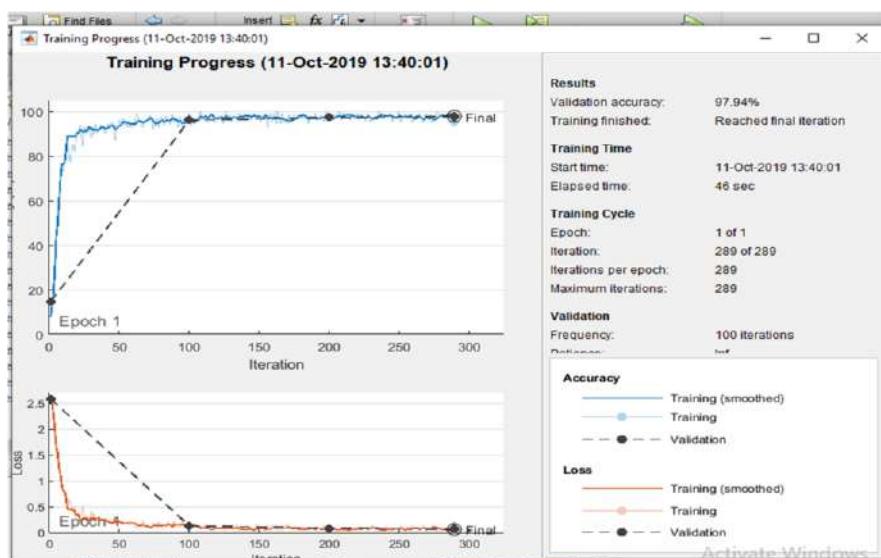
Figure 4. Training progress of CNN

There are 50 different samples in each sign. In total, there are 2500 samples. The accuracy of the proposed system's convolutional neural network is shown in Table 1.

Table 1. Convolutional Neural Network Accuracy

Trial Cases	Total no of Samples	Training samples	Testing Samples	Accuracy of CNN
Trial case 1	2500	1500	1000	89%
Trial case 2	2500	2000	500	**97.8%**

In contrast to current systems, the proposed system has improved accuracy. The accuracy of the proposed system is compared to that of the current system in Table 2.

Table 2. Accuracy comparison between other existing systems and proposed system.

Author	Used Methodology	Achieved Accuracy
Su Min Lee [5]	Sensor fixed gloves	92.5%
Gwang Soo Hongb [6]	Sensor Gloves	93.9%
Pamela Godoy-Trujill [10]	Sensor gloves	93.3%
Nishith A. Kotak [11]	Sensor gloves	91.5%
R Jafari [14]	Wrist-Worn, SEMG	91.2%
Proposed system	Sensor gloves	97.8%

6. Conclusion

A smart glove device is proposed in this project for categorizing Indian sign words. This smart glove captures the gesture created by deaf and dumb people. For identifying each sign word, the sensor values of the respective signs are processed and categorised using a CNN. This network helps to identify the word with the aid of training data. For each value of the 3 types of sensors, the network is trained with a collection of 18 neurons as inputs. The word is recognised based on the training results.

References

[1]   Mittal, A., Kumar, P., Roy, P. P., Balasubramanian, R., & Chaudhuri, B. B. (2019). A Modified LSTM Model for Continuous Sign Language Recognition Using Leap Motion. *IEEE Sensors Journal*, *19*(16), 7056–7063. https://doi.org/10.1109/JSEN.2019.2909837

[2]   Cui, R., Liu, H., & Zhang, C. (2019). A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training. *IEEE Transactions on Multimedia*, *21*(7), 1880–1891. https://doi.org/10.1109/TMM.2018.2889563

[3]   Kaur, J., & Rama Krishna, C. (2019). An efficient indian sign language recognition system using sift descriptor. *International Journal of Engineering and Advanced Technology*, *8*(6), 1456–1461. https://doi.org/10.35940/ijeat.F8124.088619

[4]   Bantupalli, K., & Xie, Y. (2019). American Sign Language Recognition using Deep Learning and Computer Vision. In *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018* (pp. 4896–4899). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/BigData.2018.8622141

[5]   Lee, B. G., & Lee, S. M. (2018). Smart Wearable Hand Device for Sign Language Interpretation System with Sensors Fusion. *IEEE Sensors Journal*, *18*(3), 1224–1232. https://doi.org/10.1109/JSEN.2017.2779466

[6]   Kim, J. H., Hong, G. S., Kim, B. G., & Dogra, D. P. (2018). deepGesture: Deep learning-based gesture recognition scheme using motion sensors. *Displays*, *55*, 38–45. https://doi.org/10.1016/j.displa.2018.08.001

[7]   Bhujbal, V. P., & Warhade, K. K. (2019). Hand Sign Recognition Based Communication System for Speech Disable People. In *Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems, ICICCS 2018* (pp. 348–352). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICCONS.2018.8663054

[8]   Shaheen, H., & Mehmood, T. (2018). Talking Gloves: Low-Cost Gesture Recognition System for Sign Language Translation. In *2018 IEEE Region 10 Symposium, Tensymp 2018* (pp. 219–224). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/TENCONSpring.2018.8692069

[9]   Jadhav, D., & Tripathy, A. (2018). Gesture Aided Speech for Deaf and Mute. In 2018 3rd International Conference for Convergence in Technology, I2CT 2018. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/I2CT.2018.8529411

[10] Rosero-Montalvo, P. D., Godoy-Trujillo, P., Flores-Bosmediano, E., Carrascal-Garcia, J., Otero-Potosi, S., Benitez-Pereira, H., & Peluffo-Ordonez, D. H. (2018). Sign Language Recognition Based on Intelligent Glove Using Machine Learning Techniques. In *2018 IEEE 3rd Ecuador Technical Chapters Meeting, ETCM 2018*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ETCM.2018.8580268

[11] Jani, A. B., Kotak, N. A., & Roy, A. K. (2018). Sensor Based Hand Gesture Recognition System for English Alphabets Used in Sign Language of Deaf-Mute People. In *Proceedings of IEEE Sensors* (Vol. 2018-October). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICSENS.2018.8589574

[12] Oliveira, T., Escudeiro, P., Escudeiro, N., Rocha, E., & Barbosa, F. M. (2019). Automatic sign language translation to improve communication. In *IEEE Global Engineering Education Conference, EDUCON* (Vol. April-2019, pp. 937–942). IEEE Computer Society. https://doi.org/10.1109/EDUCON.2019.8725244

[13] Galka, J., Masior, M., Zaborski, M., & Barczewska, K. (2016). Inertial Motion Sensing Glove for Sign Language Gesture Acquisition and Recognition. *IEEE Sensors Journal*, *16*(16), 6310–6316. https://doi.org/10.1109/JSEN.2016.2583542

[14] Wu, J., Sun, L., & Jafari, R. (2016). A Wearable System for Recognizing American Sign Language in Real-Time Using IMU and Surface EMG Sensors. *IEEE Journal of Biomedical and Health Informatics*, *20*(5), 1281–1290. https://doi.org/10.1109/JBHI.2016.2598302

[15] Hsu, Y. L., Chu, C. L., Tsai, Y. J., & Wang, J. S. (2015). An inertial pen with dynamic time warping recognizer for handwriting and gesture recognition. *IEEE Sensors Journal*, *15*(1), 154–163. https://doi.org/10.1109/JSEN.2014.2339843

Dementia People Tracking System

S. DEEPA [a,1], LAVANYA DHANESH [b], DANUSHA V [c], DIVYA DATH K [c],
PAVADHAARINI G K [c], SHOBANA SRI C [c]

aProfessor, Department of EEE, Panimalar Engineering College, Chennai
bAssociate Professor, Department of EEE, Panimalar Institute of Technology, Chennai
cUG Student, Department of EEE, Panimalar Institute of Technology, Chennai

Abstract. The number of elderly people worldwide is rigidly increasing due to decrease in birth rates and innovations implemented in medical field. Due to the increase in elderly people population diseases like dementia are also being increased year by year. Having done many kinds of research it is found that there is no permanent treatment for diseases like dementia, even if those patients come in public they look similar to normal people, however, people with dementia have abnormal behaviors like loss of patience, aggression, lack of thinking which in turn causes burden to family members and caretakers. In order to address this issues, this paper demonstrates a follow-up and rescue program for the elderly. The system includes a GPS receiver, a GSM module and a long-distance RF transmitter and receiver, real-time location. Families and care takers can obtain real-time information and history of patient location through GPS to avoid loss of elderly patients. With the help of this system, the number of losing patients will be decreased and the pressure on the caretakers and family people will be cut down to some extent.

Keywords. Patient Tracking, Health Monitoring, GPS, GSM Module, Buzzer.

1. Introduction

Safety is one of the important concerns in present days. Incidents like misbehaving with old people are also raising day by day. It has been found that nearly 30 million people are suffering from dementia, autism, Alzheimer's [1]. The people suffering from dementia if they have been lost from their residence, this device helps in locating, health monitoring, alerting the surroundings with the help of a buzzer, and tracking the patients who got lost from their homes [2].

2. Existing System

Even in the past years, there are many tracking and monitoring systems for vehicle tracking and children monitoring [3]. A person named Kennedy in 2007 implemented alert notification through text messaging which is extracted from Amber alert system and that system helped in finding kids who got kidnapped [4]. and they were also indicated by a map through some icons and visual characteristics along with the locality of the vehicle in 2005 [5]. King and Yancy these two persons innovated that if

---

[1] S.Deepa, Department of Electrical and Electronics Engineering, Panimalar Engineering College, India Chennai, Email: deesoms123@gmail.com.

any vehicle had an emergency, it will be reached to the destination safely and quickly and there was also a location scheme

3. Proposed System

To solve this problem, we propose a technique based on the Arduino Uno board and RF receiver and transmitter modules. If the RF receiver and RF transmitter both have strong signals, it signifies the dementia patient is in range. The buzzer will sound if the patient is out of range, and the GSM location will send a message to the pre-registered mobile number with GSM location. So that dementia patient can be easily located. We also use a heartbeat sensor and a temperature sensor to monitor patients [6].

The heartbeat sensor monitors heartbeats, whereas the Dallas temperature sensor monitors temperature. And if the patient has a problem, it will send an SMS through GSM and a buzzer will sound to alert the patient. The data will be displayed in the LD. Here using Arduino Uno, RF receiver and RF transmitter modules, Buzzer, heartbeat sensor, Dallas temperature sensor, GSM, GPS, LCD. RF transmitter and receiver modules are used [7]. If the RF transmitter and RF receiver are both in range, it means the dementia patient is in range. If the patient is out of range, the buzzer will alarm, and the GSM location will be sent to the per-registered mobile number. As a result, you'll be able to easily locate the dementia patient. We're also monitoring patients using a heartbeat sensor and a temperature sensor. Heartbeat sensor and Dallas temperature sensor are for monitoring heartbeat and temperature, respectively. And if the patient has a problem, it will send an SMS through GSM as well as a buzzer to alert them. In the LCD, data will appear.
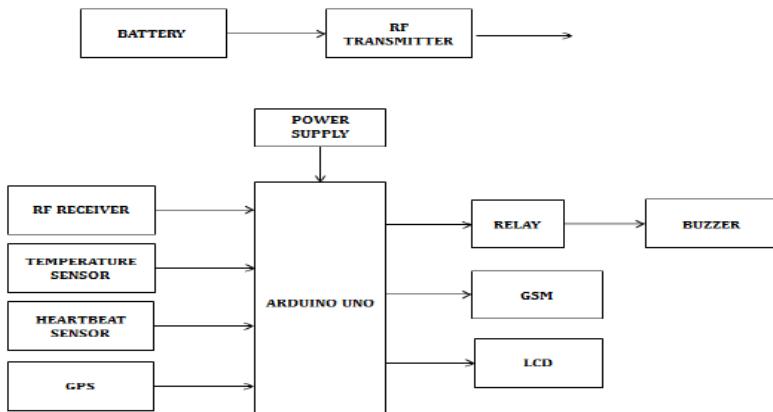
Figure 1. Block Diagram of Proposed System

4. Hardware Description

Arduino UNO:

It is a microcontroller ATmega328P which was developed by Arduino.cc and it is an open-source. It consists of analog input/output pins and digital pins. It has 14 digital pins and 6 analog pins. Arduino IDE software is used in Arduino UNO board using the USB cable or 9-volt battery

Power Supply:

A power supply that is regulated converts the alternating current (AC) to a constant direct current or voltage (DC). In this regulated power supply even though input changes output remains constant.

LCD (Liquid Crystal Display):

LCD   modules are very commonly used because of its cheap price. It is also programmer-friendly. 16×2   LCD is mostly used in recent days. It will have   32 characters and each one of the character will be of   5×8 Pixel Dots.

GPS (Global Positioning System):

It is a satellite navigation system and is 20,000km far from the earth. This provides location and time information. It requires a minimum of 24 satellites and now 33 satellites work in GPS and work for 24 hours a day.

GSM:

GSM full form is global system for mobile communication. SIM800 is a quadband GSM SIM800 current consumption is as low as 1.2mA. SIM800 is very helpful for data transfer instructions as it integrates TCP/IP protocol and extends TCP/IP AT commands.

RF module – Transmitter & Receiver:

They are small in dimension but have a wide voltage range that is   3V to 12V. These modules are 433 MHz They consist of two microcontrollers for data transfer.

Temperature Sensor:

This is a   digital sensor like DS18B20. It is a single wire protocol.it is used to measure temperature with an accuracy of +-5%.

Relay:

Relay is similar to switch. Relay is same as a switch that connects or disconnects two of the circuits. Relay is used with the electrical signal instead of manual operation.

Heartbeat Sensor:

It is a device that helps to measure the speed of the heartbeat. Heart rate can be measured or monitored using two ways 1. Manually check the pulse at the wrist or neck. 2. Use a heartbeat sensor.

Buzzer:

It is a magnetic device that produces a magnetic field when a signal and that too it must be oscillating, moves across the coil which helps in wavering the disk at a frequency that is similar to the drive signal.

5. Software Description

Arduino IDE:

The Arduino Integrated development environment is software that can be used for any Arduino boards this Arduino IDE helps in writing the code easily and dump it to the board. It is open-source software. The programming language used in it is java. Arduino is a microcontroller ATMEGA 328P for the UNO.

6. Advantages

- Don't take much time as in the traditional method.
- Can easily monitor the dementia patient.

7. Future Scope

In our proposed system, we demonstrated an innovative proposal of a low-cost GPS tracking system that relies on mobile using GPS module, which at this present and the future point will be helpful for Alzheimer diseased patients [8]. The solution for this is that the architecture consists of five components web platform, web server, web service for tracking location, mobile application, and database. In the future, we can add a feature that allows caretakers to share their experiences too so that others will be knowing and learning from them.

8. Results and Discussion
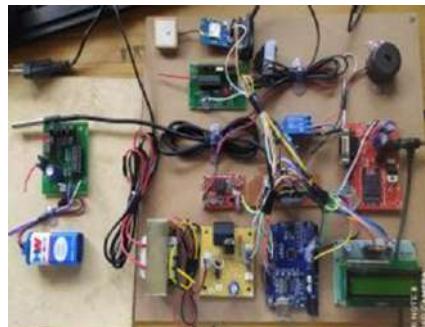
Figure 2. Outputs of Transmission and Reception

Figure 2 show the entire kit of transmitter and receiver. Figure 2 tells the pulse rate of the person using a heartbeat sensor. Figure 2 shows the temperature value using the Dallas temperature sensor and it also shows that the patient is in range. Figure 2 tells about the location of the patient with the help of GPS. Figure 2 shows the patient is out of range. These readings and values will be sent to the caretakers or family members through messages to the prescribed phone number using GSM and alert them about the patient's condition and place. There is also another feature of the buzzer system which alerts the surroundings that the patient around them is suffering from some kind of dementia or some chronic diseases. Through this system, we can trace out the patients and reduce the burden of family members.

9. Conclusion

The system we proposed is a tracking device using RF receiver and RF transmitter modules for those people who are fighting every day from diseases like dementia and it can also be used for Alzheimer's disease. This system is affordable as it is of low cost and low complexity and the best part is it consumes low power. The people suffering from these diseases can purchase it as it is of low cost and it reduces the burden of family members, caretakers. We have done the trials of our prototype and it is successful in showing the patients' location, temperature, and pulse rate through the messages which will be sent to the family members mobile phone using

GSM. We can add the databases like NGO phone numbers and nearby police stations which will be not difficult for the caretakers to tell to the police and NGO people to find and get hold of the patient at their location.

References

[1] Sara Paivaa, Carlos Abreua, "Low Cost GPS tracking for the Elderly and Alzheimer Patients", HCIST 2012 – International Conference on Health and Social care Information Systems and Technologies, Procedia Technology 5 (2012) 793 – 802.

[2] Ms. Smita S. Auti, Prof. Nagnath B. Hulle, "Advanced shoes with embedded position tracking and path guidance to keep track of Alzheimer's patients", International Journal of Scientific and Research Publications, Volume 5, Issue 1, January 2015.

[3] Robert P. O'Brien, Srinivas Katkoori, and Meredith A. Rowe, "Design and Implementation of an Embedded System for Monitoring At-home Solitary Alzheimer's Patients", Circuits and Systems (MWSCAS), 2015 IEEE 58th International Midwest Symposium.

[4] Yao-Jen Chang, Chien-Nien Chen, Li-Der Chou, "A Novel Indoor Way finding System Based on Passive RFID for Individuals with Cognitive Impairments", Pervasive Computing Technologies for Healthcare, 2008. Second International Conference.

[5] Al-Anssari, H. and Abdel-Qader, "Vision Based Monitoring System for Alzheimer's Patients Using Controlled Bounding Boxes Tracking", Electro Information Technology (EIT), 2016 IEEE International Conference.

[6] Dhanesh, L., Deepa, S., Elangovan, P., & Prabhu, S. (2020). Enhanced and Energy-Efficient Program Scheduling for Heterogeneous Multi-Core Processors System. In *Lecture Notes in Electrical Engineering* (Vol. 665, pp. 737–747). Springer. https://doi.org/10.1007/978-981-15-5262-5_5

[7] Maheswari, V., Elangovan, P., Baranidharan, M., Deepa, S., & Dhanesh, L. (2019). Theoretical and simulation analysis of first generation DC-DC converters. *International Journal of Advanced Science and Technology*, *28*(19), 72–78.

[8] C. Jain, G. V. S. Sashank, V. N and S. Markkandan, "Low-cost BLE based Indoor Localization using RSSI Fingerprinting and Machine Learning," *Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2021, pp. 363-367, https://doi.org/10.1109/WiSPNET51692.2021.9419388

GSM Based Smart Energy Meter System

KANNAN P [a,1], ALLWİN DEVARAJ [a], ESAKKİ RAJAVEL. S [a], NISHARNI B[b],
SANCTA A [b], MUTHU LAKSHMİ M [b]

[a]*Assistant Professor, Francis Xavier Engineering College,*
Tirunelveli, TN, India
[b]*UG Student, Francis Xavier Engineering College, Tirunelveli, TN, India*

Abstract. This project aims to develop a smart energy meter which has many advantages like monitoring and controlling the energy meter at anywhere and at any place. Our project also called as pay first and then use it because here the user first has to pay the bill after that only they can consume the electricity. We know that each user has different kind of load demands and based on it we can recharge in number of units (KWh). The energy we are consuming and based on those pulses from the energy meter, microcontroller decrements the number of units and displays on LCD the remaining units we can use. The load gets disconnected when the units become zero and it is reconnected when the user recharges it. Our project will be very useful for the electricity board and also the user because this project gives the live update about the energy usage of the user through the message. GSM (Global system for mobile communication) will help to send and receive this information to the user. This system helps us to identify the energy losses of the electricity board to a large extent.

Keywords. Kilo Watt Hour, Liquid Crystal Display.

1. Introduction

In this modern world we are living in the "Automation world" because nowadays all are depending and addicted to this automation instead of the man power [1]. This kind of sustenance is now spreading in all nations particularly like our developing nations. Also, one of the most imminent thing in this world is Electricity which is highly used in many places such as for agricultural, industrial and also for household purposes. Although we enriched with electricity sources, till we are suffered with many problems in energy distribution, measuring the bill and controlling the energy consumption [2]. we know that electricity is playing an important role by giving coziness to our life so it should be used very carefully by without wasting it Electricity is one of the vital requirements for the sustainability of life's comforts and as such it must be used very wisely for its proper use [3]. But in our country, we have a lot of areas where we have surplus electricity supply while many areas cannot even get to it. Their distribution policies are also partly responsible for this because we still were unable to properly estimate us [4].

---

[1] KANNAN P, Assistant Professor, Francis Xavier Engineering College, Tirunelveli, TN, India; Email: kannamuthu1987@gmail.com.

The implementation is done in such a way that SMS is delivered to GSM MODEM whose readability must be noted, and this metric is then returned to the server in SMS format and SMS costs are known to be very low [5]. GSM AMR takes advantage of available nationwide infrastructure coverage GSM and the cell broadcast feature of SMS system to order and retrieve individual homes and build reading power consumption back to the power supply wirelessly [6].

Goal of the proposed machine are

- To make energy intake prepaid.
- To centralize the manipulate of all energy meters.
- To prevent power thefts at consumer premises.
- To lessen fee of manpower for billing.

2. Proposed System

In the Figure 1, microcontroller, LCD and the GSM unit is interfaced with the power Meter of each house. every meter has a separate RR quantity, which is given with the aid of the electricity providing organizations [7]. the heart beat from the energy meter is given to optocoupler and output of the optocoupler is given to the microcontroller. The microcontroller constantly decrements the quantity of gadgets (1 unit= 1 kWh) as consistent with electricity consumption and shows the available gadgets on the LCD connected to the strength meter [8].

Output of power meter is attached to the weight via the relay. right here relay is used to attach or disconnect the supply to the burden. while the server (electricity supplying organisation) sends a request to the GSM modem, the modem commands the microcontroller to carry out the required action. The microcontroller in response sends the reply to the modem and the mod, in turn, sends the facts to the server. MAX232 is used for records switch between microcontroller and GSM modem [9].
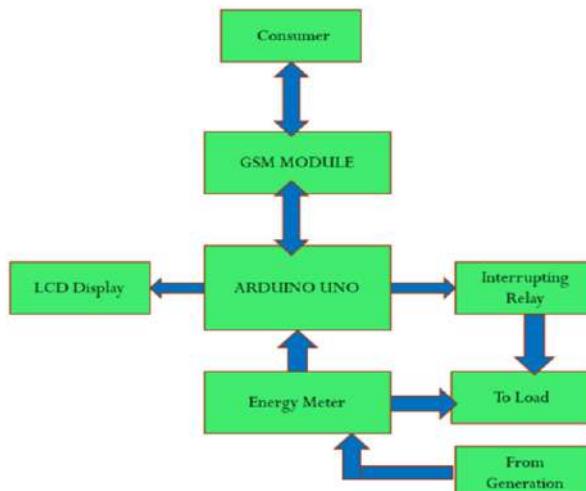
Figure 1. Block Diagram of GSM Module

GSM helps to deliver the Electricity fed on data to the application organization in addition when it is needed to the client. Aerial, connected close to meter container, which is mainly helped for development of signal energy in GSM. Automatic metering

communique is the imminent thing in meter reading, so the maintenance staff from electricity board don't need to visit each consumer for meter reading [10]. Nevertheless, checking out and upkeep meters may additionally want to study from time to time. The primary responsibility of smart meter is to measure the energy that we have consumed and have to send it to software while call for also to consumer [11-13]. The Effective values of voltage is measured by the voltage detector which is given to the micro controller, and inside the micro controller active and reactive energy calculations are carried out. In our project we have used current and voltage transformers and simultaneously as contradict to voltage detectors. A observation from the software organization is gained from the automatic energy meter which is a programmable one and the activities done by the meter is highly matched with the given enumeration a main advantage of automatic Energy meter is that we can disconnect and reconnect the power supply of the user without the help of the man power [14-15]. We can also reconnect the power supply once it is disconnected. The main characteristic of the Automatic Energy meter is that it alerts the client by ringing an alarm whilst the consumer load crosses the threshold restriction. If the consumer does not reduce his power usage then the meter automatically cut off the connection. GSM helps to deliver the energy ate up facts to the application management in addition to the consumer while demanded. Aerial, connected onor near meter container, maybe usedfor improvement of sign energy in GSM communication.

3. Results and discussion

We can also check the correctness of our proposed system by comparing the values which may be displayed in LCD display of Automatic energy meter and the units we have consumed also received through SMS. We can also check this system by connecting and disconnecting the client connections. Also we connected different loads to it and checked its overall performance. The checking out of SEM provided correct consequences, as a result verifying the performance and accuracy of the device. As shown in Figure 2, a basic connection is set up with a bread board with all additives like clever electricity Meter, lcd display, GSM Module, Load and so on.

Figure 2. Overall setup with program dumped in GSM Module

Then program is dumped into the GSM Module in C language and then a kit is prepared and all soldering take place and the project is ready to execute. So once the program is dumped from the GSM Module a message alert saying System Ready reaches the person whose mobile number is registered. Once the system gets ready the

load will automatically ON and the units and total balance is shown and it decreases as the power is consumed. The outputs are shown in Figures 3,4 and 5.

Figure 3. System Ready   **Figure 4.** LCD Display of total units and balance

Figure 5. Final Output

4. Conclusion

The evolved GSM primarily based energy meter reading and billing are useful for each electricity provider supplying utility and customers. This machine overcomes the drawbacks of the conventional meter reading system and affords extra offerings together with energy cut and tampering alert. The evolved system additionally gives records about daily, month-to-month, and yearly electricity utilization details regarding day-by-day power intake will assist the customer to control their strength usage. This developed gadget is dependable and relaxed as best legal individuals can access the device.

References
[1] Amruta Chore, Prasad Mali, Dinesh Vyanjane, Vijay Karewar, 2018, "IOT Based Smart Electricity Meter and Billing System", IRJET journal, vol: 05, no: 10, pp: 916-919.
[2] Yogesh B.Mandake, Abhishek Pandey, Mohit Sareen, Pratyush Prasoon, Rishav Raj, Ayush Agrawal, 2020, "GSM Based Smart Energy Meter", IRJET journal, vol:07, no: 06, pp: 7618- 7621.
[3] Tejas Chate, Aniket Kahalkar, Prashant Soni, Dinesh Gawande, 2017, "A Review on GSM Based Monthly Electrical Energy Billing through SMS", IRJET journal, vol: 04, no: 03, pp: 1068-1069.

[4] Shital S. Bobade, Vibhuti S. Jadhav, Anuja A. Mokal, K. D. Mahajan, 2016, "Automatic Energy Meter Billing System with Theft Detection", IRJET journal, vol: 03, no: 04, pp: 237-240.

[5] Swapnil V. Vanmore, Prathamesh Avalekar, Indrajeet Patil, 2016, "Prepaid and Postpaid Energy Meter for Billing System Using Microcontroller and GSM", IRJET journal, vol: 03, no: 02, pp: 1214-1216.

[6] K. Sindhuja, M. Sravya, Y. Rajkumar, P. Anjali rao, J. Ravichander, 2020, "Electricity Energy Meter using IoT", IRJET journal, vol: 07, no: 04, pp: 1177-1182.

[7] Vairaperumal K, Dinesh M, Senthilkumar P, Mubeena S, 2018, "Effective Energy Measurement and Billing System Using Arduino GSM Shield", IRJET journal, vol: 05, no: 11, pp: 836-840.

[8] Darshit S. Patel, H.B. Patel, 2017, "GSM Based Energy Meter Monitoring and Load Control" IRJET journal, vol: 04, no: 03, pp: 360-363.

[9] Seema More, Arati Kumbhar, Jyoti Chavan, Prof. Ghewari M.U, 2019, "Smart Energy Meter Billing, Monitoring and Controlling System", IRJET journal, vol: 06, no: 03, pp: 3420-3424.

[10] Kiran Mohite, Nisha Kadam, Aishwarya Kadam, Ramchandra Gurav, 2019, "Smart Prepaid Energy Meter Using GSM Technology", IRJET journal, vol: 06, no: 03, pp: 5079-5082.

[11] Shivam Roy, Shreyans Patel, Mahir Udepurwala, 2019, "IOT Based Prepaid Energy Meter", IRJET journal, vol: 06, no: 03, pp: 1158-1161.

[12] Tejas Gujrati, E Terence, 2019, "ESP8266 Based Pre-Paid Electricity System", IRJET journal, vol: 07, no: 06, pp: 682-685.

[13] Pooja S. Vaidya, Chanchal S. Kature, Sonali G. Waghmare, Ms. S. A. Dhumane, 2019, "IOT Based Energy Meter Billing System", IJASRET journal, vol: 04, no: 05, pp: 4-6.

[14] A.Subba Rao, Sri VidyaGarige, 2019, "IOT Based Smart Energy Meter Billing Monitoring and Controlling the Loads", IJITEE journal, vol: 08, no: 4S2, pp: 340-344.

[15] Hannoon, N., Vijayakumar, V., Vengatesan, K., & Hidayat, N. (2019). Small Signal Fault Analysis for Renewable Energy (Wind) Power System Distributed Generation by Using MATLAB Software (Simulink). Journal of Computational and Theoretical Nanoscience, 16(2), 537-543.

[16] Shivprasad S. More, Ketaki K. Korane, Akshaykumar K. Khamkar, Aditya V. Jadhav, Nachiket C. Keripale, Ajay B. Shinde, 2020, "Prepaid Energy Meter System", IRJET journal, vol: 07, no: 06, pp: 4771-4775.

Grid-Tie Rotating Solar Rooftop System Using Atmega

K.Ushanandhini [a,1], A.Kiruthika [a], Dharmaprakash [b]
[a]UG Student, Department of Electrical and Electronics Engineering,
[b]Professor, Department of Electrical and Electronics Engineering,
Panimalar Institute of technology, Chennai, TN, India

Abstract. This paper presents a grid-tie rotating solar rooftop system solar power project which is powered by using Atmega 328 microcontroller. It includes solar panel, LCD display, and battery charging circuit and an inverter circuit with sun tracking capability. This project represents whether a particular industrial or residential load would be powered by the photovoltaic panel or the company. This project is based on Atmega 328 micro-controller which controls the solar array by rotating it consistently with the position of sun. This energy obtained from the solar array is then stored in battery which is then sent back to power the domestic or industrial area. The remaining energy is then reverted to the power house through the gird-tie system. Hence with the assistance of this project, power usage can be reduced by the renewable source of energy and profit can be earned with the help of the power which is fed back to the grid.

Keywords. Solar panel, Inverter, Atmega 328 Microcontroller, sun tracking capability, Grid-tie system, Profit

1. Introduction

Solar energy has been used for thousands of years in many different ways by people all over the world. This renewable energy source is cost free, non-polluting, greenhouse gases cannot be emitted while producing electricity and it is feasible, inexhaustible. However, with the efficient experimentation of scientists around the world, the solar energy is not efficiently or insufficiently used for the demand. According to recent survey of International Energy Agency, India will face the rising demand for energy over the next two decades, due to 25% of global growth, and is set to over lead the European Union by 2030. This project makes use of Atmega 328 microcontroller which controls the solar array by rotating it according to the position of sun. The sun tracking of the panel can be viewed through the LCD display and it automatically switches to the position in which maximum energy can be generated. The energy from the solar panel can be stored in battery or directed to the inverter which is then used to power the residential or industrial area. You can reduce energy costs, which will rise the value of your home or office and reduce toxic waste without losing the safety of the public power grid. The remaining energy is returned to the power station with the help of transformers through the grid-tie system [1][2]. Hence with the help of these system lots of power intake can be reduced.

---

[1] K.Ushanandhini, Department of Electrical and Electronics Engineering, Panimalar Institute of technology, Chennai; Email: ushanandhinius29@gmail.com.

2. Existing System

An incorporation of solar power technologies into the existing energy structure depends on an in-depth understanding of the solar resource. The roof must be positioned so that it will be easier for the shingles to catch the sun. There must be sufficient area to put in solar shingles to gather enough light to supply electricity.

3. Proposed System

Our project presents a better and economical way to synchronize the solar panel output with the utility grid is shown in Figure 1. The Atmega 328 microcontroller was well-programmed such that the energy from the solar panel would provide electricity to the inverter circuit and also as it charges the battery circuit if needed.

The energy from the inverter circuit is then used to power loads as the primary source of electrical energy [1]. If the load acquainted excess energy either directly from the battery or inverter, available power is then fed back to the power station through the secondary transformer. Otherwise, if the electricity demand hang-back below the energy available by the solar panel, the microcontroller would cut the power line source [3].

As a result, an individual solar panel system owner can act as an electricity provider to the national power grid along with the residential grid. This method is usually carbon free or carbon neutral and intrinsically doesn't emit greenhouse gases during its operation.
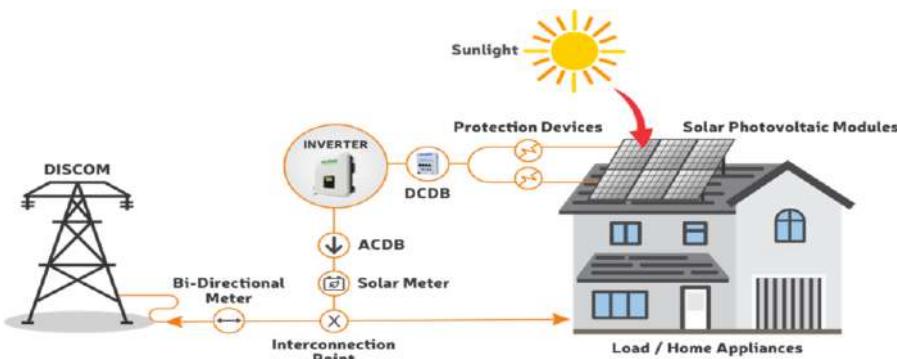
Figure 1. Proposed system

4. Working

It is an innovative system which can be used to store the abundant solar energy into a battery following in which excess energy can be returned to the power station using grid-tie. The population growth has led to increase in demands of people by both commercial as well as residential. The exhaustible resources are getting used-up at a rapid rate if this trend continuous we may face the energy crisis in future. In order to address this issue, our system comes to use [6-7].

The system begins by monitoring solar energy at different positions. It identifies the position with maximum power and reaches that position.

There are three modes of operation
- Battery charging mode
- Load mode
- Grid mode

The system is powered using Atmega 328 microcontroller. It includes a solar panel and LCD display, battery charging circuit and an inverter circuit. Initially, the system monitors the solar power at different positions and angles and it automatically reaches the position with maximum incident solar energy. The outputs are shown in figure 2, If the inverter is off, the system is in battery charging mode. The energy from the solar panel used to charge the battery [2]. If the inverter is ON, now the load mode is ON. The energy accumulated in the battery is converted into AC power, which is used to run AC load.

In Grid-mode, if the load gets turned OFF, all the excess energy is reverted to the power station with the help of grid-tie system. The system monitors power supplied to the grid and simultaneously keep running the total energy returned to the grid [4-5].
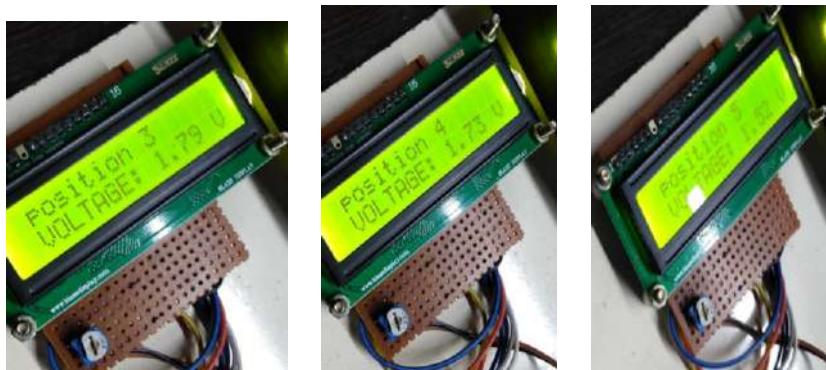
Figure 2. Load display of charging

4.1 Battery charging mode

The Figures 3 and 4 depicts the prototype model built. In battery charging mode, a 12V battery is mutually coupled with the transformer and an inverter. And an inverter is connected with the load, which comes under load mode [1-2].

As mentioned above, the electricity is generated by solar panel connected with a transformer which is then passed to inverter that converts the DC power to AC power, then fed through the MOSFET and IC drives, for the internal operation of Atmega 328 [2-4].

Without inverter operation, the energy available on the panel on the roof becomes unused. So, the role of inverter becomes predominant. The operation of battery and inverter system doesn't require the electrical grid to remain active, and that is the main advantage of off-grid residential and industrial installations [3].

Figure 3. Prototype Model

4.2 Grid mode

In an AC coupled system, the grid-tied inverter is paired to the transformer, load and the battery bank [5]. The off-grid inverter provides a alternate source, which effectively tricks the grid-tied inverter into staying online. This enables you to charge your batteries and run essential appliances during an influence outage [3]

Figure 4. Grid mode

5. Illustrations

Figure 5 shows the inverter circuit and Figure 6 shows the overall circuit diagram.

6. Conclusion

The presented paper represents a photovoltaic system to produce electricity to domestic applications to reduce the over-reliance on the power generating company, which would reduce the energy costs. The given model includes PV solar array system with sun tracking capability, battery storage, a DC to AC inverter and an Atmega 328 microcontroller, which determines whether a certain residential load would be powered by the solar panel or the power generating company. This project monitors the power supplied to the grid and simultaneously keep running the total energy returned to the grid. It is an innovative system which can be used to store the abundant solar energy
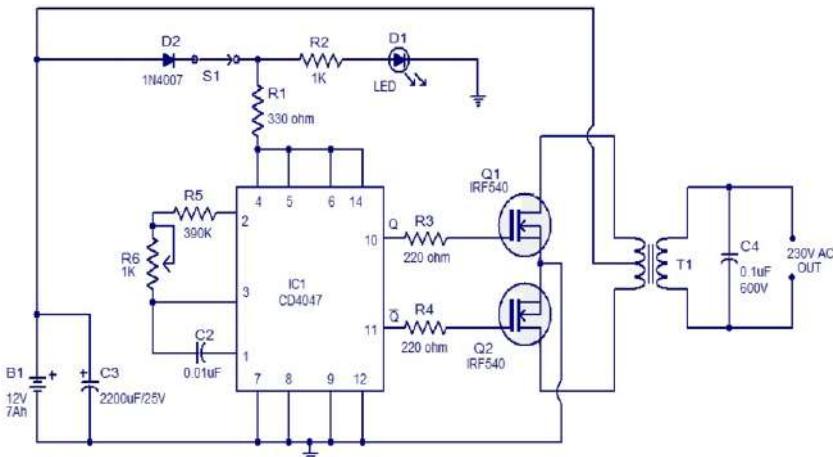
Figure 5 Inverter circuit
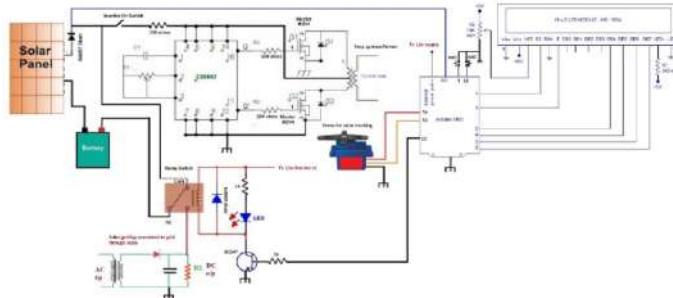
Figure 6 .Overall circuit diagram

into a battery following in which excess energy can be returned to the power station using grid-tie. This system may be executed in real residential applications too. However, both the photovoltaic system and Atmega 328 microcontroller have to be modified to fit the targeted applications fulfilling the power demand.

References

[1] Pearce, J. M. (2002). Photovoltaics - A path to sustainable futures. *Futures*, *34*(7), 663–674. https://doi.org/10.1016/S0016-3287(02)00008-3

[2] Jahn, U., & Nasse, W. (2004). Operational performance of grid-connected PV systems on buildings in Germany. *Progress in Photovoltaics: Research and Applications*, *12*(6), 441–448. https://doi.org/10.1002/pip.550

[3] Denholm, P., Margolis, R., & National Renewable Energy Laboratory. (2008). Supply Curves for Rooftop Solar PV-Generated Electricity for the United States. *National Renewable Energy Laboratory*, (November), 1–23. Retrieved from papers://a0cac570-6072-44e1-9d44-cdd15f758707/Paper/p2083

[4] Mackamul, K. and Wieting, R. The Shell Solar 245 kW Grid-Connected CIS Thin Film PV Rooftop Array: System Design and First Year Performance, Proceedings of the 19th European PVSEC, 2004, pp. 3084-3086.

[5] Payne, A., Duke, R., & Williams, R. H. (2001). Accelerating residential PV expansion: Supply analysis for competitive electricity markets. *Energy Policy*, *29*(10), 787–800. https://doi.org/10.1016/S0301-4215(01)00014-3

[6] Ganesh Babu R., Karunakaran A., Manikandan G., Kalimuthu Kumar S., Selvameena R. (2021) IoT in Smart Automation and Robotics with Streaming Analytical Challenges. In: Singh K.K., Nayyar A.,

Tanwar S., Abouhawwash M. (eds) Emergence of Cyber Physical System and IoT in Smart Automation and Robotics. Advances in Science, Technology & Innovation (IEREK Interdisciplinary Series for Sustainable Development). Springer, Cham. https://doi.org/10.1007/978-3-030-66222-6_7

[7] R. G. Babu, V. V. Nathan, J. Bino, C. Amali and S. Ganesh, "IoT Security Enhancement with Automated Identification Device using IOT SENTINEL," *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2021, pp. 518-523, https://doi.org/ 10.1109/Confluence51648.2021.9377165

Removal of Specular Reflection on Handheld Camera Devices

R.PRATHIPA [a,1], M.PREMKUMAR [b], M.ARUN [c], S.SATHIYA PRIYA [b]

[a]Associate Professor, [b]Professor, [c]Assistant Professor
Department of Electronics and Communication Engineering,
Panimalar Institute of Technology, Chennai, TN, India

Abstract. Handheld mobile devices are a dime a dozen in the industry, thanks to rapid technological advancements. However, none of the algorithms currently available on the market can remove reflections in a hand-held camera system in real time. Our Algorithm is intended to mitigate this loop as it is designed especially for smart phones, we're taking two photos of the same goal here. One shot will be taken with the flashlight turned OFF, while the other will be taken with the flash turned ON. After that, we'll perform a simple pre-processing phase called aligning the image, which involves extracting features. We will eventually get the final desired picture of good quality after all the planned operations are completed.

Keywords. Features Extraction, SIFT, RANSAC, Color Transformation, Image splicing.

1. Introduction

Although it is the industry standard to remove specular reflection with several flashes, our aim is to remove the castback with just two pictures. [1]. A multiple flash camera is shown in figure 1. By optimising the process, the information required to perform the scuttle can be obtained by examining just two images instead of several. The product of the illumination's spectral energy distribution and the surface reflectance is used to compute the spectral energy distribution of light reflected from an object.
Based on Dichromatic Reflection Model, the reflected light can be split into two parts because of specular and diffuse reflections. [8]. Because specular reflection is unavoidable in the real world and captures valuable scene details such as surface shape and source characteristics, the study strictly adheres to the use of specularity regions. [9]. Earlier methods for segregating reflection components based on the quantity of images used can be divided into two groups. The first group employs many images, while the others employ a limited number of images.
Many algorithms have been created to achieve this aim. These algorithms can be classified as either Lambertian assumption-based algorithm [2] or color-based algorithm or polarization-based algorithm or ICA (Independent Components Analysis) based algorithm.

---

[1]R.PRATHIPA, Department of Electronics and Communication Engineering Panimalar Institute of Technology, Chennai, TN, India; Email:prathipa.srini@gmail.com.

For instance, ICA-based reflection removal algorithm [3], requires images obtained from a linear polarizer at various angles and these pictures must be linear combination of the original light and the reflected lights. Principle Component Analysis is used to obtain information about original light [4]. Reflection removal is also required in camera-equipped smart phones. When we try to capture image of objects behind windows, landscapes behind glass covers, or information on monitors with our phones, the reflections often extremely degrade the quality of the image and we will not be able to use it. Worse, none of the algorithms mentioned above can be used to solve the problem. The knowledge available to the algorithms processing a single image is insufficient to recreate the image behind the glass. The task of taking photos with different polarizations is impractical with a mobile.
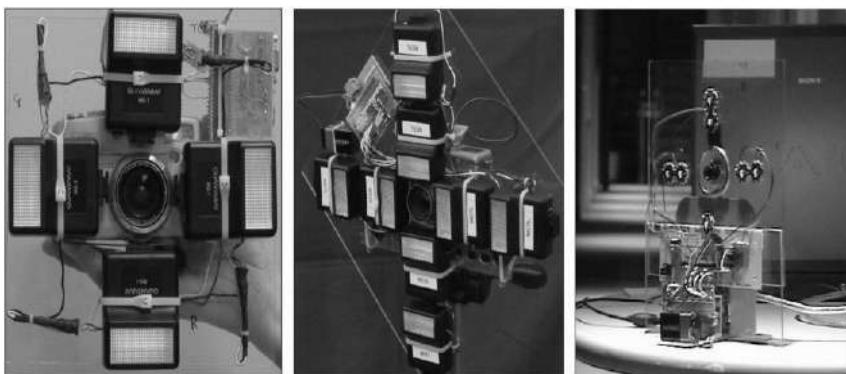
Figure 1. Multiple Flash Camera

It's also not feasible to request users to take images of the object from various angles. Here, we are introducing a reflection removal algorithm designed for mobile with camera. This algorithm requires two pictures - one should be taken with the flashlight and the other without the flashlight. After obtaining the two pictures, the proposed algorithm merges them to produce a new picture with minimal reflections. There is no restriction about the position in which the pictures are taken. It can be obtained from same or at different angle. Researches in the field of Digital Image Processing always aims to enhance the quality of the image. Whenever it comes to cast back removal or specular reflection removal denoising through multiple snaps of images is the most recent development, which is widely known by all. Other than this technique, colour based reflection removal, polarisation-based reflection removal [7] and Independent Component Analysis [10] Techniques are very popular for reflection removal [11]. Not just the field of engineering but, even medicinal field requires Digital Image Processing for inspecting and analysing certain vital internal organs through catheter imaging. The main drawback of the tongue image analysis through total variation-based image inpainting is the texture replacement, which becomes tedious. The reason being, the authors have resorted to converting the RGB image in the processing to indexed image. To do this they follow minimum variance quantization and image dithering techniques. In doing so, the space and color resolution gets hampered. But we have penned down a mitigation strategy for this problem by converting the RGB image to a gray scale image. So, the hue and saturation get removed but luminescence is still retained. We will just obtain two snaps of images in our proposal. First one will be with the

flashlight OFF and second one will be with the flashlight ON. The one without the flashlight will only have a poor image quality. Second one will have a very good light quality but, the undesired flash will be present. So, we just replace the information lost due to the flash with the information from the First image. The user must click once and the app will capture two pictures consecutively, with the flashlight on for the second picture alone. This algorithm has a sequence of operation as shown in Figure 2.
 The second part will examine the features of the two images obtained and clarify our algorithm's overall concept and sequence. Third part describes how to recognise and identify the flash spot in the photograph taken with the flash turned on. The fourth segment demonstrates how to transform the colour of one source image to reduce the impact of different lightings. The fifth section discusses the process of combining the two pre-processed source images to create a natural-looking image. Finally, the sixth segment will summarise the algorithm's benefits and drawbacks.

2. Sequence of Operation

The sequence begins with the image acquisition process, which is done twice, once with the flashlight turned off and once with the flashlight turned on. These two images will provide the necessary amount of data to identify and remove unwanted information. Then, using the SIFT and RANSAC algorithms, features are extracted. After studying the characters, we move on to the crucial portion of colour transformation, which involves copying the weighted average value of nearby pixels using several reference points. Finally, the skewed information from the flash is softly combined with the weighted average value.

Figure 2. Flowchart of Sequence Operation

The example of two input pictures for the algorithm is shown in the diagram below. The two photos were taken at night in an indoor setting from two slightly different angles. The algorithm could have a wide range of applications in medicine [5]. Figure 3 shows a photograph taken without the use of a flash. As can be seen, the glass has a lot of reflections on it, which seriously reduced the image quality. Figure 4 shows a photo taken with the flash turned on. First, we must determine the location and magnitude of the flash spot in the photograph. Second, since the two input images may be taken from different positions, alignment of pictures should be done. Third, the two images have colour inconsistencies due to the different light sources. Before merging the two images, we must first change the colour of one to match the colour of the other. Finally, combining the flash-spot region of Figure 3 with the rest of Figure 4 results in an artificial image. To get a natural-looking result, we'd have to find out how to soft-combine the two images.

Figure 3. Picture without Flashlight

Figure 4. Picture with Flashlight

3. Flashspot Analysis

The goal now is to use the information from figure 3 to replace the affected region in figure 4. We will need to locate the location and size of the flash spot in figure 4 to accomplish this.

3.1 Identification of Flashspot using Matched Filter

As we have seen many photos with flash spots, we can understand that the flash spots have a particular pattern – that is, they are very bright in the middle and fade away when you move away from it. This resemblance indicates that using a matched filter to locate the flash spot in a photograph taken with the flash turned on is a smart idea. We were able to conclude that using a Gaussian Filter is the best way to get rid of this castback by analysing the flash points.

The following question is how we determine the size of this Gaussian filter. Since we can't tell the extent of the flash spot ahead of time, it's unlikely that we will be able to find a Guassian filter that suits it. However, we discovered through experiments that constraining the sigma of the Gaussian filter to 1/200 to 1/20 of the width of the picture has no effect on the matching result in real time applications. Therefore, we simply set sigma = (1/100) *Width in our algorithm.

3.2 Flashspot Estimation

After you've figured out where the flash spot's core is, you'll need to figure out how big it is. The diameter we're talking about is the equivalent diameter of the flash spot, which diminishes from its middle slowly. We binarize the image with a threshold of 0.85 to exclude most of the bright areas in the image, since the flash spot is very bright in the image. Then, we obtain the binarized image's connected components and select the components that contains the flash spot centre which we got earlier. Finally, we can estimate the size of the flash spot by calculating the equivalent diameter d of this identified connected part.

SIFT

As SIFT can find distinguishing key points which are constant to spot, size and rotation, and strong to affine transformations, as well as variations in lighting, it is useful for object identification. The algorithm extracts SIFT features from the source image. These characteristics are compared to the SIFT function database created from the training images. Even though the distance ratio test mentioned above eliminates several false matches caused by background clutter, there are still matches that belong to different items.

4. Color Transformation

Our primary goal is to ensure that the original is improved while remaining normal [6]. In this case, colour transformation is extremely important. We've decided to remove the details from the second image's flash site and replace it with data from the first figure, which was taken without the flashlights turned ON. This can seem to be child's play at first, but it really works. We also try to increase the number of reference points from which we calculate the value of pixels in the target region. In addition to the colour space size, there are two other considerations to consider when locating neighbour pixels. The first consideration is the distance between the neighbouring pixel and the original point on the picture. To reduce the impact of irregular illumination in the frame, we'd like to highlight the points that are near to the original point. The distance between the neighbour point and the flash spot centre is the second element. We want the neighbour points to be far away from the flash point core because the flash point is the undesirable information in Figure 4.

5. Image Splicing

Finally, we are going to see how to combine the two images together. As we mentioned earlier,the only area which we don't require in figure 4 is the flash point i.e the bright spot. So, our idea is to use the flash point area of Figure 3 in Figure 4 to overcome the bright spot. So, with the help of Gaussian mask, we can get the combined image.

6. Results

In this paper we have discussed the pipeline of proposed algorithm. Initially, we find the position and extent of the flash point in the image with the help of template matching and the connected component method. Then the images are aligned by extracting the features and matching it. Next, we resolve the color inconsistency in the two pictures due to difference in the light source. At last, we combine both images in order to get the resultant image with maximum reflections removed. But still there is scope for improvement in Digital Image Processing methods. Whatever is the extent to which an algorithm is said to remove the reflections reaching ideal result is a far cry. However, by experimenting. We understood that in real time applications, if we set sigma value of Gaussian filter between 0.005 to 0.05 of the width of the image, it doesnt have impact on the matching result. Therefore, here we use sigma as 0.01 of width of picture. As we have already discussed through the course of our project, two images can also be taken from any angle. As the images may be of different angles, aligning them before combining is must. We have used standard algorithms like SIFT

descriptors and RANSC with holography model. Both SIFT and RANSAC best serves the purpose. Thus, we can only conclude that out of all the novel techniques available, cast back scuttle on handheld device produces the best results.

7. Conclusion

Thus, the process of cast back scuttle has been divided into various stages and each segment is dealt with individually. First the image is obtained, then it is aligned, followed by reading in flash image, converting it into a binary image, and once we convert it to a binary image, we resort to creating a Gaussian mask to perform colour transformation to do reflection removal and eventually combine the images to obtain the processed or desired image.

References

[1]   Tsuji, T., & Inoda, R. (2010). Removal of Specular Reflection Based on High-Speed Camera Images. IEEJ Transactions on Industry Applications, 130(3), 261–267. https://doi.org/10.1541/ieejias.130.261

[2]   Basri, R., & Jacobs, D. W. (2003). Lambertian reflectance and linear subspaces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(2), 218–233. https://doi.org/10.1109/TPAMI.2003.1177153

[3]   Zhong Zhang, Takeshi Enomoto, Tetsuo Miyake, & Takashi Imamura. (2007). Blind source separation by combining indepandent component analysis with complex discrete wavelet transform. 2007 International Conference on Wavelet Analysis and Pattern Recognition, 549–554. https://doi.org/10.1109/ICWAPR.2007.4420731

[4]   Schölkopf, B., Smola, A., & Müller, K.-R. (1998). Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural Computation, 10(5), 1299–1319. https://doi.org/10.1162/089976698300017467

[5]   Bo Pang, Zhang, D., & Kuanquan Wang. (2005). The bi-elliptical deformable contour and its application to automated tongue segmentation in Chinese medicine. IEEE Transactions on Medical Imaging, 24(8), 946–956. https://doi.org/10.1109/TMI.2005.850552

[6]   Sato, Y., & Ikeuchi, K. (n.d.). Temporal-color space analysis of reflection. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 570–576. https://doi.org/10.1109/CVPR.1993.341073

[7]   Wolff, L. B. (n.d.). Using polarization to separate reflection components. Proceedings CVPR '89: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 363–369. https://doi.org/10.1109/CVPR.1989.37873

[8]   Yoon, K., Choi, Y., & Kweon, I. S. (2006). Fast Separation of Reflection Components using a Specularity-Invariant Image Representation. 2006 International Conference on Image Processing, 973–976. https://doi.org/10.1109/ICIP.2006.312650

[9]   Chaurasiya, R., & Ganotra, D. (2021). An Overview of Learning Approaches in Reflection Removal (pp. 411–425). https://doi.org/10.1007/978-981-15-7907-3_31

[10] Farid, H., & Adelson, E. H. (1999). Separating reflections from images by use of independent component analysis. Journal of the Optical Society of America A, 16(9), 2136. https://doi.org/10.1364/JOSAA.16.002136

[11] Naejin Kong, Yu-Wing Tai, & Shin, J. S. (2014). A Physically-Based Approach to Reflection Separation: From Physical Modeling to Constrained Optimization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(2), 209–221. https://doi.org/10.1109/TPAMI.2013.45

Fundus Image Classification Using Convolutional Neural Network

U.Savitha [a], Kodali Lahari Chandana [a,1], A.Cathrin Sagayam [a], S. Bhuvaneswari [a]

[a] *Department of Electronics and Communication Engineering,*
Vel Tech HighTech Dr Rangarajan Dr Sakunthala Engineering College, Chennai, TN,
India

Abstract. Different eye disease has clinical use in defining of the actual status of eye, in the outcome of the medication and other alternatives in the curative phase. Mainly simplicity, clinical nature are the most important requirements for any classification system. In the existing they used different machine learning techniques to detect only single disease. Whereas deep learning system, which is named as Convolutional neural networks (CNNs) can show hierarchical representing of images between disease eye and normal eye pattern.

Keywords. Fundus image disease, glaucoma, Macular edema, Eye lesion, machine learning, Deep learning technique, image processing.

1. Introduction

The Eye, it is the visual organ and the most important of the special senses. We know the situation is very horrible when we don't have sight. Blindness knows no geographic, economic, cultural bonds. It will have no concern for age or gender [1]. Blindness cases are progressing on a daily basis. The NPCB was appeared in 1976, to reduce prevalence of blindness from 1.4 to 0.3%. The vision 2020 "The Right to Sight" may be a worldwide scheme to scale back averted blindness [2]. They mainly targeted cataract, refractive error childhood blindness, corneal blindness, glaucoma is shown in Figure 1. Some eye problems may be small, but some other may lead to loss of vision [3-6]. Common eye problems which were mainly refractive errors, cataracts, optic nerve disorders, Glaucoma. Retinal disorders etc. Mainly eye diseases were due to use of modern medicines and treatment methods, this was noted by Edward Kondort MD, CCH, DHt (human homeopathic ophthalmologist)

1.1 Macular Edema: It is the build from fluid within the macula, which is that the area within the centre of the retina. The retina is present at the rear of the attention which is light sensitive tissue and macula is that the a segment of the Retina, which is liable for precise and straight-forward vision. The macula stretches and elongates as a result of the fluid buildup, in that we would lose our sight.

U.Savitha, Department of Electronics and Communication Engineering, Vel Tech HighTech Dr Rangarajan Dr Sakunthala Engineering College, Chennai, India

Email: laharichandanakodali@gmail.com.

*1.2 Glaucoma***:** It is disease that damages the eye's nervous optics. It happens when fluid builds up ahead a part of the attention. This fluid buildup a pressure, which is understood as intraocular pressure (IOP), is high. Increase within the vital sign which causes the loss of blindness. it's a grave disease and each case that's untreated ends in blindness.
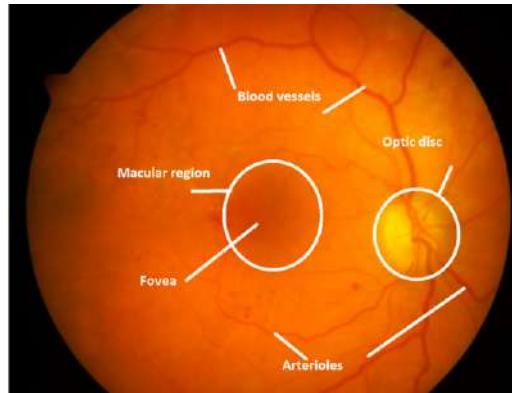
Figure 1. Eye glaucoma

*1.3 Eye lesion***:** Begin lesion have several characteristics that differ them from malignant lesions, a number of the kinds begins which include squamous papilloma, keratoacanthoma, hereditary begin intraepithelial dyskeratosis etc. It is often treated by surgical excision, cryotherapy or topical chemotherapy**.**

2. Existing System

Diabetes may be a disease during which, an individual fails to respond to insulin released by their pancreas, or pancreas doesn't produce sufficient insulin [7]. Early detection of diabetic disease is completed through advanced machine learning techniques over manual detection. This paper presents a scientific survey of automated approaches to diabetic disease from several aspects, which are named as Available data sets, image preprocessing techniques, deep learning models and performance evaluation metrics. The synopsis of diabetic disease detection is given by our survey.

ADVANTAGES:

1) Fast treatment results.
2) Accurate results.

DISADVANTAGES:

1) It detects only analysis of a diabetic disease.

3. Proposed System

In Our paper, we presented a deep learning technique designed to detect the disease. In this proposed system, deep learning model is implemented to detect various eye diseases like macular edema, glaucoma and eye lesions [8]. One of the foremost famous deep learning algorithms most commonly utilized in image classification is a

convolutional neural network. In general, this consists of three layers namely convolutional layer, pooling layer, and fully connected layer is shown in Figure 2. The CNN design picks up an image as input, which is then carried via the layers to detect features and identify the image, and yielding a response.

ADVANTAGES:
Here we classify different disease like macular edema, eye lesion and glaucoma detection.
1) Highly reliable.
2) Fast treatment results.
3) Accurate results.
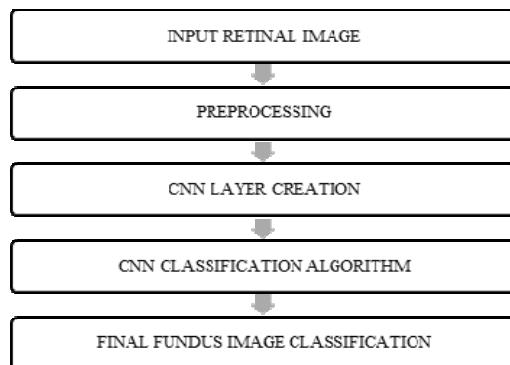
Figure 2. Proposed System Block Diagarm

3.1.1 Input Retinal Image:
We begin by downloading the images into MATLAB. Dataset's area unit holds on in many alternative file varieties. This image is stored as binary files that MATLAB will quickly use and reshape into pictures**.**

3.1.2 Preprocessing:
Data pre-processing may even be a method of making ready the information and creating it appropriate for a machine learning model [9-16]. The data typically consists of noises, mission values, maybe unusable formats that can't be used directly for machine learning models. Pre-processing is used to cleansing the information and creating it appropriate for a machine learning model. This will increase the accuracy and potency of the machine learning models.

3.1.3 CNN Layer Creation:
The Convolutional Neural Networks (CNN) is one in all the foremost notable deep learning algorithms and therefore, this was ordinarily utilized in image classification applications. This algorithmic software takes an input image and runs it via the levels to classify choices and recognize the image, resulting in a classification outcome. This design usually involves alternating convolutional layers and pooling layers, after that it follows by a slew of fully connected layers. At regular intervals, the output is the input of the next layer. In Figure 3, the CNN input is also a three-dimensional image (width $\times$ height $\times$ depth), the breadth, peak area unit is the size of the images. The depth is the number of input nodes, which are three color channels Red, Green, and Blue (RGB).

This layer brings out options via pictures. Every convolutional has weighted matrices and the area unit referred to as filters or kernels that float across the input part of image to explicit data. The CNN's first layers look for colors and simple patterns. Thus, at intervals the next layers step by step find a lot of complicated patterns. To generate a function map, each filter performs a convolution process.

Figure 3. Proposed CNN Layer Creation

4. Equations

We have a picture "x", which may be a 2D array of pixels with different color channels (Red, Green and Blue-RGB) and that we have a feature detector or kernel "w" then the output we get after applying a mathematical process is named a feature map.

$$s[t]=(x * w)[t]=\sum_{a=-\infty}^{a=\infty} x[a]w[a+t]$$

Feature map   Input   kernel

(1)

This process aids in determining the correlation between two signals. The feature detector will recognize edges within the images. Thus, the operation can assist us in identifying the sides within the image until such a filter is applied to the image. These functions are usually assumed to be zero everywhere but their values are stored at a finite number of points. Thus, this helps us to implement a sum of a given range of array element.

$$S(i,j) = (I * K)(i,j) = \sum_{m}\sum_{n} I(m,n)K(i-m,j-n)$$

(2)

Where I is two dimensional array is Kernel function
The above equation is rewritten as cross correlation function, since the convolution function is independent

$$S(i,j) = (K * I)(i,j) = \sum_{m}\sum_{n} I(i-m,j-n)K(m,n)$$

(3)

For Example ,if we saw a person or an image in a dark how could we identify the image? We would just turn on our flashlight and sacn across the image.That is why we neutralize convolutional layer.

5. Results
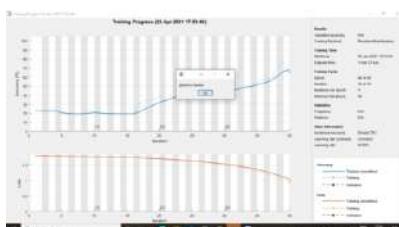
Figure 4. Glaucoma
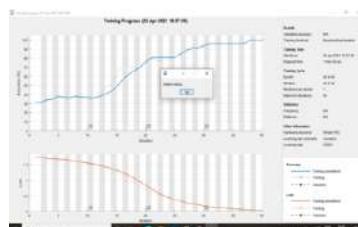
Figure 5. Healthy Eye

Figure 6. Macular Edema

Figure 7. Eye Lesion

Eye lesion: In the above Figure 7 results, eye lesion can be found out by seeing the pop-up message as diabetic lesion. Accuracy obtained 92% and the loss is 0.25.

Macular Edema: In the above Figure 6 results, macular edema can be found out by seeing the pop-up message as severe.
Accuracy obtained 100% and the loss is 0.

Glaucoma: In the above Figure 4 results, glaucoma can be found out by seeing the pop-up message as glaucoma disease.
Accuracy obtained 69% and the loss is 1.1.

Healthy Eye: In the above Figure 5 results, healthy eye can be found out by seeing the pop up message as healthy stating.
Accuracy obtained 100% and the loss is 0.

6. Conclusion

In our proposed system we have implemented a method for the classification of diabetic eye diseases using Convolutional Neural Network. Our method primarily identifies diseases such as macular edema, glaucoma, eye lesion and healthy eye. We classify the images as regular, extreme or abnormal based on the measurements of exudates from the macula. We also tested many datasets, which gave us excellent results. Our proposed method, which is integrated into a real-time patient management system, will perform automated diagnoses on medical images for diabetic complications

References

[1]   Park, K. (2015). Park's Textbook of Preventive and Social Medicine 23rd edition, page no.723. *Bansaridas Bhanot*, 404–13. Retrieved from http://www.parasredkart.com/index.php?route=product/ product&product_id=4544

[2]   Sandhar Singh Harbakhash. Textbook of Opthalmology and Otology with Homeopathic Therapeutics, 1st edition, New Delhi, B.Jain Publishers, 2003.

[3] Todd Cooney. Homeopathy for eye disorders info (internet) (cited 9 feb 2017) available from: https://ivcjournal.com/homeopathy-eye-disorders

[4] Book Review: Handbook of Opthalmology. (1951). *Canadian Journal of Occupational Therapy*, *18*(3), 95–95. https://doi.org/10.1177/000841745101800306

[5] Sathye, S. (2017). Effect of homoeopathic preparation of Ruta graveolens on the progression of childhood myopia before, during and after cessation of treatment: A retrospective study. *Indian Journal of Research in Homoeopathy*, *11*(4), 249. https://doi.org/10.4103/ijrh.ijrh_72_16

[6] Verma. S.P. Practical Handbook of Surgery with Homeopathic Therapeutics, Reprinted edition, New Delhi, B. Jain Publishers(P)Ltd, 2007.

[7] Norton A.B. Opthalmic Diseases and their Homeopathic Therapeutics, 3rd edition, New Delhi, B.Jain Publishers (P) Ltd., 2005.

[8] Chand DH. Role of Homoeopathy in Opthalmological Conditions, Indian J Opthalmol, 1982, Volume: 30, Page 245- 248.

[9] A.Kunwar, S.Magotra, and M.P.Sarathi, "Detection of high –risk macular edema using texture features and classification using SVM classier," in proc. Int. Conf. Adv. Comput., Commun. Inform. (ICACCI), Aug. 2015, pp. 22852289

[10] S. J. J. Kumar and C. G. Ravichandran, ``Macular edema severity detection in colour fundus images based on ELM classier,'' in Proc.   Int.Conf. I-SMAC (IoT Social, Mobile, Anal. Cloud) (I-SMAC), Feb. 2017, pp. 926933.

[11] Wu, W., Zhang, H., Pirbhulal, S., Mukhopadhyay, S. C., & Zhang, Y. T. (2015). Assessment of Biofeedback Training for Emotion Management Through Wearable Textile Physiological Monitoring System. *IEEE Sensors Journal*, *15*(12), 7087–7095. https://doi.org/10.1109/JSEN.2015.247063

[12] Pirbhulal, S., Zhang, H., Wu, W., Mukhopadhyay, S. C., & Zhang, Y. T. (2018). Heartbeats Based Biometric Random Binary Sequences Generation to Secure Wireless Body Sensor Networks. *IEEE Transactions on Biomedical Engineering*, *65*(12), 2751–2759. https://doi.org/10.1109/TBME.2018.2815155

[13] Sodhro, A. H., Pirbhulal, S., & Sangaiah, A. K. (2018). Convergence of IoT and product lifecycle management in medical health care. *Future Generation Computer Systems*, *86*, 380–391. https://doi.org/10.1016/j.future.2018.03.052

[14] Wu, W., Pirbhulal, S., Sangaiah, A. K., Mukhopadhyay, S. C., & Li, G. (2018). Optimization of signal quality over comfortability of textile electrodes for ECG monitoring in fog computing based medical applications. *Future Generation Computer Systems*, *86*, 515–526. https://doi.org/10.1016/j.future.2018.04.024

[15] Jamal, I., Akram, M. U., & Tariq, A. (2012). Retinal Image Preprocessing: Background and Noise Segmentation. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *10*(3), 537. https://doi.org/10.12928/telkomnika.v10i3.834

MIMO Using Silicon Patch Antenna

Jagadeesh D [a,1], Karthik E [a], Gautam K [a], Gopi [a]

[a] *Department of Electronics and Communication Engineering,*
Vel Tech High Tech Dr. RR Dr. SR Engineering College, Avadi, Chennai,
TN, India

Abstract. The growth of mass communication system results the need for the high gain compactable multi-antenna base station by employing the recently introduced Multi-Input Multi-Output (MIMO) concept. A dual-band silicon patch antenna for MIMO operating at 2.4GHz and 5GHz has been designed and stimulated results are presented in this paper. The antenna designed with the dielectric constant 2.2 of material RT-duroid results in high efficiency and low substrate loss. ADS software is used for the stimulation of antenna RF-properties and other parameters. With the use of online calculator, the antenna has been designed for 33mm of length and 35mm of width.

Keywords. MIMO, ADS Software, silicon, Microstrip patch antenna.

1. Introduction

Multi-Input Multi-Output is a radio antenna communication technology uses multi antennas to receive and transmit data. Wi-Fi, LET, radio and other RF technologies are evolving MIMO concepts [8]. It provides increased link capacity and spectral efficiency. This system provides the advantages of time diversity, frequency diversity and space diversity [1]. It works under the phenomenon of multipath, i.e., the same signal is transmitter at different path through antenna and different frequency [2].

This system has divided into several types based on the number of antenna: SISO (Single Input Single Output) is the simple method which works with single rectenna. It has no diversity and other process. The SIMO (Single Input Multi Output) uses single antenna to transmit data and multiple antennas to receive it. The MISO (Multi Input Single Output) has multiple antenna to transmit data and single antenna to receive data. And the MIMO has multiple antenna to transmit and receive data.

2. Patch Antenna

Patch antenna also known as microstrip antenna, which can be printed on circuit board itself [3]. They are at low cost, low profile, lighter in weight and ease to fabricate. This antenna has two layers of fabrication, one for transmission and another for ground. The antenna is measured by length (*l*) and width (W) of the patch and thickness (h) of the substrate with permittivity($\in_r$). The height of the substrate must be smaller than the operating wavelength. The frequency of the antenna is determined the length of the

---

[1] Jagadeesh D, Department of Electronics and Communication Engineering, Vel Tech High Tech Dr. RR Dr. SR Engineering College, Avadi, Chennai, TN, India, Email: jagadeeshdamodaran@gmail.com.

patch through the formula $f_c = \frac{1}{2L\sqrt{\epsilon_0 \epsilon_r \mu_0}}$. The input impedance is controlled by the width of the patch.

3. Silicon

Silicon (Si) is a nonmetallic carbon family element. It is available at 27.7% of earth atmosphere, which is second most abundant [4]. It has the atomic number of 14, melting point of 1,410˚C and density of 2.33g/cm^2. Pure silicon is more reactive to the atmosphere, it's found as common in all rocks as silica (silicon di-oxide) with mixture of oxygen. It has only 0 and +4 oxidation state. It is mainly used as a semiconductor. By the property of conduction silicon is divided into two types. **N-type silicon,** it's chemically doped with phosphorous for the conduction. It has four electrons with our outer electron bond tightly together. When voltage applied single electron of phosphorous moves and makes negative charge. **P-type silicon,** it's doped with boron gas for receiving electron.

4. ADS

Advance Design System software, It's an   electronic design software by Path Wave Design on Keysight Technology [5]. It has a platform to stimulate computer-oriented report of radiofrequency with appropriate antenna [6-7]. The software allows to draw the layout of the antenna and the number of layer or coating with substrate. The software generates the graph with respect to frequency of the antenna working in Figure 3 to Figure 6. The software allows the dimensional analysis of the proposed system.

Antenna design

The patch antenna was sketched for 33*35mm^2 with the substrate of dielectric constant 2.2 at the thickness of 1.6mm. The patch and the ground are fabricated with silicon in Figure 1.
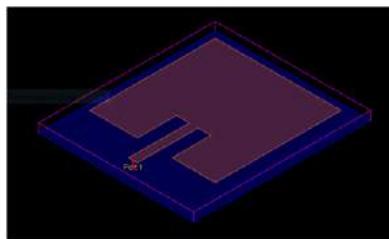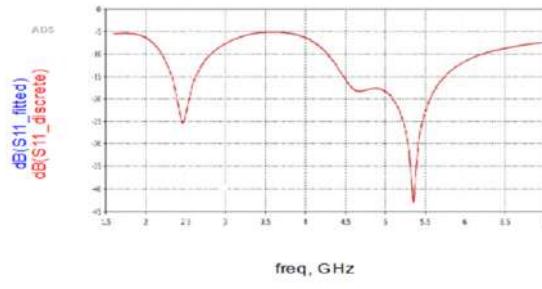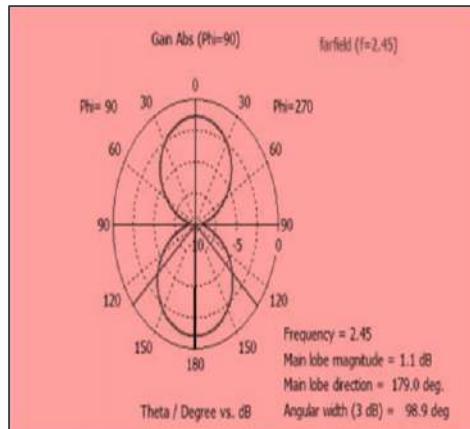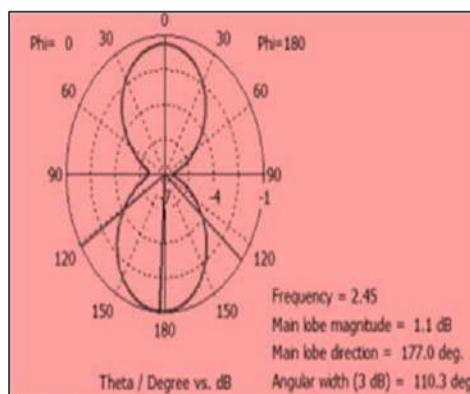
Figure 1. 2D view of antenna

RT-duroid 5880 material is used for substrate, which has good thermal and electric properties.

Figure 2. S parameter

The characteristics of the patch antenna designed have been stimulated by ADS software. Figure 2 show the graph between frequency and S_{11}.

Figure 3. 2D radiation pattern in E-plane for the frequency of 2.4GHz

Figure 4. 2D radiation pattern in H-plane for the frequency of 2.4GHz.
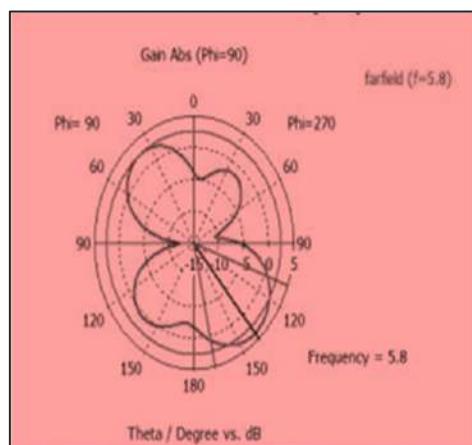
Figure 5. 2D radiation pattern in E-plane for the frequency 5.8GHz.
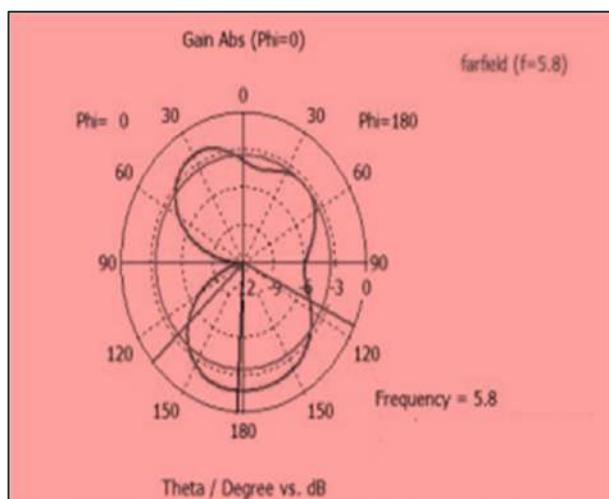
Figure 6. 2D radiation pattern of H-plane for frequency of 5.8GHz.

References

[1] Rangan, S., Rappaport, T. S., & Erkip, E. (2014). Millimeter-wave cellular wireless networks: Potentials and challenges. *Proceedings of the IEEE*, *102*(3), 366–385. https://doi.org/10.1109/JPROC.2014.2299397

[2] Chanclou, P., Cui, A., Geilhardt, F., Nakamura, H., & Nesset, D. (2012). Network operator requirements for the next generation of optical access networks. *IEEE Network*, *26*(2), 8–14. https://doi.org/10.1109/MNET.2012.6172269

[3] Fontaine, N. K., Doerr, C. R., Mestre, M. A., Ryf, R. R., Winzer, P. J., Buhl, L. L., … Lingle, R. (2012). Space-division multiplexing and all-optical MIMO demultiplexing using a photonic integrated circuit. In *National Fiber Optic Engineers Conference, NFOEC 2012*. Optical Society of America (OSA). https://doi.org/10.1364/nfoec.2012.pdp5b.1

[4] C Wiesel, A., Mestre, X., Pages, A., & Fonollosa, J. R. (2004). Efficient implementation of sphere demodulation. In *IEEE Workshop on Signal Processing Advances in Wireless Communications, SPAWC* (Vol. 2004-January, pp. 36–40). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/SPAWC.2003.1318917

[5]   Winzer, P. J. (2014). Making spatial multiplexing a reality. *Nature Photonics*. Nature Publishing
       Group. https://doi.org/10.1038/nphoton.2014.58
[6]   Miller, D. A. B. (2013). Self-configuring universal linear optical component [Invited]. *Photonics
       Research*, *1*(1), 1. https://doi.org/10.1364/prj.1.000001
[7]   Markkandan, S., Malarvizhi, C., Raja, L., Kalloor, J., Karthi, J., & Atla, R. (2021), **"Highly compact
       sized circular microstrip patch antenna with partial ground for biomedical applications"**,
       Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2021.04.480
[8]   Markkandan, S., Logeshwaran, R., & Venkateswaran, N. (2021). **Analysis of Precoder
       Decomposition Algorithms for MIMO System Design**. *IETE Journal of Research*, 1–8.
       https://doi.org/10.1080/03772063.2021.1920848

Powerful and Novel Tumour Detection in Brain MRI Images Employing Hybrid Computational Techniques

K Lakshmi Narayanan [a,1], R Niranjana [a], E Francy Irudaya Rani [a], N Subbulakshmi [a], R Santhana Krishnan [b]

[a]*Assistant Professor, Francis Xavier Engineering College, Tirunelveli, India*
[b]*Assistant Professor, SCAD College of Engineering and Technology, Tirunelveli, India*

Abstract. Brain tumour detection is an evergreen topic to attract attention in the examination field of Information Technology innovation with biomedical designing, in view of the gigantic need of proficient and viable strategy for assessment of enormous measure of information. Image segmentation is considered as one of the most vital systems for visualizing tissues in an individual. To robotize image segmentation, we have proposed a calculation to get global optimal thresholding esteem for a specific brain MRI image, utilizing OTSU+Sauvola binarization strategy. The fundamental reason for feature collection is to diminish the quantity of structures utilized in classification while keeping up satisfactory classification exactness. One of the most extra-customary procedures applied for feature extraction is Discrete Wavelet Transform (DWT). Adequately it anticipates the estimation space on a plane to such an extent that the fluctuation of the information is ideally protected. We propose a justifiable model for brain tumours discovery and classification i.e., to classify whether the tumour is benign or malignant, utilizing SVM classification. SVM utilized here deals with basic hazard minimization to group the images for the tumour extraction, and a Graphical User Interface is created for the tumour classification operation, using the MATLAB platform.

Keywords. Brain tumour, MRI image, Discrete Wavelet Transform, SVM

1. Introduction

Normal specific forms of brain tumour include gliomas and meningiomas. Be that as it may, whether they are helpful or negative is the basic category within them. MRI is an expert technique in imaging for identifying brain tumours. In either case, simple manual analysis by people of these MRI images cannot be sufficient either to reliably assess a tumour's closeness or to classify the suspected tumour as benevolent or dangerous. This could spur human errors. To maintain a strategic distance from such an erroneous inquiry and to provide a reliable procedure to break down the brain pictures naturally and to determine specifically whether or not there is a tumour and also trained algorithms are also used to determine whether a tumour is dangerous or not. Premalignant tumour is a cancer stage where cancer can be treated while ignorance can lead to cancer. The disease is known as malignant tumour.

---

[1]K.Lakshmi Narayanan, Assistant Professor, Francis Xavier Engineering College, Tirunelveli, India; Email: drkln86@gmail.com.

When it occurs, it starts to expand around neighbouring tissues after a particular period. Malignant tumours develop very rapidly and can cause a person's death. Early detection of malignant tumour can be of great benefit to the patient. That is why malignant tumour is immediately identified and marked a new way of achieving optimum precision is being investigated.

2. Methodology

2.1 Pre-Processing

The overall block diagram that depicts the process flow in the proposed technique is shown in Figure 1.
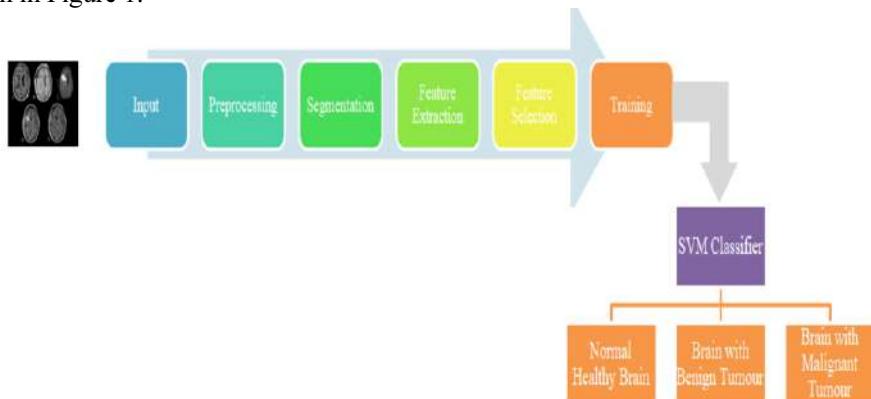
Figure 1. Process flow of proposed system

MRI images by large are subjected to alterations by a distortion of the bias field, which makes the asset of the same tissues vary throughout the picture. In order to make right this situation N4ITK was employed [1]. Whatsoever, this did not seem to justify the fact that intensity dispersal of a tissue type is a worthy match to intensity that various volunteers proved to show for similar series of MRI images [2]. In addition, it may also swerve from expected results, even if the patient is subjected to scans using the same imaging modality at different times or only when abnormality is detected[3]. So intensity normalization tactic advised by Nyúletal[3] was employed and analysed on each series in order to improve the sudden changes in intensity ranges which would be more distinct among different numerous cases. Through this form of intensity normalization, a sequence of depth symbols can be imbibed from training sets of images and are selected as defined in [4], for each MRI sequence.

2.2 Segmentation & Feature Extraction

This paper suggests a more traditional approach which is a combination of two strategies, the binarization strategy OTSU+Sauvola. The Otsu method of segmentation is now well known for its capabilities in segmentation. But when combined with Otsu the Sauvola binarization results in an excellent segmented image. In particular, this combination works amazingly well for non-uniform context, as is the case with brain MRI image. It will detect the melanoma but it also takes into account a lot of

distracting surrounding. Several image processing applications deploy Otsu's system to perform histogram technique based thresholding or perform binary thresholding [5]. The algorithm presented takes into account that the image accepts bi-modal histograms and further calculates the optimal threshold, giving rise to the above two classes in such a way that their joint distribution is negligible. This taken up to more levels leads to Multi Otsu method [6]. This is an essential part of procedure, which helps to make the algorithm more efficient and quick. Only those features which directly mark the territories for possible presence of tumor cells alone will be extracted from the images.

2.3 Classification via Machine Learning

The Support Vector Machine is the neural network employed to perform the classification process as, SVM classifier is outstandingly good for brain tumour detection [7]. Using the features extracted during the preceding development period, the SVM is trained for a broad data collection [8]. Having to reconsider the situation that might arise wherein the data under analysis has 2 individual and distinct classes, Support Vector Machine (SVM) should be the choice undertaken [9]. We use a collection of new brain images of the MRIs is shown in Figure 2. According to SVM technique, most applicable hyper plane is selected in such a way that it lets go of entire set of data dots, belonging to a single class, from the other class.
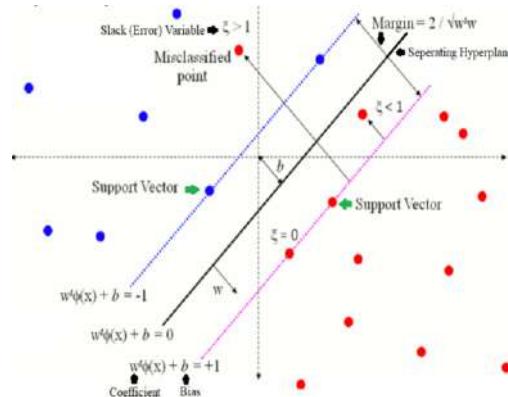
Figure 2. Support Vector Machine

2.4 Accuracy Features

Radial Basis Function Accuracy- For data that is dispersed in all dimensions, interpolation of the radial basis function can be very efficient [10-16]. RBFs are capable of offering exceedingly precise numerical resolutions for a wide range of classes.

Linear Accuracy - When it comes to employing linear motion, accuracy and replicability are two features that have to be taken into account. For this a detailed account of the element causing the linear motion and its parameters must be taken into consideration.

Polygonal Accuracy - Instead of employing pixel dots, polygon structures can be used for data for drilling and testing the algorithms. The polygonal data can be transformed into raster standards, which in turn should be within the limits of the features. Across each individual polygon structure the resident accuracy measures can be obtained.

Quadratic Accuracy –m For medical image analysis, the quadratic accuracy measures proved to be more precise than the linear interpolation based accuracy. Rationalizing the numerator of the quadratic equation provides certain level of precision.

3. Experimental Results

To train the computational algorithm, a sufficiently large dataset of the MRI brian images is used. The algorithm is being implemented in version MATLAB 2017. To train the computational algorithm for effective classification a discerning matrix is created in Figure 3 to Figure 5. Classification, most essential parameters like sensitivity, specificity and accuracy were premeditated using following carefully designated formulas which constitute the decision matrix

- TP- True Positive: this involves the case where affected brain is accurately recognized as affected.
- TN- True Negative: this involves the case where normal brain is properly recognized as being normal and unaffected.
- FP - False Positive: this involves the case normal unaffected brain is wrongly classified as affected by tumor brain.
- FN- False negatives: this involves the case brain affected by tumor, whether benign or malignant has been wrongly classified as normal and unaffected.

$Sensitivity = TP/(TP+FN) (*100\%)$

$Specificity = TN/(TN+FP) (* 100\%)$

$Accuracy = (TP+ TN)/(TP+ TN+FP+FN) (*100 \%)$

These are the parameters chiefly employed to analyse the classifiers efficiency in terms of performance in Table 1.
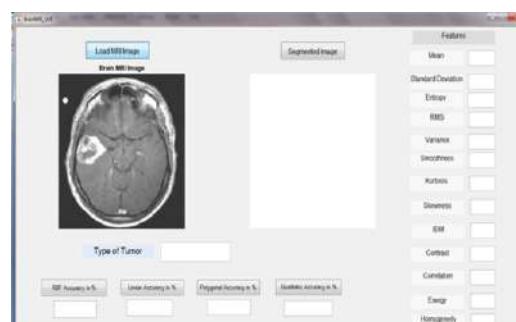
Figure 3. MRI Image Loaded in to GUI

Figure 4. MRI image Segmentation using OTSU & Classification
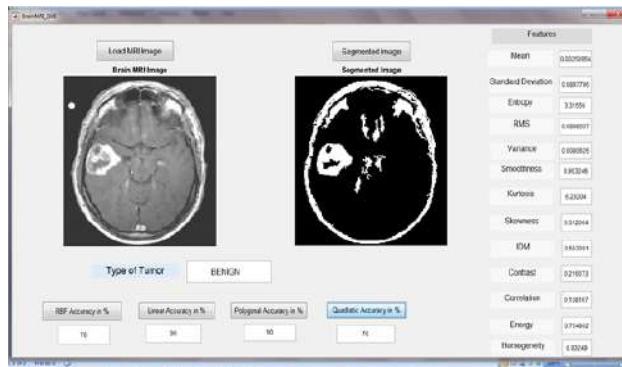
Figure 5. Accuracy Calculation

Table 1. SVM Classifier outputs

Kernel Function	Accuracy	Sensitivity	Specificity
Linear	90.87%	84.43%	98.9%
Quadratic	85.43%	74.33%	99.2%
Polynomial	86.5%	76.23%	99.39%

4. Conclusion

Brain MRI images of this new method have proven to be an effective means of identifying brain tumour. The hybrid methodology of combining contemporary Otsu with Sauvola algorithm proved to provide better segmentation of areas of interest, in the images. Support vector machine for classification provides accurate results for the identification of the brain tumour. The result proves that with precise training input SVMs can distinguish between tumour spots and fittingly categorize them as a benign tumour, malignant tumour or healthy brain tissue. Computational advantages for SVMs are noteworthy. In future research, various data mining modules could be employed for training various kernel functions. This would greatly lift up the efficiency of classifier, and more number of dataset images can be employed for training and testing purposes.

References

[1] Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

[2] Shah, M., Xiao, Y., Subbanna, N., Francis, S., Arnold, D. L., Collins, D. L., & Arbel, T. (2011). Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. *Medical Image Analysis*, *15*(2), 267–282. https://doi.org/10.1016/j.media.2010.12.003

[3] Nyúl, L. G., Udupa, J. K., & Zhang, X. (2000). New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging*, *19*(2), 143–150. https://doi.org/10.1109/42.836373

[4] Nyú, L. G., & Udupa, J. K. (1999). On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine*, *42*(6), 1072–1081. https://doi.org/10.1002/(SICI)1522-2594(199912)42:6<1072::AID-MRM11>3.0.CO;2-M

[5] Zhu, N., Wang, G., Yang, G., & Dai, W. (2009). A fast 2D otsu thresholding algorithm based on improved histogram. In Proceedings of the 2009 Chinese Conference on Pattern Recognition, CCPR

2009, and the 1st CJK Joint Workshop on Pattern Recognition, CJKPR (pp. 319–323). https://doi.org/10.1109/CCPR.2009.5344078

[6]   LIU Jian-zhuang, Li Wen-qing, "The Automatic threshold of gray level pictures via Two-dimentional Otsu Method", *Acta Automatic Sinica*,1993.

[7]   Haralick, R. M., Dinstein, I., & Shanmugam, K. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-3*(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314

[8]   Srivastava, D. K., & Bhambhu, L. (2010). Data classification using support vector machine. *Journal of Theoretical and Applied Information Technology*, *12*(1), 1–7.

[9]   Himanshu makkar and Aditya Pundir, "Image Analysis using improved Otsu's thresholding method", *Int. Journal on Recent and Innovation Trends in Computing andCommunication*, pp. 2122-2126.

[10]   Narayanan, K. L., & Ramesh, G. P. (2017). VLSI architecture for multi-band wavelet transform based Image compression and Image reconstruction. *Journal of Engineering and Applied Sciences*, *12*(Specialissue2), 6281–6285. https://doi.org/10.3923/jeasci.2017.6281.6285

[11]   K.Lakshmi Narayanan, Dr.G.P.Ramesh "Image Compression using Frequency Band Suppression in VLSI Design based Discrete Wavelet Transform" in *Int. Journalof Control Theory and Application*, **10** (36), 133-139, 2017.

[12]   Niranjana.R, Francy Irudaya Rani.E, Manoj P, Naveen Nagarajan S, Raaja P, "Analysis And Detection Of WBC Cancer Cells Using Random Forest Classifier",*Int. Journalof Emerging Technology and Innovative Engineering*, vol **A5**, Issue 9, September 2019 (ISSN: 2394 – 6598)

[13]   Francy Irudaya Rani. E, Niranjana. R, Abirami.M, Grace Priya.A, Indiranatchiyar.A "Brain Tumor Detection Using Ann Classifier", *Int.Journal of Emerging Technology and Innovative Engineering*, vol **5**, Issue 9, September 2019 (ISSN: 2394 – 6598)

[14]   E.Francy Irudaya Rani, S.Ramu, T.Nisha, S.Kavi Sundari, "Predicting a Superlative Classifier for the Identification of Blood Cancer in Human", *Int. Journal of Advanced Research Trends in Engineering and Technology* (IJARTET), vol. **7**, Issue 9, September 2020

[15]   E.Francy Irudaya Rani, M.Srimathidevi, E.Sherina, G.Sivasankari, M.Saranya, "A Competent Algorithm for the Detailed Analysis of Skin Images from Dermoscopy", *Int. Journal of Advanced Research Trends in Engineering and Technology* (IJARTET), vol. **7**, Issue 9, September 2020.

[16]   Lakshminarayanan, K., Muthukumaran, N., Robinson, Y. H., Shanmuganathan, V., Kadry, S., & Nam, Y. (2021). Deep Learning-Based Hookworm Detection in Wireless Capsule Endoscopic Image Using AdaBoost Classifier. Computers, Materials and Continua, 67(3), 3045–3055. https://doi.org/10.32604/cmc.2021.014370

Firefighter Safety Using IoT

Caroline Jebakumari S [a,1], Aarthi K [a], Darshan S [a], Karan S [a]

ªDepartment of Electronics and Communication Engineering, Easwari
Engineering College, Chennai, TN, India

Abstract. From ancient times, firefighter's faces a high risk of burns which may sometimes lead to death, especially those who enter the burning building first and those who are holding the front end of the nozzle. In order to avoid this kind of risk, an enlightened physical admonition arrangement for firefighter was put forward. Corresponding controller is developed with low power consumption on the display screen terminator which could inspect numerous conflagration in the surroundings and the site where firefighter reside as substantially as their critical bodily data so that the admonition might be provided with a low transmitting error rate. Moreover, the suitable information processing computation is sorted out to assure the preciseness. Consequently, competent firefighter's individual protection is augmented considerably with strong anti-interference. In addition to this, a buzzer will be there on the side of firefighter like alarm when parameters like temperature, blood pressure, heart rate becomes abnormal so that they might not take the risk considering their body conditions.

Keywords. Multi-sensor, physical warning, low power consumption, cloud

1. Introduction

The difficulty of fire control became hopeless as the firefighter uniform were somewhat retrogressive as substantially as the shutdown edifice are augmented at present. The conflagration control became recalcitrant even in recent times as two firemen lost their lives while fighting the flames in the Madurai textile shop. In order to reduce this kind of incidents this system has been preferred. The Fire department uniform has to be upgraded as this was crucial to assure the firefighter protection, which has been examined in the earlier instance. Thus, an enlightened physical caution scheme for firefighter premised on IoT has been envisaged. The physiological information of the fireman and the site where the fireman remains will be collected by this system and will be forwarded to display screen via cordless procedure. Omni bearing oversees task for the firefighter individual protection, preferable protection factor and servicing execution are accomplished by utilizing this network. As considerably as such method might arrive at the worldwide sophisticated tier to the point of a firefighter functioning orderliness and existence aid. While integrating sensing element in a firefighter's uniform, the conflagration surroundings and the site might be recognized consequently the ones might evade damage to the utmost. In order to give a detailed study about the controllers, this paper has been organized into the following sections.

- Segment 1 set out the introduction for this proposed methodology and it describes the related works which deal with various types of sensors and its

---

[1] Caroline Jebakumari S, Department of Electronics and Communication Engineering, Easwari Engineering College, Chennai, TN, India; Email: caroline.s@eec.srmrmp.edu.in.

advantages and disadvantages compared to the previous designs and performance.

- Segment 2 describes the proposed methodology and its advantages.
- Segment 3 characterizes outcome and examination of the projected system.
- Segment 4 account termination and prospect work for fire fighter's safety using IoT.

Existing System

T. Brunschwiler [1] proposed a system which measures the pressure in firefighter while learning in heated workout locale and provided a task to obtain specialist mark. It is also concentrated on Chronic-Pulmonary-Disease. While this system measures the stress in firemen which is the major advantage but still there are many disadvantages this system hasn't proposed which includes physiological parameters when the firefighter is in danger.

Sarita Gupta [2] insists that fire is the major cause for accidental in this World so a system was proposed that were destined to live up to consumer guideline that has a micro - controller, RF modules and two types of sensors [7]. If a flame were ascertained, the information and Uniform resource locator address of damaged locale is forwarded to the conflagration supervision department via The Global System for Mobile Communications unit and without delay the salvage functioning will take place. The above technology does not reveal the nature or physiological parameters of firefighter.

2. Proposed System

Temperature spotting, pulse oximeter, concentration of carbon monoxide spotting and IoT module are the physiological confluence that this paper reflects on as dominant core [3]. The warning system keeps track of a firefighter's physical environment in real time through a focal point confluence the particulars are communicated to display screen fumigator. Apart from the earlier points, cloud keeps track of a firefighter's physiological conditions and the surroundings where firefighter remains. The latitude and longitude where the firefighter stays will also be monitored and when there is any abnormal state in any one of the physiological activities, the person from the outside will go for rescue operation and there will be a buzzer insisting that the firefighter is in danger, so that the firefighter gets indicated about his physiological activities. The algorithm that this system follows, advantages that this system possesses and the components used are explained below.

2.1 Fire Fighter Algorithm

The steps specified below is the algorithm for this proposed system,

• The Firefighter who is inside the building or a place which is on fire will be having different kinds of sensors inside firefighter's suit in order to monitor different body parameters.

• The values from the sensor will be updated in a cycle at a certain period of time in an URL link. The one who is monitoring from outside can monitor those values.

• The temperature sensor (LM35) has been set to a limit of 40°C. When the surrounding temperature of the firefighter is below 40°C it displays as "TEMP_NORMAL" in the URL link. When the surrounding temperature goes above

40°C, it displays as "TEMP_ABNORMAL".

• Gas sensor (MQ-02) will show a value of 1 on the screen and displays "GAS_NORMAL" when there are no harmful gases in the surrounding environment of the firefighter. When it detects any harmful gas, then it displays "GAS_ABNORMAL" and the value changes to 0.

• Pulse Oximeter (MAX30100) displays "HB_NORMAL" when the heart rate > 10 and displays "HB_ABNORMAL" when the heart rate < 10. Blood pressure sensor (Mercury sphygmomanometer), displays "BP_NORMAL" when the blood pressure > 80 and displays "BP_ABNORMAL" when blood pressure < 80.

• The locale of the firefighter gets updated in the URL link at a certain interval of time. It displays location as latitude and longitude coordinates in Figure 1. All the sensor values which are mentioned in the condition can be changed and can be used for different circumstances. While showing these values in a table the date and time gets updated.
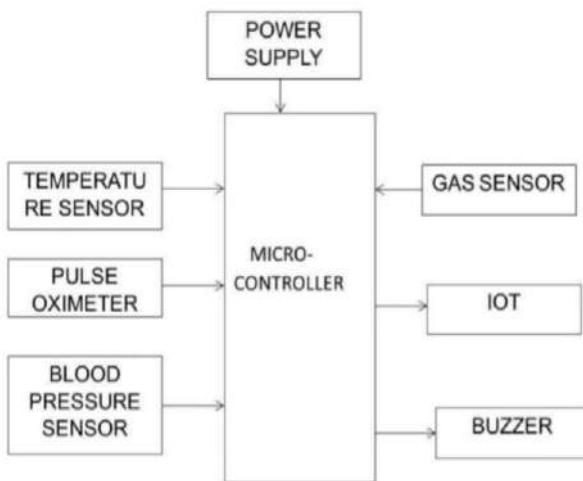
Figure 1. Block Diagram of Proposed System

2.2 Advantages

The advantages that this paper hold are when the firefighters in scorch forests, it is not easy to know the location when he is in the hazard [4]. Using this system, the locale can be easily found, so that someone from the external can salvage the firefighters who are in peril. When deleterious gases like Carbon monoxide or some other gases is sensed via gas sensor, then it will be updated to the person external. So that later on, no one will get in that spot or will walk inside with some indispensable pre-emptive actions [5]. When pernicious gases surround the fire fighters or when the encircling temperature goes very severe the person who is overseeing from the outside will know that, but to specify that to the firefighter inside, a buzzer is kept within the firefighter [6].

2.3 Components Used

The components used here are power supply, temperature sensor (LM35), blood pressure sensor, gas sensor (MQ-02), pulse oximeter (MAX30100), as it operates from

4V to 30V, the scale factor for LM35 is 0.5°C and ranges from -55°C to 150°C. ATmega328P can be used in real time bio-metrics, robotics and academic applications. MAX30100 is a well conceived pulse oximeter sensor. Mercury sphygmomanometer being used for measuring blood pressure sensor. MQ-02 is the best for gas sensor as it detected the hydrogen gas leakage and portable gas.

3. Results and Analysis

While the power supply is given, the ARDUINO board shows the HB, G, BT, BP where HB indicates the Heartbeat value, G represent the gas sensor value, BT represents the temperature value, BP represents the blood pressure value. According to the situation the value varies. Initial values that are specified in the controller in Figure 2.

Figure 2. Display image of the micro-controller.

Figure 3 shows the updated sensor values of the system. The Table 1 has the following entries which show whether temperature, gas, pulse rate; blood pressure is normal or abnormal for the fire fighter. In case of any abnormal condition sensor 5 in the Table 1 indicates an EMERGENCY situation so that the person who is monitoring from the external will get alerted. Here, the Date & Time will also be pointed out while the sensor values are getting processed.

Table 1. Displays the physiological nature of the fire fighter

#	Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5	Sensor 6	Sensor 7	Sensor 8	Date & Time
1	TEMP_NORMAL	GAS_NORMAL	HB_ABNORMAL	BP_ABNORMAL	EMERGENCY	null	null	null	2021-03-05 13:33:45
2	TEMP_NORMAL	GAS_NORMAL	HB_ABNORMAL	BP_ABNORMAL	EMERGENCY	null	null	null	2021-03-05 13:33:14
3	TEMP_NORMAL	GAS_NORMAL	HB_ABNORMAL	BP_ABNORMAL	EMERGENCY	null	null	null	2021-03-05 13:31:35
4	TEMP_NORMAL	GAS_NORMAL	HB_ABNORMAL	BP_ABNORMAL	EMERGENCY	null	null	null	2021-03-05 13:30:41
5	TEMP_NORMAL	GAS_NORMAL	HB_ABNORMAL	BP_ABNORMAL	EMERGENCY	null	null	null	2021-03-05 13:29:43

Latitude and longitude of the firefighter are laid out in the Table 2, so that the local of the firefighter can be easily found when the firefighter is in peril.

Table 2. Displays the latitude and longitude along with date and time

Date & Time	Longitude	Latitude
2021-03-02 16:53:56	80.2116077	13.0341263
2021-03-02 16:54:07	80.2116077	13.0341263
2021-03-02 16:54:49	80.2116077	13.0341263
2021-03-02 16:55:04	80.2116077	13.0341263
2021-03-02 16:55:18	80.2116077	13.0341263
2021-03-02 17:27:50	80.2117393	13.0348844
2021-03-02 17:29:51	80.2117393	13.0348844
2021-03-02 17:29:53	80.2117393	13.0348844
2021-03-02 17:43:47	80.212179	13.0343248
2021-03-02 17:53:42	80.212179	13.0343248

Finally, the complete hardware setup of the proposed system is shown in Figure 3. The complete setup which includes power supply adapter, power supply, Arduino board, temperature sensor (LM35), blood pressure sensor, pulse oximeter, gas sensor, IoT and a buzzer. Here, the result of the proposed system is discussed and analyzed.
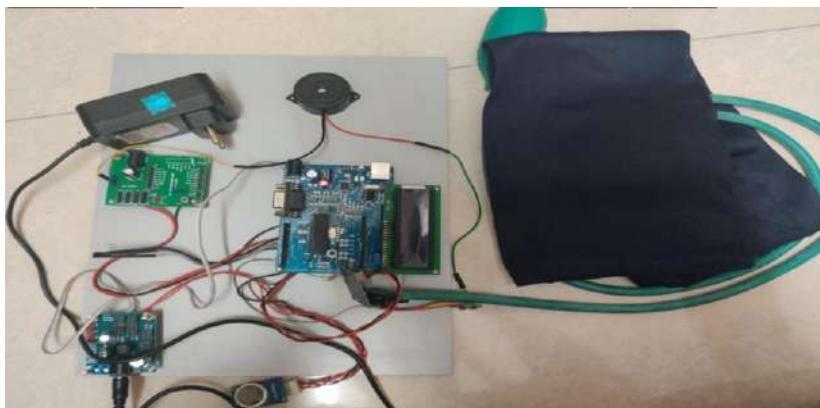
Figure 3. Project Kit Image

4. Conclusion

The physical warning system for firefighter is done in this paper with less network technology, IOT, sensor, embedded and computer technology to produce this intelligent system. It collects information from the firefighter like temperature, blood pressure, pulse and the gas which firefighter is inhaled and according to the values the buzzer rings to the firefighter simultaneously the values are shown to the one who is monitoring from outside those values will be saved in an URL address. The result of this system has many edges which will help the firefighter who are working in a dangerous position like building, forest and quarries. Thus, an equipment for firefighter

safety using IOT has been done to protect the firefighter who are in danger. In the future, GPS which is in-built at present can be improved like even when there are many floors in buildings in which floor or in which room the firefighter is present can be found. The sensors can also be covered with thick sheet or metal to withstand more temperature. These are the expected improvisation in the future for the proposed system.

References

[1] Hackett, S., Cai, Y., & Siegel, M. (2019). Activity Recognition from Sensor Fusion on Fireman's Helmet. 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 1–6. https://doi.org/10.1109/CISP-BMEI48845.2019.8966005

[2] Brunschwiler, T., Weiss, J., Paredes, S., Sridhar, A., Pluntke, U., Chau, S. M., Gerke, S., Barroso, J., Loertscher, E., Temiz, Y., Ruch, P., Michel, B., Zafar, S., & van Kessel, T. (2019). Internet of the Body - Wearable Monitoring and Coaching. 2019 Global IoT Summit (GIoTS), 1–6. https://doi.org/10.1109/GIOTS.2019.8766409

[3] Gupta, S., Mudgil, A., Bhardwaj, P., & Gupta, M. (2016). Design and Development of Automatic Fire Alert System. 2016 8th International Conference on Computational Intelligence and Communication Networks (CICN), 632–636. https://doi.org/10.1109/CICN.2016.130

[4] N. Brushlinsky, M. Ahrens, S. V. Socolow, P. Wagner, "World Fire Statistics. Technical report 24" Centre of Fire Statistics Int. Assoc. of Fire and Rescue Services, 2019.

[5] Jiang, W., Chen, C., & Cai, Y. (2016). The design on intelligent physical warning system for fireman. 2016 International Conference on Audio, Language and Image Processing (ICALIP), 694–698. https://doi.org/10.1109/ICALIP.2016.7846531

[6] R. G. Babu, A. Nedumaran and A. Sisay, "Machine Learning in IoT Security Performance Analysis of Outage Probability of link selection for Cognitive Networks," *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2019, pp. 15-19, https://doi.org/10.1109/I-SMAC47947.2019.9032669

[7] R. G. Babu, V. V. Nathan, J. Bino, C. Amali and S. Ganesh, "IoT Security Enhancement with Automated Identification Device using IOT SENTINEL," *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2021, pp. 518-523, https://doi.org/10.1109/Confluence51648.2021.9377165

Shopping Wagon: A Smart Shopping System Using RFID for Shopping Malls

S.Deepa [a,1], Amala Nihila.A [b], Prabhavathi.J [b], Meenatchi.M [b], Varsha.M.J [b]

[a]Professor, Department of Electrical and Electronics Engineering, Panimalar Engineering College, Chennai, TN, India
[b]UG Student, Department of Electrical and Electronics Engineering, Panimalar Institute of Technology, Chennai, TN, India

Abstract. This project describes a supermarket automation trolley based on an RFID reader. .The trolley is equipped with an RFID reader and an electronic hardware system to make the transaction more convenient. The RFID card, whose price is set into the reader, is used to correct those items that are above a certain number. The value of the item is added to the sales bill and shown on the LCD monitor when the item is shown in front of the reader. The trolley car is programmed in such some way that it'll move consistent with the user command. It additionally has the supply for removing the things from the trolley car wherever price is aloof from the overall cost. The user can view their bill through IOT along with the number of items purchased and total bill amount. The user can also pay the bill using his card provided and the system will give an alert if the total amount exceeds the amount in the card. The system will also suggest the user whether the purchased product is suitable for their health condition or not through IOT app provided to the user.

Keywords. Microcontroller ATMEGA328P, RFID Tags, RFID reader, Liquid Crystal display, IOT.

1. Introduction

The Robotized Shopping Cart, "Shopping wagon", is a modern patron shopping product this is designed to assist shoppers' fast-tune their buying revel in. The idea of this clever buying wagon will revolutionize the shopping revel in of each buyer [1]. In these days each grocery store and shopping center uses purchasing baskets and purchasing carts to accumulate the gadgets from the racks. In the actual global we are facing many issues at the same time as status in a queue due to the fact we need to wait for a protracted time [2]. When we're in buying shops it's far simpler for us to take the buying cart and save anything we need however whilst we need to wait in a queue for billing the product that's a busy process. As a result, billing takes longer [3]. The RFID reader and IR sensors will be used in the wagon which is able consequently scan the item being entered into the cart. Once the item has been put into the Wagon, the information related to the item will be shown on the LCD screen introduced on the wagon or the smartphone of the shopper [4]. The data of the items put in cart will be

---

[1] Amala Nihila.A, UG Student, Department of Electrical and Electronics Engineering, Panimalar Institute of Technology, Chennai, TN, India; Email: amalanihila1998@gmail.com.

communicated to the charging counter of the shop by means of IOT. This will lead to automatically billing up the items the client has put into the cart. Once the shopping of the client has been over, the ultimate charge will be shown on his smartphone [5]. This will radically diminish the shopping time of the client and make the shopping involvement less demanding [6]. The said conception will not only change the shopping experience of the buyer but it will also have great impact on the shop keepers. They will have access to real-time information about particulars that are being bought. The above idea will enhance consumer courting management. Introduction of this clever carts can be quite attractive withinside the retail zone and could result in a brand-new generation of purchasing and promoting of goods. It may be taken into consideration an affordable desire for capability retails for the reason that key hassle it addresses is saving the consumers time [7-9].

2. Existing System

In existing system, automation system reaches its goal only until the billing system at shopping malls even in developed cities. There is no any automated system to get and give all kinds of groceries to the customer based on his/her grocery list. To solve this kind of problem and take the automation to next level, we are proposing an idea by that customer can order from at any place by using smart phone app [10-12].

3. Proposed System

In the residing scene invention is growing grade by grade in numerous fields like faux astute, AI, digital reality, touch commerce, net of things, etc. The precept purpose of the paper is middle to the purchaser wishes and purpose because time is greater essential to anybody in reality. But the humans spend the greater time within the supermarket. Consumers, for instance buy a certain amount of goods in the store using a trolley. After making a purchase, they can encounter issues such as standing in a long line in the billing section and not knowing the total cost of the products purchased. So, we implement a system using IOT which consists of RFID reader and LCD display, while customer takes an object and placed within the trolley, that time, the reader scan the item and the amount of the product is added to the bill through IOT Server and it will automatically updated to the store in charge through IOT server and consumer also knows the value of money they spend on the items before billing [13]. So, they can add or remove according to the convenience and also they can pay the bill from their card itself using indoor localization [14]. The system also has automatic trolley set up which can be operated using IOT control switches. The system will also suggest the user whether the purchased product is suitable for their health condition or not through IOT app provided to the user in Figure 1.
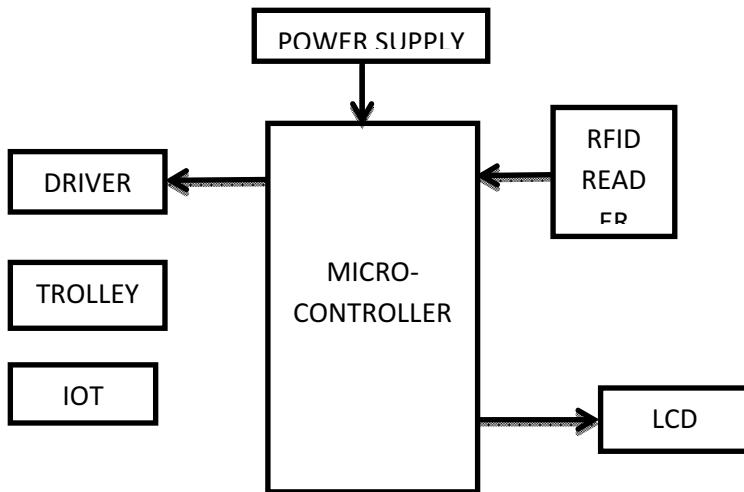
Figure 1. Shows the block diagram of the proposed work

4. Results and Discussion

In the arrangement there is a driver circuit - streetcar instrument present. This driver circuit has a 12 V ground supply. It has 4 information and 4 yield pins. The info is acquired from the regulator and the yield is associated with the engine. So relying upon the regulator input the engine yield changes. The code is given so that the engine pushes ahead, in reverse, turns and stops. . So for the proficient running of engines, driver circuit is utilized. Next RFID is available and it deals with UART communication (serial communication).

Figure 2, a UART communication, two UARTs discuss straightforwardly with one another. The sending UART changes over equal information from a controlling gadget like a central processor into sequential structure, communicates it in sequential to the accepting UART, which at that point changes over the sequential information back into equal information for the getting device in Figure 3. In this arrangement every one of the item's subtleties is perused utilizing RFID peruser and every one of the items has a secret string esteem remarkable to it. This information is then passed to the regulator and if the put away information is gotten the subtleties get shown in the LCD show. The code for expansion in includes and sum is put away in the microcontroller and it is shown in a LCD show of size 16x2(2 lines &16 coloumns) in Figure 4. The application chips away at IOT and the inside association is given utilizing a username and secret key. Here likewise information transmission happens through UART protocol (serial communication). Here remote exchange to cloud happens. A battery with 12 V is utilized as force supply and the force for all segments is given from it. The live updates of the item subtleties can be seen in the application.

Figure 2. Shows the final output of developed product

Figure 3. Homepage of IOT application

Figure 4. Real Time Sensor Values

Figure 5. Real Time Sensor Product Values

Figure 6. Next Product Sensor Values

Next product is added and number of products gets updated to 2 and then the total amount gets incremented by 100 after biscuit is purchased. Now the total amount is Rs.200. In Figure 6, the last product which the user adds is a paste of Rs.50. Now the total amount is incremented to Rs.250 and the number of products is 3. Since the user is done with the shopping she can pay Rs.250 via the RFID card provided to her.

5. Conclusion

The advanced product is comfortable to operate, low-budget and do now no longer need any unique training. Using RFID technology, this venture simplifies the billing process, makes it faster, and improves protection. This elevates the general shopping experience to new heights. A design for automated retail shop rested on RFID is established in this journal. This ground plan uses RFID technology to simplify, speed up, and improve the security of the billing process. Require the right direction toward the per user to be perused and any harm makes them mixed up. Also utilizing standardized tags is a human-escalated movement at the retail location. Consequently, RFID innovation is the better way to deal with defeat these deficiencies.

References

[1] Gross, H. M., Boehme, H., Schroeter, C., Mueller, S., Koenig, A., Einhorn, E., … Bley, A. (2009). TOOMAS: Interactive shopping guide robots in everyday use - Final implementation and experiences from long-term field trials. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009* (pp. 2005–2012). https://doi.org/10.1109/IROS.2009.5354497

[2] Endres, H., Feiten, W., & Lawitzky, G. (1998). Field test of a navigation system: Autonomous cleaning in supermarkets. In *Proceedings - IEEE International Conference on Robotics and Automation* (Vol. 2, pp. 1779–1781). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ROBOT.1998.677424

[3] The elements of user experience: user-centered design for the Web and beyond. (2011). *Choice Reviews Online*, *49*(01), 49-0321-49–0321. https://doi.org/10.5860/choice.49-0321

[4] Gross, H. M., Meyer, S., Scheidig, A., Eisenbach, M., Mueller, S., Trinh, T. Q., … Fricke, C. (2017). Mobile robot companion for walking training of stroke patients in clinical post-stroke rehabilitation. In *Proceedings - IEEE International Conference on Robotics and Automation* (pp. 1028–1035). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICRA.2017.7989124

[5] Gross, H. M., Scheidig, A., Müller, S., Schütz, B., Fricke, C., & Meyer, S. (2019). Living with a mobile companion robot in your own apartment - Final implementation and results of a 20-weeks field study with 20 seniors. In *Proceedings - IEEE International Conference on Robotics and Automation* (Vol. 2019-May, pp. 2253–2259). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICRA.2019.8793693

[6] Marin-Hernandez, A., Hoyos-Rivera, G. D. J., García-Arroyo, M., & Marin-Urias, L. F. (2012). Conception and implementation of a supermarket shopping assistant system. In *Proceedings of Special Session - Revised Papers, 11th Mexican International Conference on Artificial Intelligence 2012: Advances in Artificial Intelligence and Applications, MICAI 2012* (pp. 26–31). https://doi.org/10.1109/MICAI.2012.21

[7] Gruen, T. W., Corsten, D. S., & Bharadwaj, S. (2002). *Retail Out of Stocks : A Worldwide Examination of Extent , Causes , and Consumer Responses*. Business (pp. 1–62). Retrieved from http://gbspapers.library.emory.edu/archive/00000035/01/GBS-MKT-2002-001.pdf

[8] Paolanti, M., Sturari, M., Mancini, A., Zingaretti, P., & Frontoni, E. (2017). Mobile robot for retail surveying and inventory using visual and textual analysis of monocular pictures based on deep learning. In *2017 European Conference on Mobile Robots, ECMR 2017*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ECMR.2017.8098666

[9] Kumar, S., Sharma, G., Kejriwal, N., Jain, S., Kamra, M., Singh, B., & Chauhan, V. K. (2014). Remote retail monitoring and stock assessment using mobile robots. In *IEEE Conference on Technologies for Practical Robot Applications, TePRA*. IEEE Computer Society. https://doi.org/10.1109/TePRA.2014.6869136

[10] Frontoni, E., Mancini, A., Zingaretti, P., Contigiani, M., Bello, L. D., & Placidi, V. (2018). Design and test of a real-time shelf out-of-stock detector system. *Microsystem Technologies*, *24*(3), 1369–1377. https://doi.org/10.1007/s00542-016-3003-3

[11] Dhanesh, L., Deepa, S., Elangovan, P., & Prabhu, S. (2020). Enhanced and Energy-Efficient Program Scheduling for Heterogeneous Multi-Core Processors System. In *Lecture Notes in Electrical Engineering* (Vol. 665, pp. 737–747). Springer. https://doi.org/10.1007/978-981-15-5262-5_55

[12] Maheswari, V., Elangovan, P., Baranidharan, M., Deepa, S., & Dhanesh, L. (2019). Theoretical and simulation analysis of first generation DC-DC converters. *International Journal of Advanced Science and Technology*, *28*(19), 72–78.

[13] Deepa, S., Dhanesh, L., & Elangovan, P. (2019). Optimal fuzzy controller forpower quality improvement of dynamic voltage restorer using bacterial foraging algorithm. *International Journal of Advanced Science and Technology*, *28*(19), 10–15.

[14] C. Jain, G. V. S. Sashank, V. N and S. Markkandan, "Low-cost BLE based Indoor Localization using RSSI Fingerprinting and Machine Learning," Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2021, pp. 363-367, https://doi.org/10.1109/WiSPNET51692.2021.9419388

IoT Based Solar Panel Tracking System with Weather Monitoring System

K. Dinesh [a,1], Lakshmi Priya. A [b], Preethi. T [b], Sandhya. M [b], Sangeetha. P [b]

[a]*Professor, Department of Electrical and Electronics Engineering*
[b]*UG Student, Department of Electrical and Electronics Engineering*
Panimalar Institute of Technology,Chennai,TN, India

Abstract. Solar power is the burgeoning method of continual energy. The assignment is designed and carried out the use of dual axis sun tracker system. In order to maximise power era from solar, it's important to introduce sun ray monitoring systems into solar electricity production. A dual-axis tracker can boom power through monitoring solar rays from switching photovoltaic cells in various directions. These photovoltaic cells can rotate in all directions. The LDR (Light Dependent Resistor) have been used to feel the depth of mild at 30 degree every or at 180 degree general and ship the information to microcontroller. This assignment also can be used to experience rain drop, temperature and humidity using sensor and they may be displayed on LCD. We can save the Solar energy in battery.

Keywords. Photovoltaic cells, LDR (Light Dependent Resistor), IOT, Microcontroller ATMEGA 328P, Temperature sensor, Humidity sensor, Rain drop sensor.

1. Introduction

Solar trackers are growing in popularity, however now no longer all people know the entire advantages and capability hazards of the framework. Photovoltaic cells tracking following solutions are a prominent predominant innovation for mounting photovoltaic boards [1]. Stationary photovoltaic board, which maintain photovoltaic boards in a constant position, will have their productiveness compromised whilst the solar passes to a less-than-most excellent slant. In order to overcome this, sun trackers robotically pass to "track" the development of the solar throughout the atmosphere, thereby magnify the result. Tracker generate greater energy than their stationary opposite numbers because of elevated direct publicity to sun rays [2]. This increase may be as a great deal as 10 to 25% pending at the geographic region of the monitoring system. There are many special styles of sun trackers, such a single-axis and dual-axis trackers, all of which may be the perfect healthy for a completely unique jobsite. Installation size, neighborhood climate, diploma of range and electric necessities are all crucial concerns that can impact the form of sun tracker excellent ideal for a selected photovoltaic cells initiation.

---

[1] K. Dinesh, Professor, Department of Electrical and Electronics Engineering Panimalar Institute of Technology, Chennai, TN, India; E-mail:mailforkdinesh@gmail.com.

Solar trackers generate greater energy in kind of the same quantity of area wanted for constant-tilt systems, making them perfect for optimizing of land.

2. Existing System

Sun is the crucial supply of power. It is the renewable source of electricity production and it's distant liberated from esteem [3]. Now day's fossil fuels are limited. Fossil gas is high-priced and that they motive plenty of pollutants compared to solar power [4]. As we recognize that solar power is freed from price and does now no longer motive any type of pollutants. To keep away from this disadvantage of fossil gas we use sun power. Solar power is turning into a vital power supply as petroleum-primarily based totally sources grow to be extra expensive. Solar power is now ample and sun generation is developing as increasingly more human beings put sun electricity to work. Previous kind of sun monitoring machine was fixed device [5]. Solar panels which might be utilized in those structures are most effective in a single way direction. And this framework creates moo quality on the yield. As we will see there are numerous troubles that arise withinside the present sun device.

3. Proposed System

To overcome these drawbacks of existing system we propose a new Solar tracking system can rotate 180 degree on both horizontal and vertical position of panel which can optimize the solar power generation [6]. This assignment makes use of five sensors in horizontal & vertical route to experience the route of most depth of luminescence [7-10]. Its sensor will face 30 degree so the overall perspective that the device can experience is one hundred eighty degree. This device will use pic microcontroller as an important electronic to function this device and motor to rotate the solar batteries. The perspective/route of solar batteries will be exhibited on the LCD. Then LDR sensor detects the solar rays and offers the facts to microcontroller. Then microcontroller compares the depth of luminescence and it offers the command to motor. As steady with the commands motor activities the solar batteries to multiple directions. The output voltage is stored in battery and we can monitor the voltage level in IoT also displayed on the LCD. Controller can also be used to monitor the rain, temperature and humidity present in the air. In Figure 1, Automatically cleaning mechanism for the solar panel is implemented to remove the dust partials present in the surface of panel.
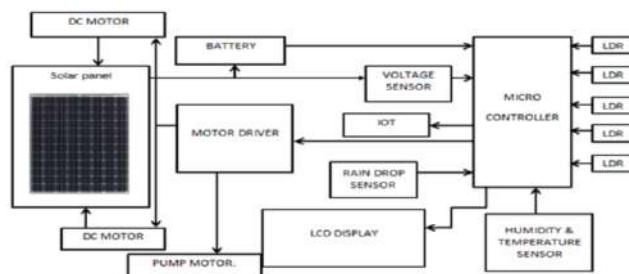
Figure 1. Represents the block diagram of the proposed work

Figure 2. Vertical direction 180 degree   **Figure 3.** Towards sun ray intensity

Figure 2 and Figure 3, It displays the vertical 180^0 direction of the photovoltaic cells respective to the sun ray intensity on the LDR.
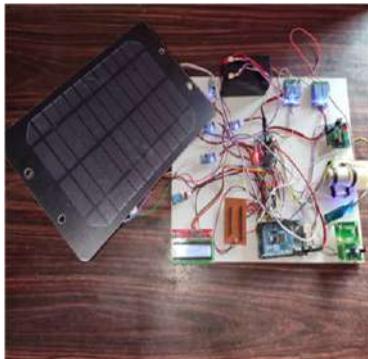
Figure 4. Horizontal 180 degree Direction   **Figure 5.** Direction to Sun ray intenstiy

Figure 4 and Figure 5, It shows the horizontal 180^0 direction of the photovoltaic cells corresponding to the sun ray intensity on the LDR.

4. Results and Disscusion

In this arrangement , the battery initially has 9V and provides power to the compnonents in the prototype. The LDR(Light Detetector Sensor) detects the intensity of solar ray in multiple direction and pass on the information to the microcontroller ATmega328P. The ATmega328P detects the highest range of solar ray direction and command the DC motor to change the direction of solar batteries corresponding to the direction.

The power produced from the photovoltaic cells is stored in the battery. Additionally this arrangement consists of humidity sensor, water level sensor, temperature sensor and voltage sensor in Figure 6. These sensor detects the value of temperature, humidity, water level in the atmosphere and the voltage produced by the panel. Figure 7, the resultant value from these sensors are displayed in the IOT server. They are also displayed on the LCD in digital mode. Additionally the prototype also

consist of pump motor which sprays water and clean the solar panel every 2 minutes in Figure 8.

Figure 6. The prototype of developed product

Figure 7. LCD
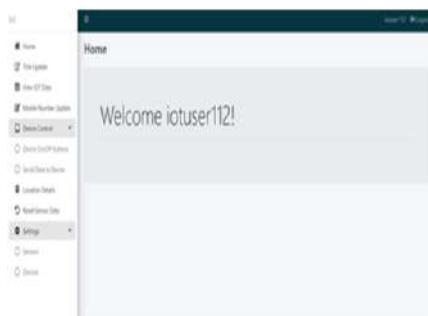
Figure 8. Homepage of IOT application

Figure 9. Server page

In Figure 9, the IoT server displays the temperature sensor, humidity sensor and water level sensor in tabular column. The second column shows the value of temperature sensor,the fourth column displays the value of humidity sensor and the sixth column displays the water level sensor count. The server also displays the date and time of the value measured.

6. Conclusion

This Created framework can be executed in several solar plant stations to measure the precise strength created from the ones devices. Also, this assignment may be carried out within side the far flung desolate tract areas. Solar tracking system is greater efficient than each other regular solar board. It is feasible to lessen the control catastrophe to many tremendous extent.

References

[1]   Marinescu, D., & Marinescu, C. (2006). Control Optimizing Algorithm for Soft Sun-Trackers. 2006 IEEE International Conference on Automation, Quality and Testing, Robotics, 1, 54–57. https://doi.org/10.1109/AQTR.2006.254496

[2] Cotton, M. F., & Rabie, H. (2015). Maintenance of antiretroviral efficacy in children. The Lancet HIV, 2(4), e120–e121. https://doi.org/10.1016/S2352-3018(15)00040-5

[3] Visconti, P., Ventura, V., Tempesta, F., Romanello, D., & Cavalera, G. (2011). Electronic system for improvement of solar plant efficiency by optimized algorithm implemented in biaxial solar trackers. 2011 10th International Conference on Environment and Electrical Engineering, 1–4. https://doi.org/10.1109/EEEIC.2011.5874639

[4] Angelkov, D., Koceska, N., & Koceski, S. (2014). Low-cost dual-axis system for solar tracking. 2014 3rd Mediterranean Conference on Embedded Computing (MECO), 169–172. https://doi.org/10.1109/MECO.2014.6862685

[5] Veligorskyi, O., Kosenko, R., & Stepenko, S. (2014). High-efficiency solar tracker development and effectiveness estimation. 2014 IEEE International Conference on Intelligent Energy and Power Systems (IEPS), 153–158. https://doi.org/10.1109/IEPS.2014.6874169

[6] Hammad, B. K., Fouad, R. H., Ashhab, M. S., Nijmeh, S. D., Mohsen, M., & Tamimi, A. (2014). Adaptive control of solar tracking system. IET Science, Measurement & Technology, 8(6), 426–431. https://doi.org/10.1049/iet-smt.2013.0293

[7] Kulkarni, A., Kshirsagar, T., Laturia, A., & Ghare, P. H. (2013). An Intelligent Solar Tracker for Photovoltaic Panels. 2013 Texas Instruments India Educators' Conference, 390–393. https://doi.org/10.1109/TIIEC.2013.76

[8] Alqarni, M., & Darwish, M. K. (2012). Maximum power point tracking for photovoltaic system: Modified Perturb and Observe algorithm. 2012 47th International Universities Power Engineering Conference (UPEC), 1–4. https://doi.org/10.1109/UPEC.2012.6398443

[9] Hinov, N. (2017). New method for designing serial resonant power converters. 060019. https://doi.org/10.1063/1.5014013

[10] Hensel, S., & Marinov, M. B. (2018). Comparison of Algorithms for Short-term Cloud Coverage Prediction. 2018 IX National Conference with International Participation (ELECTRONICA), 1–4. https://doi.org/10.1109/ELECTRONICA.2018.8439356 D. Marinescu, C. Marinescu. Control Optimizing Algorithm for Soft Sun-Trackers. Automation, Quality and Testing, Robotics, 2006.

Quantification of Blood Cells and Blood Disease Detection Using Image Processing

K.Rahimunnisa [a,1], V.Aparna [a], R.K. Harrini [a], K.Kamalini [a]

[a] Department of Electronics and Communication Engineering, Easwari Engineering College, Chennai, TN, India

Abstract. RBC (Red Blood Cells) and WBC (White Blood Cells) are the main constituents of blood. WBC fight infections by attacking bacteria and viruses, that invade the body, while RBC transports oxygen in the body. Many blood diseases can be detected using RBC and WBC count values. Immunity-related blood diseases like Leukopenia and Leukocytosis can be easily detected using the WBC count value. The manual counting method of blood cells in laboratories takes at least one day to get the blood results, which becomes a major drawback for healthcare sectors to diagnose the disease at the right time. More expensive pathological tests are also a major drawback. Accurate counting of blood cells is essential in the accurate diagnosis of the disease. The proposed system is used to calculate the RBC and WBC Count, Total blood Count, RBC and percentage and the blood disease (Leukocytosis, Leukopenia) from the input blood smear image. This will help laboratories to perform the counting of blood cells with high accuracy and less workload. This is achieved by pre-processing that involves grayscale conversion, image enhancement, noise removal, binary conversion of input image, followed by plane extraction and threshold-based Segmentation. The blood disease (Leukocytosis and Leukopenia) is detected using WBC percentage-based classification methodology. This approach obtained an accuracy of 98.4%, specificity of 88.889%, precision of 99.58%, F - Measure of 99.50%. Morphological operations are implemented using MATLAB software.

Keywords. Threshold-based Segmentation, Image enhancement, Blood cell Quantification, Blood disease detection, Image Processing, Leukocytosis, Leukopenia.

1.   Introduction

Blood transports gases, nutrients, and other macromolecules throughout the body. Blood constitutes of blood cells (RBC, WBC, Platelets) and plasma. Blood cells play a major role in detecting blood disease. RBCs contains hemoglobin, a protein that carries oxygen to the tissues and collects carbon dioxide. WBCs are a key part of the immune system and defend against infection. The WBCs are larger than RBCs. WBCs are purple in color and RBCs are red in color. Disease based on WBC count directly deals with immunity-related diseases and detecting only the WBC count-related blood diseases from the input blood smear image is the scope of this paper.

---

[1] V.Aparna, Department of Electronics and Communication Engineering, Easwari Engineering College, TN, India; Email: aparnaradha05@gmail.com.

A lower WBC count indicates Leukopenia, a blood disease due to lower immunity in the body that may result in the increased risk factor of getting a high level of infections. A Higher WBC count may indicate Leukocytosis, a blood disease due to inflammatory response that leads to cancerous condition in the blood.

2. Existing Methodology

In existing methodology [1], Image is pre-processed and blood cell segmentation is done. Feature extraction is done with the help of morphological operations and boundary is detected using Sobel edge detection. Counting is done using Circular Hough Transform method. The drawback of this methodology is the Sobel edge detection will not detect accurate edges for smooth and thin edges. It detects one or more blood cells in the image based on similarity criterion and leads to a less efficient system. The conventional method of finding out RBC and WBC count is done by collecting small sample of blood. The manual method using Hemocytometer and the automated method using Automated Hematology Analyzer. Based on the count value given from the CBC tests, doctors will diagnose the blood disease and treatment is given accordingly.

3. Proposed Methodology

The proposed methodology involves quantification of blood cells and detection of WBC-related blood diseases using image processing techniques wherein the input image is obtained from the blood-smeared samples and is converted to a grayscale image using the Image processing toolbox in MATLAB. Image Enhancement is done for better segmentation of the cells, by removing the noise and thus the quality of the image is improved. Threshold-based segmentation is carried using Otsu's Method. The blue plane is extracted from the image to extract WBC and the red plane is extracted from the image to extract RBC and other planes are not considered as they contain less information about the WBC present in the image and it is converted to binary scale through thresholding. The area of region and boundary is calculated for the detected RBCs and WBCs respectively. The RBC and WBC count, total count, RBC and WBC percentage values are found .Based on the WBC percentage value, the WBC-related blood diseases are detected. The block diagram is depicted in Figure 1.
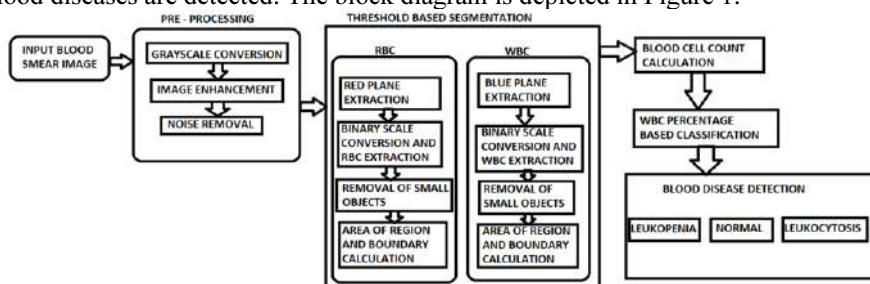
Figure 1. Block diagram of the proposed methodology

3.1 Pre-processing

 Blood drops are smeared across a microscope slide and the images are converted to digital when the microscope is interfaced with the computer. A dataset containing 250

blood-smeared digital images is taken as input images. Pre Processing improves the image data which suppresses noise and enhances features for further processing. Grayscale images are much easier to work with and perform morphological operations and segmentation. Image enhancement is done with the help of histogram equalization technique to improve the contrast the effectively spreads the most frequent intensity values in the image. The noise in the blood smear image degrades the performance of image processing algorithms and may lead to inaccurate counting of blood cells. The median filter is used for noise removal in the image.

3.2 Threshold based Segmentation (Otsu's Method)
Threshold-based segmentation is applied to segment the WBC and RBC accurately, which is achieved by implementing Otsu's Method of threshold-based segmentation [2]. As the intensity distribution of RBC and WBC are distinct, the global threshold technique is used to ideally segment the cells with high accuracy. Otsu's method [3], is used to calculate a single threshold value which separates pixels into foreground and background. When WBC is segmented, then those corresponding pixels become the foreground and the RBC becomes the background and vice versa. This threshold is determined by maximizing the inter-class variance. The algorithm can also be extended to multi-level Thresholding.

3.2.1 Plane Extraction, Binary scale conversion and Blood cells extraction
The blue plane is extracted from the image to extract WBC and the red plane is extracted from the image to extract RBC and other planes are not considered as they contain less information about the WBC present in the image.
In order to extract WBC from the image, purple pixels are considered. Based on Otsu's algorithm, a threshold value of 29 in the blue plane is fixed to extract the purple pixels. Similarly, to extract RBC from the image, a threshold value of 6 in the red plane is fixed to extract the red pixels.

3.2.2 Area of Region and Boundary Calculation
The area property of region props tool is used to calculate area by returning the number of pixels for every 8-connected object for each detected WBCs and RBCs in the binary image. The ismember function creates a binary image with objects whose area exceeds 100 pixels. The perimeter of the detected WBC and RBC in the binary image is found using bwperim the boundary of the detected WBC and RBC are drawn in green color to highlight the detected cells to the viewer. The number of WBCs is found using bwconncomp, by counting the number of connected components. It uses default connectivity of 8 pixels for binary image and returns the connected components as a structure. And thus, the WBC count value is found and displayed. The BoundaryBox property in the regionprops is used to create a bounding box around the detected RBCs in the binary image and its coordinates are returned in the form of an array. The size of the array returns the actual detected RBCs count in the binary image.Total blood cell count is calculated by the following formula,
TBC Count = WBC Count + RBC Count
Based on the total blood cell count, RBC and WBC percentage are calculated. In each and every case, the round-off value is only taken and stored in the corresponding terms.

3.3 WBC Percentage Based Classification

The WBC Count in mm3 is related to WBC count in the given blood smear image as follows [4], $= C$ x 3000, where N denotes WBC count in cubic millimeter and C denotes Count of WBC in an image. The normal range of WBC count is from 5000/mm3 to 10000/mm3 and is equivalent to a range of 20% to 40% of WBC present in the blood sample. Leukopenia is detected when WBC count is lesser than 5000/mm3 or equivalent WBC percentage lesser than 20%. Leukocytosis is detected when WBC count is greater than 10000/mm3 or an equivalent WBC percentage greater than 40%. When the count lies between 5000/mm3 and 10000/mm3 or between WBC percentage of 20% to 40%, then the blood sample is free from any disease and is considered normal. Based on this, the input blood-smeared images are classified into Leukocytosis detected image, normal and Leukopenia detected image.

4. Results and Discussion

The proposed system is implemented in the image processing toolbox of MATLAB software. BCCD v3 Database of blood cells is used wherein 250 blood smear images are used. The input blood-smeared image as shown in Figure 2, is converted to grayscale and is enhanced using the Histogram equalization technique. Noise present is removed by the median filter and is further processed. The blue and red planes are extracted to segment WBCs and RBCs separately as shown in Figure 3 and Figure 7. Based on the threshold value set by Otsu's Method, the image is converted to binary as shown in Figure 4 and Figure 8. Smaller detected blood cells (RBC and WBC) are removed as shown in Figure 5 and Figure 9 and the area and boundary is calculated, and detected cells are highlighted in green shown in Figure 6 and Figure 10. The RBC and WBC count, TBC Count, WBC and RBC Percentage and the detected blood disease (Normal, Leukopenia, Leukocytosis) are mentioned in the output GUI as shown in Figure 11, Figure 12, and Figure 13.
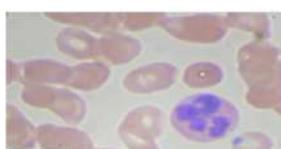
Figure 2. Input image   **Figure 3.** WBC extraction   **Figure 4.** Binary Scale conversion
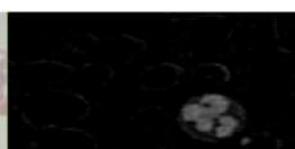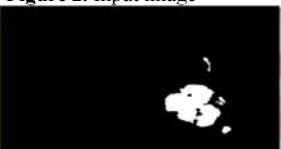
Figure 5. Noise removal   **Figure 6.** WBC highlighted in Green   **Figure 7.** RBC extraction
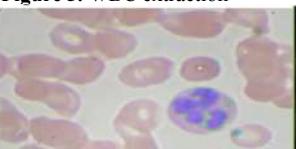
Figure 8. Binary Scale conversion   **Figure 9.** Noise removal   **Figure 10.** WBC highlighted in Green
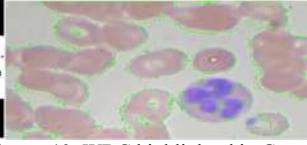
Figure 11. Normal condition **Figure 12.** Leukopenia condition **Figure 13.** Leukocytosis condition

5. Performance Analysis

Accuracy is calculated using the overall performance of the threshold-based segmentation of segmenting the blood cells and producing accurate blood cell count in Table 1. Specificity is ratio of expert identified blood cells to the total blood cells present in the image. Precision is the measure of correctly identified blood cells from all the predicted blood cells. F – Measure shows that the algorithm correctly identifies real threats and is not distributed with the falsely predicted values. The accuracy ,specificity, precision and F-Measure can be calculated by identifying the number of images as correctly identified positive pixels -True Positive (TP), correctly identified negative pixels - True Negative (TN), falsely identified positive pixels - False Positive (FP), and falsely identified negative pixels - False Negative (FN).The accuracy, specificity, Precision and F-Measure is calculated as follows,

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN), \text{Specificity} = TN / (FP + TN)$$
$$\text{Precision} = TP / (TP + FP), \text{F} - \text{Measure} = 2 / ((1/\text{Precision}) + (1/\text{Sensitivity}))$$

Table 1. Comparison of various performance metrics of the existing and proposed method.

Method	Accuracy (%)	Specificity (%)	Precision (%)	F - Measure (%)	
Existing Method [1]	91.667	87.5	-	-	
Existing Method [5]	94.58	-	-	-	
Existing Method[6]	92	-	-	-	
Existing Method [7]	98.4 - WBC	-	89.7 - WBC	93.9 – WBC	
Proposed Method	98.4	88.889	99.58	99.50	

6. Conclusion and Future Works

Immunity is one of the most important factors in the human body to fight infections in the body. Immunity-related blood diseases can be detected using the WBC count, and when left unnoticed, this condition may lead to Leukemia (Blood cancer). This method helps the health care to detect the disease accurately with fewer efforts in a cost-effective manner. The future scope of this work is to detect additional blood diseases based on the cell count value, so that early diagnosis of asymptomatic blood diseases can also be found and treated before it becomes worse. This approach obtained an accuracy of 98.4%, specificity of 88.889%, precision of 99.58%, F - Measure of 99.50%.

References

[1]. Chadha, G. K., Srivastava, A., Singh, A., Gupta, R., & Singla, D. (2020). An Automated Method for Counting Red Blood Cells using Image Processing. Procedia Computer Science, 167, 769–778. https://doi.org/10.1016/j.procs.2020.03.408

[2]. Zaitoun, N. M., & Aqel, M. J. (2015). Survey on Image Segmentation Techniques. Procedia Computer Science, 65, 797–806. https://doi.org/10.1016/j.procs.2015.09.027

[3]. Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

[4]. Alomari, Y. M., Sheikh Abdullah, S. N. H., Zaharatul Azma, R., & Omar, K. (2014). Automatic Detection and Quantification of WBCs and RBCs Using Iterative Structured Circle Detection Algorithm. Computational and Mathematical Methods in Medicine, 2014, 1–17. https://doi.org/10.1155/2014/979302

[5]. P. R. P. (2014). COUNTING OF WBCS AND RBCS FROM BLOOD IMAGES USING GRAY THRESHOLDING. International Journal of Research in Engineering and Technology, 03(04), 391–395. https://doi.org/10.15623/ijret.2014.0304071

[6]. Di Ruberto, C., Loddo, A., & Puglisi, G. (2020). Blob Detection and Deep Learning for Leukemic Blood Image Analysis. Applied Sciences, 10(3), 1176. https://doi.org/10.3390/app10031176

[7]. Gulpreet Kaur Chadha, Aakarsh Srivastava, Abhilasha Singh, Ritu Gupta, Deepanshi Singla, An Automated Method for Counting Red Blood Cells using Image Processing, International Conference on Computational Intelligence and Data Science (ICCIDS 2019), Procedia Computer Science 167 (2020) 769–778.

Optical Switch Performance for Electronic Applications

V.Reji [a,1], Bhavadharani N [a], Shridhar.B [a], K.Chaitra Sri [a]

[a]*SRM Institute of Science and Technology, Chennai, Tamil Nadu, India*

Abstract. In this paper, various optically controlled switch performances are studied and analysed. Different types of optically controlled materials are considered in this paper to control the electrical frequency with a modified and simpler structure. Roger and Fr-4 materials are used as substrate material for the basic system design with the dielectric constant of 3.0 and 4.4 respectively. Materials like silicon, germanium, graphene, and polymers are considered for analysis in the proposed system and the semiconductor metals are etched in the middle of the copper strip. The thickness of the copper is 0.008 mm with the standard conductivity of 5.814e7 S/m and the length and width of the copper strip are 2.54x2.54mm.A Copper strip is printed on the substrate to test the performance of the switches. The operating frequency of the given optical signal of silicon is around 1GHz to 25GHz, graphene is around 1GHz to 30GHz, germanium is around1GHz to 30GHz and polymer is around 1GHz to25GHz. The S11 parameter of all the proposed systems is analysed by sonnet simulation software.

Keywords. Silicon, Germanium, Substrate, Polymer, Semiconductor

1. Introduction

Many kinds of research are being made in recent days to produce highly effective switches with several materials and methods [1]. Illuminated optical controlled GCPW switch has been designed with low insertion loss up to 50GHz with induced plasmas [2] and a QPSK model has been demonstrated for transverse electric polarization with 4*4 DLN non-blocking silicon switch for optical circuit and packet switching [3]. Another micro fluid switch with compressed micro fluid channels has been developed in [4] for better conductivity. Doherty power amplifier and frequency tuning turntable mechanism have been developed at 1.8 to 3.4 GHz for power amplification and matching networks [5].Graphene oxide with dispersed hydrogen bond liquid crystal was used for optical switching and optical and electrical properties are achieved by etching dry crystalline silicon with solar cell by short circuiting with extraordinary output and storage [6]. Hydrolation wave spectrometer using telecom with heterodyne detection using THz measures the phase delay result from a single sheet of paper alternatively a resonance cavity has been used in [7] with increased sensitivity.

---

[1] V.Reji, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India;
Email: rejiv@srmist.edu.in.

The proposed system compares the optical performance of semiconductors and polymers in the electrical domain. This study is useful for developing a switch or sensor that should be operated in both the electrical and optical domains. The aim of our project is to make a small optically controlled electrical switch model with different semiconductor metals and polymers and analyse their performance.

2. Basic design and materials

The basic design consists of a substrate, a strip with a switch design on the top, and infinite ground in the bottom [8-10]. The bottom ground and top layer are made up of copper material with a standard thickness of 1.6mm and a conductivity of 5.814e7 S/m. Two strip of copper films are placed on the substrate. In Between these copper films, semiconductors such as silicon, germanium, UV-polymer, and graphene are placed and simulated as shown in figure 1.

| (a) | (b) | (c) | (d) |

(e)
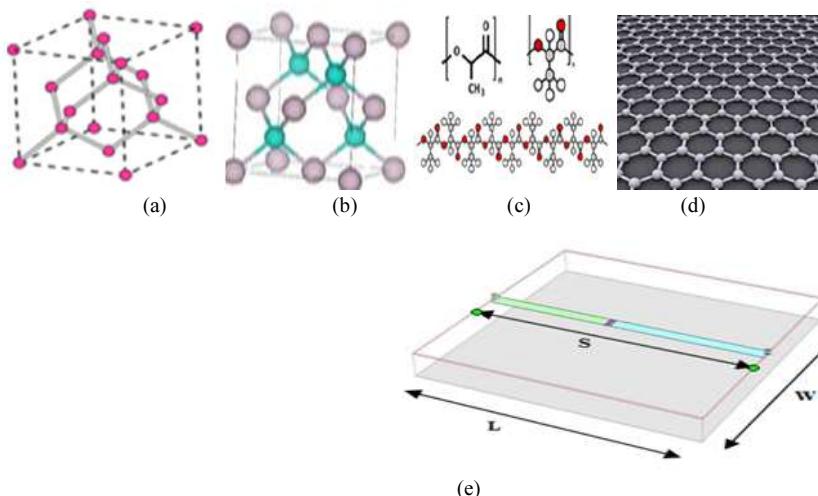
Figure 1. Semiconductors (a)Silicon (b)Germanium(c) Polymer(d)graphene (e)Basic Design(L=200miles ,W=200miles ,S=200miles)

3. Results and Discussion

3.1 Simulation with silicon switch

Silicon switch is placed in between two copper plates as shown in Figure 2.The overall structure is printed on the Fr-4 and Roger substrate is shown in Figure 2. A copper thickness of 0.03mm and conductivity of 5.8 e^6 S/m and the thickness of silicon is 0.5mm and conductivity 4.35e^{-3} S/m is selected. It is clearly shown that in Figure 3, the performance of switch with silver coating is nearly -10dB from 1GHz to 30GHz and the performance of aluminum coating is nearly -6dB from 1GHz to 35GHz for roger substrate at 850nm optical signal(room temperature). But in Fr-4 substrate the

conductivity of switch is reduced in high frequency ranges. It can be given a stable output from 1GHz to 20GHz as shown in Figure 4.
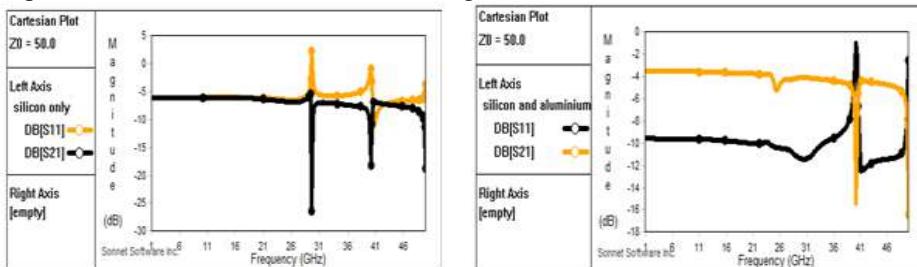
Figure 2. S11 Parameter of silicon switch on roger substrate (conductivity: 4.35e^{-3}. Left - aluminum, Right - silver)
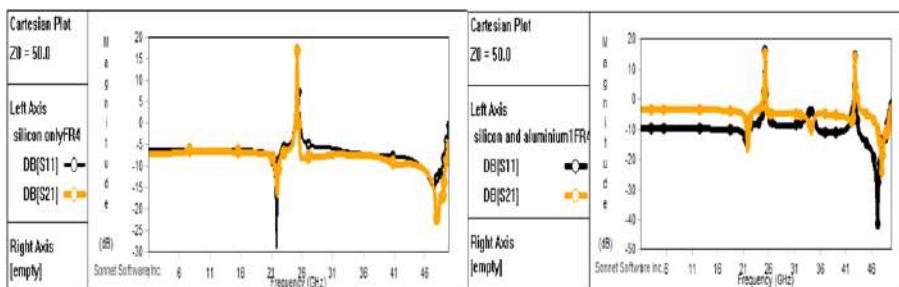
Figure 3. S11 Parameter of silicon switch on Fr-4 substrate (Conductivity: 4.35e^{-3} Left - aluminum, Right - silver)

3.2   *Simulation with polymer switch*

Figure 4 and 5 shown a simulation output of UV- polymer etched between a copper plate on Roger and Fr-4 substrate. The thickness of copper and their conductivity is same as above in all the simulations. The thickness of polymer is 0.508mm and conductivity is 50 S/m. In this simulation a stable output of -25dB is getting from Fr-4 substrate at 1GHz to 10GHz range and -50dB is getting from Roger substrate at 1GHz to 45GHz range.
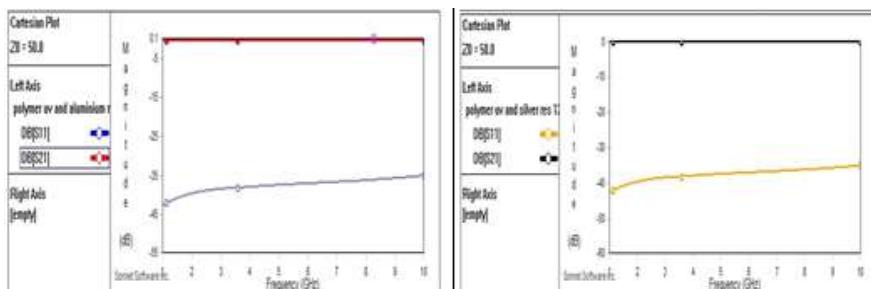
Figure 4. S11 Parameter of polymer switch on Fr-4 substrate (Conductivity:50 S/m -Left- aluminium, Right- silver)
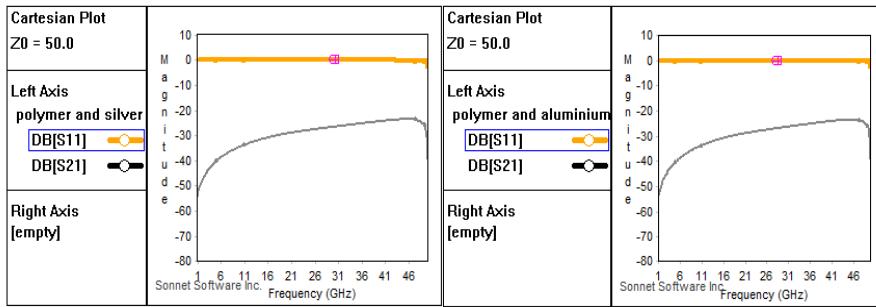
Figure 5. S11 Parameter of polymer switch on Roger substrate (Conductivity:50 S/m. Left- aluminium, Right- silver)

3.3 Simulation with germanium

Figure 6 shows a simulation of germanium is placed between the copper plate on the Roger and Fr-4 substrate are simulated. The selected thickness and conductivity of germanium is 0.5mm and 1.54e3S/m at 850nm (room temperature) optical signal for simulation. It gives a maximum of -15dB return loss at 1GHz to 20GHz at Fr-4 substrate and -20dB return loss at 1GHz to 20 GHz on Roger substrate with silver coating.
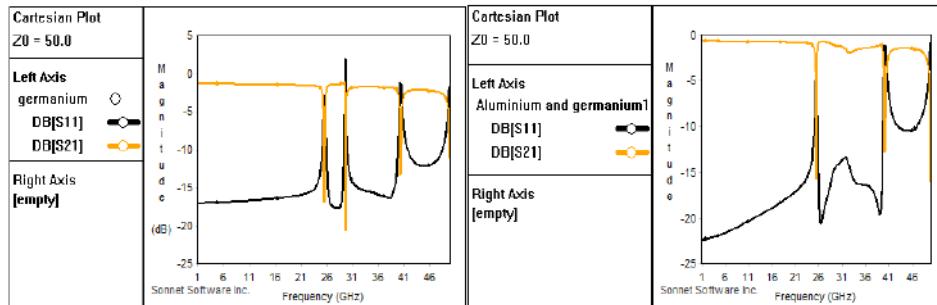
Figure 6. S11 Parameter of germanium switch on Fr-4 substrate (Conductivity:1.54 e^3S/m. Left- aluminium, Right -silver)

3.4 Simulation with Graphene

Figure 7 and 8 shows the simulation output of Graphene on Fr-4 and Roger substrate along with copper strip. Thickness and conductivity of the copper is 0.008mm, 58.14e^6 S/m and the thickness and conductivity of the graphene is 0.3mm,10^4-10^5 S/m at 130nm. It shows an excellent performance in two substrate materials with a minimum return loss value of -50dB.
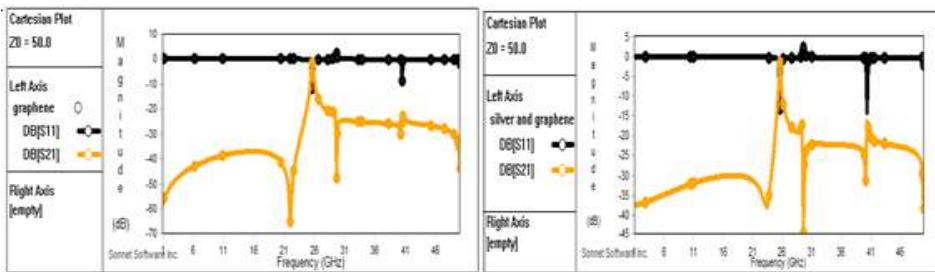
Figure 7. S11 Parameter of graphene switch on Fr-4 substrate (Conductivity: 10e^4 S/m. Left- aluminum, Right- silver)
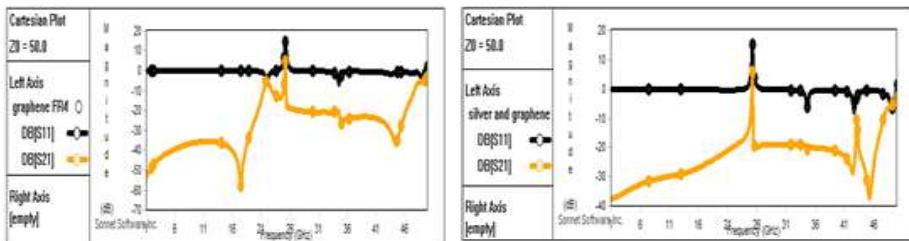
Figure 8. S11 Parameter of graphene switch on Roger substrate (Conductivity:10e^4 S/m.Left- aluminium, Right- silver)

4. Conclusion

The advantages of Graphene over other materials have boosted our interest. This highly strong yet thin and flexible material does not have an electronic band gap making it suitable to be a switch (it can switch between on and off).This encouraged us to explore its properties as a substitute for materials used for optical switches. On further study with the help of simulation shown us excellent results, however the manufacturing of Graphene is arduous and its usage is to be prevalent. Graphene is going to be a boon for the electronic domain in near the future. The above switching concept can be used for optically controlled frequency tuning reconfigurable antenna and filter circuits.

References

[1]. Pang, A. W., Gamlath, C. D., & Cryan, M. J. (2018). An Optically Controlled Coplanar Waveguide Millimeter-Wave Switch. IEEE Microwave and Wireless Components Letters, 28(8), 669–671. https://doi.org/10.1109/LMWC.2018.2840966

[2]. [2]Lu, L., Li, X., Gao, W., Li, X., Zhou, L., & Chen, J. (2019). Silicon Non-Blocking 4 × 4 Optical Switch Chip Integrated With Both Thermal and Electro-Optic Tuners. IEEE Photonics Journal, 11(6), 1–9. https://doi.org/10.1109/JPHOT.2019.2941960

[3]. Sikder, U., Usai, G., Yen, T.-T., Horace-Herron, K., Hutin, L., & Liu, T.-J. K. (2020). Back-End-of-Line Nano-Electro-Mechanical Switches for Reconfigurable Interconnects. IEEE Electron Device Letters, 41(4), 625–628. https://doi.org/10.1109/LED.2020.2974473

[4]. Krstic, M., Eren, S., & Jain, P. (2020). Analysis and Design of Multiphase, Reconfigurable Switched-Capacitor Converters. IEEE Journal of Emerging and Selected Topics in Power Electronics, 8(4), 4046–4059. https://doi.org/10.1109/JESTPE.2019.2960016

[5]. Jia, P., You, F., & He, S. (2020). A 1.8–3.4-GHz Bandwidth-Improved Reconfigurable Mode Doherty Power Amplifier Utilizing Switches. IEEE Microwave and Wireless Components Letters, 30(1), 102–105. https://doi.org/10.1109/LMWC.2019.2951215

[6]. Hindle, F., Shur, M., Abbot, D., & Ozanyan, K. B. (2013). Guest Editorial THz Sensing: Materials, Devices, and Systems. IEEE Sensors Journal, 13(1), 7–7. https://doi.org/10.1109/JSEN.2012.2226647

[7]. Pal, K., Zhan, B., Ma, X., Madhu Mohan, M. L. N., & Wang, G. (2016). Functionalized Graphene Oxide Dispersed Hydrogen Bonded Liquid Crystals Efficient Electro-Optical Switching. Journal of Display Technology, 12(3), 281–287. https://doi.org/10.1109/JDT.2015.2499326

[8]. Karanth, S. P., Sumesh, M. A., Shobha, V., Sirisha, J., Yadav M., D., Vijay, S. B., & Sriram, K. V. (2020). Electro-Optical Performance Study of 4H-SiC/Pd Schottky UV Photodetector Array for Space Applications. IEEE Transactions on Electron Devices, 67(8), 3242–3249. https://doi.org/10.1109/TED.2020.3004306

[9]. Safdar, A., Wang, Y., Reardon, C., Li, J., de Arruda, G. S., Martins, A., Martins, E. R., & Krauss, T. F. (2019). Interplay Between Optical and Electrical Properties of Nanostructured Surfaces in Crystalline Silicon Solar Cells. IEEE Photonics Journal, 11(4), 1–7. https://doi.org/10.1109/JPHOT.2019.2923562

[10]. Testa, F., Tondini, S., Gambini, F., Velha, P., Bianchi, A., Kopp, C., Hofbauer, M., Manganelli, C. L., Zecevic, N., Faralli, S., Pares, G., Enne, R., Serrano, A., Goll, B., Fontana, G., Chalyan, A., Lee, J.-M., Pintus, P., Chiaretti, G., … Stracca, S. (2019). Integrated Reconfigurable Silicon Photonics Switch Matrix in IRIS Project: Technological Achievements and Experimental Results. Journal of Lightwave Technology, 37(2), 345–355. https://doi.org/10.1109/JLT.2018.2871974

Coronary Illness Prediction Using Random Forest Classifier

Rekha G [a,1], Shanthini B [a] and Ranjith Kumar V [b]

[a] *Department of CSE, St. Peter's Institute of Higher Education & Research, Chennai, TN, India*
[b] *Department of Mechanical Engineering, Sri Sairam Engineering College, Chennai, TN, India*

Abstract. Heart diseases or Cardiovascular Diseases (CVDs) are the main cause of death on the planet throughout the most recent years and become the most dangerous disease in India and the entire world. The UCI repository is utilized to calculate the exactness of the AI calculations for foreseeing coronary illness, as k-nearest neighbor, decision tree, linear regression, and support vector machine. Different indications like chest pain, fasting of heartbeat, etc., are referenced. Large datasets, which are not available in medical and clinical research, are required in order to apply deep learning techniques. Surrogate data is generated from Cleveland dataset. The predicted results show that there is an improvement in classification accuracy. Heart disease is one of the most challenging diseases to diagnose as it is the most recognized killer in the present day. Utilizing AI algorithms, this paper gives anticipating coronary illness. Here, we will use the various machine learning algorithms such as Support Vector Machine, Random Forest, KNN, Naive Bayes, Decision Tree and LR.

Keywords. Coronary artery disease, Decision tree, K nearest neighbor; SVC, Logistic Regression, Naïve Bayes, Accuracy

1. Introduction

Cardiovascular disease is the most recognized killer in the present world. Consistently an excessive number of individuals are kicking the bucket because of heart illness. CAD can emerge because of lacking blood supply to courses. The two most common cardiac emergencies are a heart attack and myocardial infarction.
Heart disease describes a group of conditions that affect heart. Heart diseases include:

* Arrhythmias
* Congenital Heart Defects
* Heart valve disease

The aim of this study is to achieve accuracy so that it can predict a heart attack. Ages, sex, blood pressure, cholesterol, chest pain, heart rate, and so on are risk factors.

---

[1] G.Rekha, Department of CSE, St. Peter's Institute of Higher Education & Research, Chennai, TN, India; Email: rekhabensy@gmail.com.

2. Literature Review

Proposed a GUI created in python by utilizing tkinter to produce a basic dialog box which takes contribution for all the qualities essential for assessment [1]. Advanced AI techniques, for example, neural organizations and support vector machines are connected to NB, linear regression [2]. Portrays the hereditary calculation (GA)-SVR with genuine esteem GAs machine learning strategy [3]. Presents a model for predicting the next day's stock price. The interaction of intelligent agents was used in an AI model for this task [4].

A forecasting model for tourists is the proposed make use of LLSSVR (logarithm least-squares support vector regression) technologies [5]. Analyzes performance of ELMs to those of classification, support vector machine, and regression applications [6].

3. Methodology

In this paper, we have utilized our dataset for applying different machine learning algorithms for recognizing if an individual has coronary illness or not.

3.1 Data Collection

The data is acquired from the Cleveland Heart Disease database at UCI Repository. There are 14 attributes in the dataset. The dataset is narrated as follows:
   i.   Age: describes the age.
   ii. Sex: describes the gender identity; 1- male, 0 - female.
   iii. (cp): describes the form of chest.
   iv. (trestbps): describes the bp.
   v.   (chol): describes the serum cholesterol.
   vi. FBS: describes the fasting blood sugar (1 for true & 0 for false).
   vii. (restecg): describes the ECG results
   viii. (thalach): describes the heartbeat rate.
   ix. (exang): describes the angina.
   x.   Oldpeak: describes the depression mood.
   xi. (slope): describes the slope.
   xii. (ca): describes the blood vessels.
   xiii. (thal): describes thal feature (3 for normal, 6 for fixed defect, 7 for reversible effect).
   xiv. (target): describes target class.

3.2 Algorithms Applied

Gaussian Naive Bayes, Decision tree, SVM and Random Forest, KNN.

4. Results

4.1 Correlation Matrix
From the Figure 1, we can see that a few highlights are exceptionally related and some are definitely not.
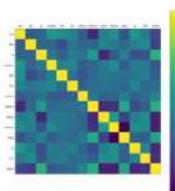
Figure 1. Features of Correlation Matrix

4.2 Histogram
It is a statistical tool for the portrayal of the appropriation of the data set in Figure 2.
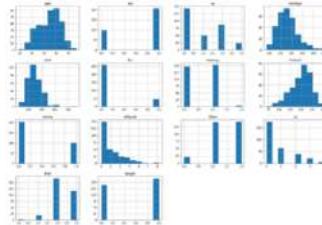
Figure 2. Histogram

Figure 3, Out of all the patients 165 patients actually have coronary illness. Now also visualize.
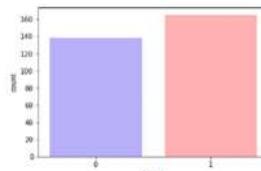
Figure 3. Count of the number of patients having heart disease (target =1)
Figure 4, classifying target variable between male and female and visualize the result.
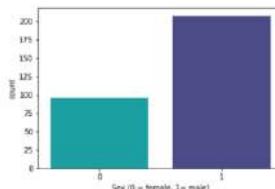
Figure 4. Count of both gender having heart disease
Figure 5, Scatter plot between "Maximum Heart Rate" and "Age" is shown below.
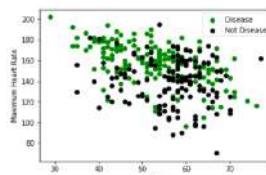
Figure 5. The Scatter plot between Age and Maximum heart rate

Exploratory Data Analysis:
The process of analyzing data, planning a new plan, test hypothesis, checking assumptions, and uncovering a hidden pattern is known as exploratory data analysis.

The Bar plot for target class (different features):
It is vital that the dataset we are utilizing ought to be pre-prepared and cleaned. This Figure 6 shows the tally of each target class.
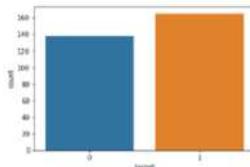
Figure 6. Count versus Target Feature

The chart above portrays the distribution of target versus count class, that is utilized to anticipate if an individual has coronary illness (0 = no coronary illness, 1 = coronary illness).
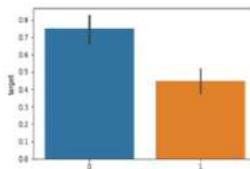
Figure 7. Target versus Sex Feature
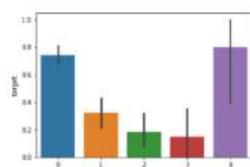
Figure 8. Target versus CA Feature
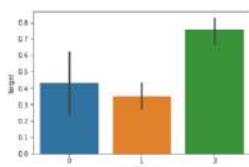
Figure 9. Target versus Slope Feature   **Figure 10.** Target versus Exang Feature
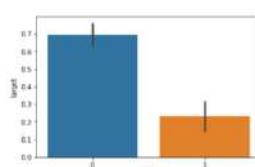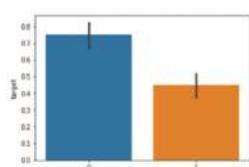
Figure 11. Target versus Restecg Feature   **Figure 12.** Target versus Fbs Feature
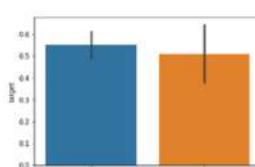
5. Machine Learning Algorithms

5.1 Logistic Regression

The supervised learning algorithm logistic regression is utilized to anticipate the parallel type of objective variable. In this paper, we achieved the accuracy of 85.25% by using this model.

5.2 Naïve Bayes Classifier

A Naive Bayesian classifier has equivalent execution with DT and other selected classifiers. In this paper, we achieved the accuracy of 85.25% by using this classifier.

5.3 K Nearest Neighbors

This calculation is utilized when the measure of information is enormous and there are non-decision boundaries between classes. In this paper, we achieved the accuracy of 67.21% by using this classifier.

5.4 Support Vector Classifier

This classifier is based on dataset containing features and labels.
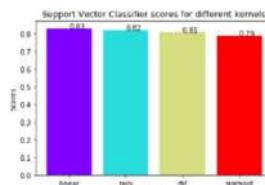
Figure 13. Support Vector Classifier scores.

This graph shows that the linear kernel is having the highest accuracy of 83 % by using this model.

5.5 Decision Tree Classifier

It can be used to solve problems including regression and classification.
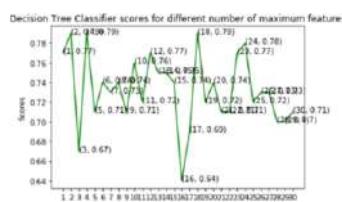
Figure 14. Decision Tree Classifier scores

This graph shows the line graph from which we observed that the maximum accuracy is 81.97%.

5.6 Random Forest

It creates decision trees from data samples chosen at random, makes expectations for each tree, and selects the best solution.

Table 1. Accuracy table

Algorithms	Accuracy
Gaussian NB Classifier	85.25%
K Nearest Neighbors Classifier	67.21%
Logistic Regression	85.25%
Decision Tree Classifier	81.97%
Support Vector Classifier	83%
Random Forest Classifier	95.08%

This table 1 shows the Random Forest Classifier which we observed that the maximum accuracy is 95.08%

6. Conclusion

This paper includes determining coronary illness dataset with legitimate data processing and execution. With an accuracy of 95 percent, Random Forest Classifier is perhaps the most successful of all the machine learning algorithms.

7. Future Enhancements

This may permit new calculation improvement to be performed off-site utilizing distributed computing programming, and afterward got back to the clinical setting as applications program interfaces (APIs) for PCs, cell phones and tablets.

References

[1] Sharma, V., Yadav, S., & Gupta, M. (2020). Heart Disease Prediction using Machine Learning Techniques. 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 177–181. https://doi.org/10.1109/ICACCCN51052.2020.9362842

[2] Hong, W.-C., Dong, Y., Chen, L.-Y., & Wei, S.-Y. (2011). SVR with hybrid chaotic genetic algorithms for tourism demand forecasting. Applied Soft Computing, 11(2), 1881–1890. https://doi.org/10.1016/j.asoc.2010.06.003

[3] Zarandi, M. H. F., Hadavandi, E., & Turksen, I. B. (2012). A hybrid fuzzy intelligent agent-based system for stock price prediction. International Journal of Intelligent Systems, 27(11), 947–969. https://doi.org/10.1002/int.21554

[4] Pai, P.-F., Hung, K.-C., & Lin, K.-P. (2014). Tourism demand forecasting using novel hybrid system. Expert Systems with Applications, 41(8), 3691–3702. https://doi.org/10.1016/j.eswa.2013.12.007

[5] Sun, X., Sun, W., Wang, J., Zhang, Y., & Gao, Y. (2016). Using a Grey–Markov model optimized by Cuckoo search algorithm to forecast the annual foreign tourist arrivals to China. Tourism Management, 52, 369–379. https://doi.org/10.1016/j.tourman.2015.07.005

[6] Al-Musaylh, M. S., Deo, R. C., & Li, Y. (2020). Electrical Energy Demand Forecasting Model Development and Evaluation with Maximum Overlap Discrete Wavelet Transform-Online Sequential Extreme Learning Machines Algorithms. Energies, 13(9), 2307. https://doi.org/10.3390/en13092307Sangya Ware, Shanu K Rakesh, Bharat Choudhary, Heart Attack Prediction by using Machine Learning Techniques, IJRTE, ISSN: 2277-3878, **Vol. 8 Issue 5**, January 2020.

Live Human Detection Robot in Earthquake Conditions

R. Kabilan [a,1], K. Lakshmi Narayanan [a], M. Venkatesh [a], V. Vikram Bhaskaran [a],
G. K. Viswanathan [a], S. G. Yogesh Rajan [a]
[a] Dept of ECE, Francis Xavier Engineering College, Tamil Nadu, 627003, India

Abstract. This report outlines a human searching device that takes the form of a robotic car and serves as a backup mechanism for saving lives in the event of a disaster. The temperature sensor, in general, detects the thermal image of the human body, and there has been extensive research into human searching with the gas and humidity sensor. In the intelligent robot device's study, achieving accurate and reliable human detection and tracking is a difficult challenge. The architecture of human detection and tracking mechanisms over non-overlapping field of views is examined in this paper. To compensate for their respective flaws, a search method is proposed. The proposed method's rate and accuracy of human detection was tested in an experimental setting. We may guide the robot's movement by commanding it to move left, right, forward, or backward. We plan to equip the robot with sensors that will enable us to track and detect humans behind the wall.

1. Introduction

Earthquakes occur in many parts of the world, causing human injury and death [1]. Almost every year in the last decade, an earthquake of magnitude 8 or greater has struck. Around 550000 people have been reported missing or killed as a result of earthquakes during this time span [2]. As a result, demand for rescue robots is growing. Rescue robotics is a relatively recent research area, with no significant advancements until the 1990s. Rescue robots are still not commonly used in real-world search and rescue scenarios, but the field is rapidly evolving [3-7]. Finding and identifying humans under debris can be a long and difficult task, and time is critical for effective rescue operations. As just a result, fast and effective search and recovery aids are critical. There are numerous techniques for conducting urban search and rescue operations today.

---

[1] R. Kabilan, Assistant Professor, Dept of ECE, Francis Xavier Engineering College, Tamil Nadu, 627003, India; E-mail:rkabilan13@gmail.com.

2. Existing System

We solved these concerns through equipping the UAV with just a camera for image data and a microphone for audio data, enabling us to identify people more accurately. A UAV equipped with a microphone/loudspeaker hangs above a disaster site, streaming an audio demand to all those below. The microphone picks up the voice of someone who answers, indicating if anyone needs to be rescued. A microphone and a hovering UAV, but at the other hand, were incompatible. The sound from nearby UAV rotors is picked up once by the microphone, masking the person's voice. Second, when you get further away from others, their voice becomes fainter.

To solve these issues, we must recommend sound-source detection technology to isolate the human voice by mixed sound through recorded methods [8]. An array microphone is used for sound source recording to estimate about role in human voice in sound-source difference [9]. They use results to localization to guide the UAV-mounted camera placed around direction towards sound source.

3. Proposed System

We'll equip the robot with sensors to track humans; if an object is detected, the observed readings will deviate, indicating the presence of a human body underneath. While radars have been observed to detect and monitor objects, it is not certain that the object being tracked is an individual [10]. A Arduino microcontroller, as well as a motor, make up the device.

Via the gas, temperature, and humidity sensors, the module displays the amount of emission, humidity, and temperature of both the person and the environment. This is possible thanks to the Ada fruit website, which receives data from the sensors directly from the module [11-14]. The following readings can be easily seen on cell phones or any other electronic device. The simplest way to know the temperature, humidity, and pollution level is based on the results we get from it. Figure 1 depicts the block diagram of the proposed scheme. We will integrate the robot with sensors to track humans; if any object is detected, there will be a deviation in the observed readings, indicating the presence of a human body underneath.

4. System Modules

4.1. Arduino UNO

The Arduino UNO is the robot module's main module, and it is only via the Arduino module that all circuits and connections are accessible. The module serves as a hub for command transmission and reception. The AT mega328P microcontroller is housed on an open-source microcontroller board. There is 1 KB of ROM, 2 KB data of RAM, and 32 KB on flash memory on the board.

4.2. Node MCU

The Arduino is linked to the node MCU, which is a strong WIFI module. It's also an IoT framework that's open source. It consists of an express-if systems system on a chip

and hardware based upon ESP-12 module. Rather than development kits, the name "Node MCU" is commonly used to refer to the firmware.

4.3. Bluetooth module

The Bluetooth module in this device allows us to connect it to a mobile phone via a mobile application, as well as monitor all of the movements, including forward, backward, left, and right, with pinpoint accuracy. The Bluetooth module and the intermediate serial port protocol module are both called Hc-05. This communicates through serial communication, which makes it simple to use.

4.4. Bluetooth module

The Dc motor is steered in either direction using the L298D motor drive IC. It has 16 pins and can drive two DC motors with a single integrated circuit.
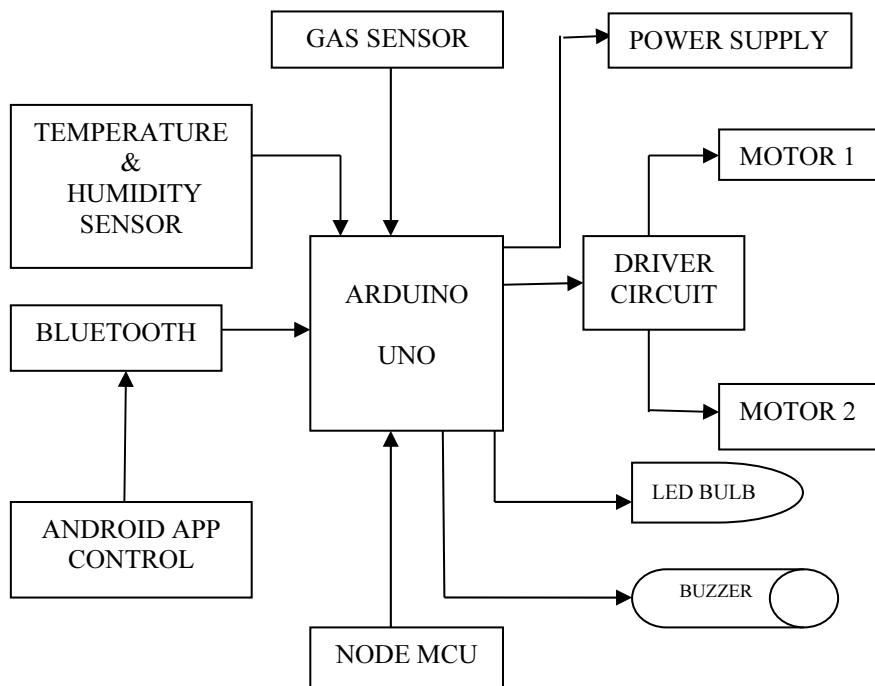
Figure 1. Block Diagram of proposed system

5. Hardware Description

There are 14 digital Input, output pins for accessing data to the Arduino board, as well as 6 analogue input pins. An Arduino's GND (#) pins serve which serves to be negative terminal for circuits or sensors attached on it. An 3.3V and 5V pins provide 5 and 3.3V of power applied to circuit, accordingly. By vast majority of sensors required set up an operate on two voltages. A power source is required for all boards to function. The board can be driven in two ways. The first method is to connected to USB cable with device. The board could also be worked by attached with the wall on power supply towards barrel jack. Additionally, code of the ATmega328 microchip is uploaded via USB.

Both Arduino device need a power supply for the code on the following chipset, which connects to the module's respective pin. On the Arduino UNO, the digital I/O pins are numbered from 0 to 13. It is used in digital inputs, such as data acceptance. The Arduino UNO, is also called as microprocessor module, ATmega328P is an microcontroller. Board is made up of various ram, rom, and flash memory components, as well as the Arduino IDE programming software. They also provide digital pins for modifying previously written data on the feature with the file of pins in the section of each Arduino digital I/O reading and writing way.

The DC motor is driven by the L298D motor drive IC, which allows this to drive in any path whereas the robot is operated. An H-Bridge circuit in a single L293D it has couple of DC motors simultaneously. Four inputs pins were in charge of control all rotation of attached motors. The input pins has either logic 1 or logic 0 which is adjust towards the direction. Normally it is 16-pin motor drive and has IC it can power 2 sets of DC motor which can operate in any direction at same time. In both directions, the L293D is modeled as bidirectional drive current to maximum upto 600mA.

6. Results

In our proposed method, we also use a multimeter to measure output voltage, input voltage and current. LUO converter used as quick charging method. The LUO converter receives a 12V AC input voltage as a single input. The output voltage must be 24 volts direct current. However, the harmonic can be reduced. When the converter is set to 10 amps, the input current will be 26.5 amps, as shown in Figure 2. The output voltage must be 24 volts direct current. Moreover, since this current is needed for fast charging, the required voltage at output will be 1.5V DC with 26.5V.
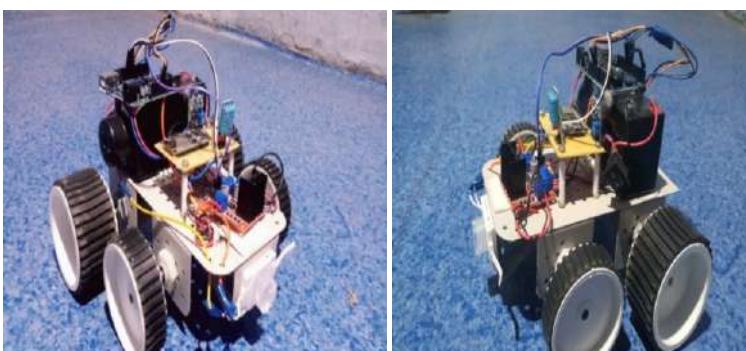
Figure 2. Shows the side view of Live Human Detection and system result

7. Conclusion

The proposed robotic device can able to capture of human lives on detection of earthquake along location tracking, environmental current condition monitoring and live streaming was developed with the Raspberry PI with an intermediate processor. The developed device has validated and has covered the benefits of many existing systems. The machine can also perform rescue operations in the event of landslides and avalanches. Since Internet access would be low in these locations, it is preferable to use radar communication in a real-time system rather than IOT. We created a low-cost autonomously operated rescue robot for this project. Search and rescue operations can be made more successful and efficient with the aid of this robot. After putting the module together, it was double-checked for functionality. A prototype robotic device capable of post-earthquake human live detection as location tracking, environmental current condition monitoring, and live streaming was developed with the Raspberry PI like intermediate processor. The developed device has been validated and has covered the benefits of many existing systems. The machine can also perform rescue operations in the event of landslides and avalanches. Since Internet access would be low in these locations, it is preferable to use radar communication in a real-time system rather than IOT. We created a low-cost autonomously operated rescue robot for this project. Search and rescue operations can be made more successful and efficient with the aid of this robot. After putting the module together, it was double-checked for functionality. The developed device has validated and has covered the benefits of many existing systems. The machine can also perform rescue operations in the event of landslides and avalanches. Since Internet access would be low in these locations, it is preferable to use radar communication in a real-time system rather than IOT. Many lives will be saved and the autonomous module could be used for several purposes in the real-time system if it is upgraded to the highest quality.

References

[1] Benli, E., Motai, Y., & Rogers, J. (2019). Human Behavior-Based Target Tracking With an Omni-Directional Thermal Camera. IEEE Transactions on Cognitive and Developmental Systems, 11(1), 36–50. https://doi.org/10.1109/TCDS.2017.2726356
[2] Booranawong, A., Jindapetch, N., & Saito, H. (2019). Adaptive Filtering Methods for RSSI Signals in a Device-Free Human Detection and Tracking System. IEEE Systems Journal, 13(3), 2998–3009. https://doi.org/10.1109/JSYST.2019.2919642

[3]   Booranawong, A., Jindapetch, N., & Saito, H. (2018). A System for Detection and Tracking of Human Movements Using RSSI Signals. IEEE Sensors Journal, 18(6), 2531–2544. https://doi.org/10.1109/JSEN.2018.2795747

[4]   Chang, F.-M., Lian, F.-L., & Chou, C.-C. (2016). Integration of Modified Inverse Observation Model and Multiple Hypothesis Tracking for Detecting and Tracking Humans. IEEE Transactions on Automation Science and Engineering, 13(1), 160–170. https://doi.org/10.1109/TASE.2015.2426712

[5]   Chen, Z., Wang, Y., & Liu, H. (2018). Unobtrusive Sensor-Based Occupancy Facing Direction Detection and Tracking Using Advanced Machine Learning Algorithms. IEEE Sensors Journal, 18(15), 6360–6368. https://doi.org/10.1109/JSEN.2018.2844252

[6]   Huang, C.-H. P., Allain, B., Boyer, E., Franco, J.-S., Tombari, F., Navab, N., & Ilic, S. (2018). Tracking-by-Detection of 3D Human Shapes: From Surfaces to Volumes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(8), 1994–2008. https://doi.org/10.1109/TPAMI.2017.2740308

[7]   Li, F., Chen, C.-H., Xu, G., & Khoo, L.-P. (2020). Hierarchical Eye-Tracking Data Analytics for Human Fatigue Detection at a Traffic Control Center. IEEE Transactions on Human-Machine Systems, 50(5), 465–474. https://doi.org/10.1109/THMS.2020.3016088

[8]   Li, H., Cui, G., Kong, L., Chen, G., Wang, M., & Guo, S. (2019). Robust Human Targets Tracking for MIMO Through-Wall Radar via Multi-Algorithm Fusion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(4), 1154–1164. https://doi.org/10.1109/JSTARS.2019.2901262

[9]   Li, H., Cui, G., Kong, L., Guo, S., & Wang, M. (2020). Scale-Adaptive Human Target Tracking for Through-Wall Imaging Radar. IEEE Geoscience and Remote Sensing Letters, 17(8), 1348–1352. https://doi.org/10.1109/LGRS.2019.2948629

[10]   Paral, P., Chatterjee, A., & Rakshit, A. (2019). Vision Sensor-Based Shoe Detection for Human Tracking in a Human–Robot Coexisting Environment: A Photometric Invariant Approach Using DBSCAN Algorithm. IEEE Sensors Journal, 19(12), 4549–4559. https://doi.org/10.1109/JSEN.2019.2897989

[11]   Will, C., Vaishnav, P., Chakraborty, A., & Santra, A. (2019). Human Target Detection, Tracking, and Classification Using 24-GHz FMCW Radar. IEEE Sensors Journal, 19(17), 7283–7299. https://doi.org/10.1109/JSEN.2019.2914365

[12]   Yang, B., & Zhang, M. (2017). Credit-Based Multiple Human Location for Passive Binary Pyroelectric Infrared Sensor Tracking System: Free From Region Partition and Classifier. IEEE Sensors Journal, 17(1), 37–45. https://doi.org/10.1109/JSEN.2016.2623325

[13]   Yun, J., & Woo, J. (2020). A Comparative Analysis of Deep Learning and Machine Learning on Detecting Movement Directions Using PIR Sensors. IEEE Internet of Things Journal, 7(4), 2855–2868. https://doi.org/10.1109/JIOT.2019.2963326

[14]   Zhang, Y. (2020). Detection and Tracking of Human Motion Targets in Video Images Based on Camshift Algorithms. IEEE Sensors Journal, 20(20), 11887–11893. https://doi.org/10.1109/JSEN.2019.2956051

Online Transformer Monitoring System

Mohan.S [a,1], S.Vinothkumar [b], K.Saravanakumar [b], M.Sudarsan [b]

[a]Assistant Professor, Department of EEE,

[b]UG Student, Department of EEE,

Sri Sairam engineering college, Chennai -600044, Tamil Nadu, India

Abstract. Power transformers, are basically used for stepping up& down the voltage levels. They are the primary equipment used in power transmission system. So it is primary to maintain all the transformers located geographically, but due to lack of man power it is impossible to monitor regularly. Due to these reasons, if a failure occurs in transformer may cause the network power shutdown. Though there are lot of protection measures that accompany a transformer, but by providing a online monitoring system will increase the reliability and reporting instantaneous fault confidently. This paper gives out the details in design and construction of an automatic monitoring system for power transformer parameters. A node-mcu module [esp8266] was enabled to monitoring of voltage, oil level and temperature (oil & winding) on a typical power transformer. With the internet of things (IoT), a self-defense system is designed and implemented for the transformer. In this system Transformer parameter are continuously marked and a graph is plotted. If the level of the parameter increases than the actual value, it gives buzzer alarm, if no action taken then the whole system will be tripped safely in power transformers.

Keywords. IoT, Transformer, moisture; high temperature.

1. Introduction

Power transformer is the single largest costliest equipment with any power sector utility which is indirectly enlightening every place of work and feeding all the means of entertainment in real life [1]. But at the time the most neglected asset, and the future of this valuable assert will be in dark if no appropriate preventive measures taken. Wireless monitoring systems are largely preferred to check various Industrial parameters from different locations. In most of the industries, there is a need to monitor different information for actual creation and all this parameter can't be monitored physically.

Wireless monitoring system is a basic in situations where accessibility to parameter to be monitored in not accessible or the situation where monitoring is a risky one [2]. Now a day's many industries are using wireless monitoring systems, but in our project which concentrate on power transformers. Power transformers are outfitted with some least amount protection that ensures their protection and dependable operation. It is very important that power transformer 's breakdown is minimized to the extreme because of their frequent maintenance, but many transformer are not maintained due to

---

[1] Mohan.S, Assistant Professor, Department of EEE, Sri Sairam engineering college, Chennai -600044, Tamilnadu, India; E-mail: mohan.eee@sairam.edu.in.

various reasons like they are present in city outer, remote area etc, by installing wireless monitoring system we can monitor any types of transformer. Wireless monitoring systems are largely preferred to check various Industrial parameters from different locations [3]. In most of the industries, there is a need to monitor different information for actual creation and all this parameter can't be monitored physically. But at the time the most neglected asset, and the future of this valuable assert will be in dark if no appropriate preventive measures taken. Wireless monitoring systems are largely preferred to check various Industrial parameters from different locations. In most of the industries, there is a need to monitor different information for actual creation and all this parameter can't be monitored physically.
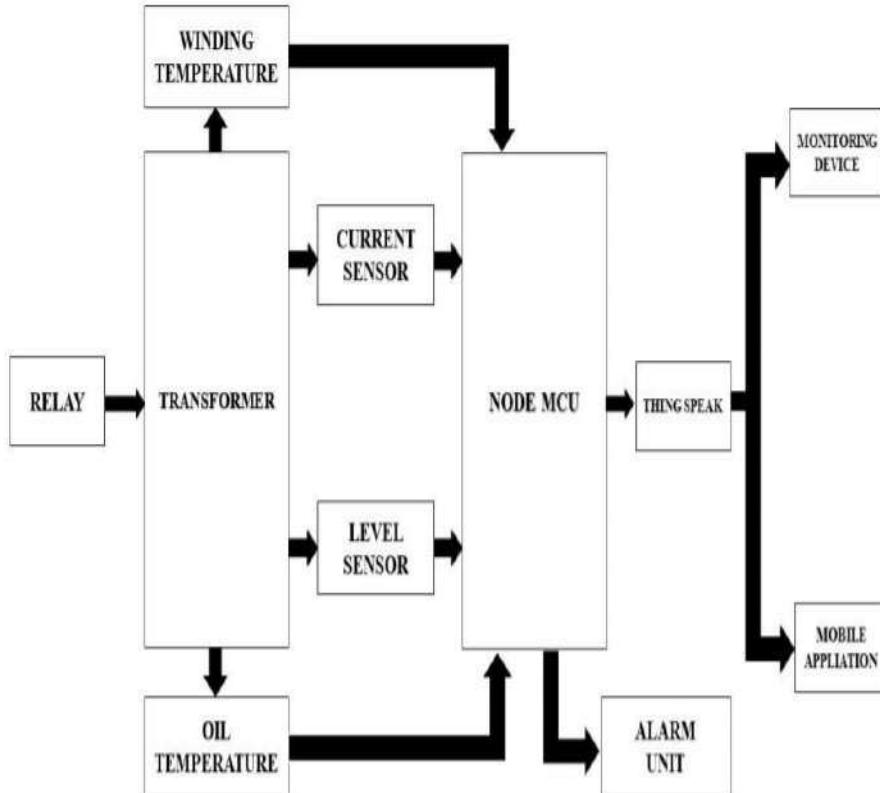
Figure 1. Block diagram

2. Working

The main objective of our project online transformer monitoring systems is to monitor the oil level, winding temperature, oil temperature and output current [4]. It also safeguards the power transformer in Figure 1. Firstly, Thermostat, current sensor and ultrasonic sensor are incorporated to various parts of the transformer [5]. Then, things view free application is installed to the system for monitoring purpose. Later the user will Have to login to the application using their login credentials Then the
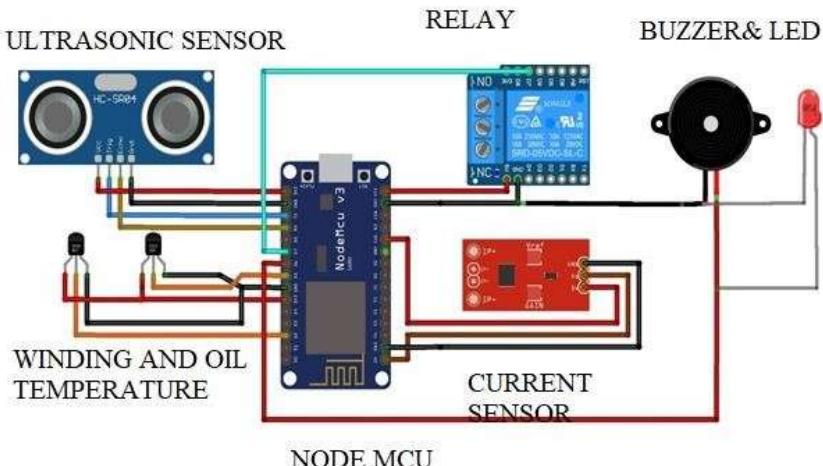
Figure 2. Circuit diagram

application will be synchronized to the sensors by the means of node-mcu. In Figure 2, the ultrasonic sensor, placed in the conservative tank of the power transformer monitors the oil level. Predetermined value of oil level is fed to the application priory.

As the sensor senses the oil level, it is simultaneously plotted into graph by the application. If the level drops or exceeds the predetermined value, the alert is thrown to the user and the alarm rings, if the alert is not noticed by the user for a period of time the particular unit will be disconnected by relay which is connected to the primary side of the transformer. A current sensor is connected in series to the load to monitor the output current

The winding temperature is monitored by thermostat which is placed with the analog meter. And similarly another thermostat is placed to analog meter corresponding to the oil temperature, the respective graphs are plotted in the application and the user is given an alert as mentioned earlier. If the temperature (oil and winding) further increases the primary side of the transformer is disconnected safely.

3. Cloud

All the data have been send to the cloud using MQTT protocol. It uses mosquito protocol. Any user can be able to access it using user id and password. According to cloud data all the actions could be taken. Cloud helps us to know about real time data analysis of sensor value and here uses the thing-speak online server which is interlinked with the Mat works. In this we have given a delay loop hole of 2 seconds for every duty cycle the system refreshes and the resultant of them is to be monitored on the graphical view in the MQTT protocol, which make sense of the data in the periodical data and the time wise.

4. Alternating System

For the sake of the transformer health the system is given a limitation of the temperature of the oil and winding which senses and in case of the winding which senses and in case of the oil and winding which senses and in case of the outbreak of the given level it intimates the user by giving a emergency alert in Figure 3 to Figure 6.

5. Results
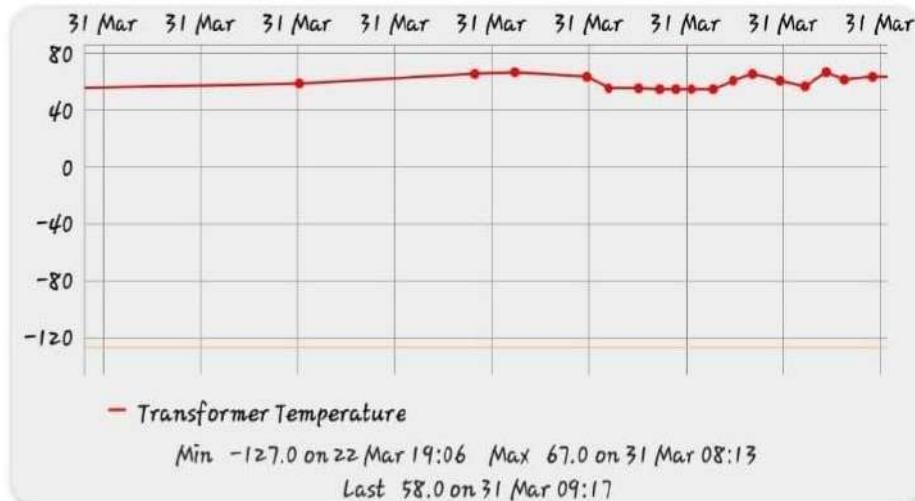
Figure 3. Transformer temperature
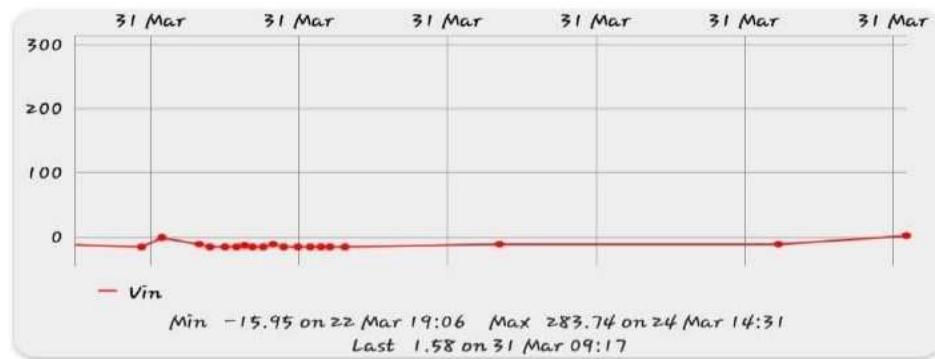
Figure 4. Voltage measured
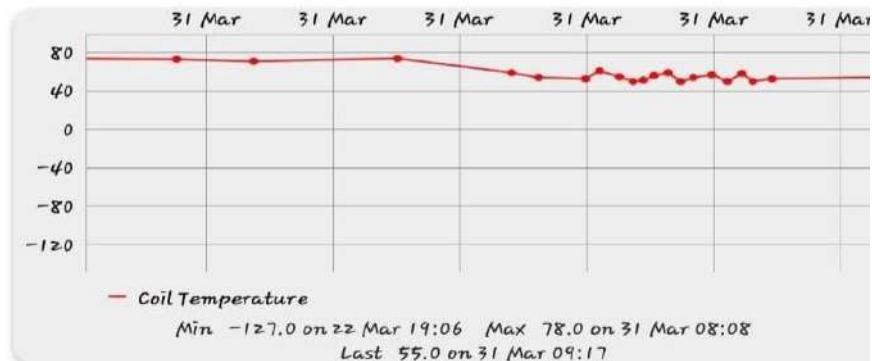
Figure 5. Finding temperature
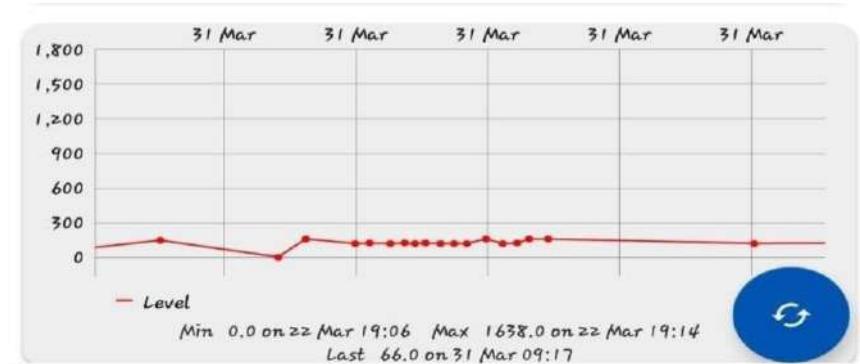
Figure 6. Oil level

6. Hardware Implementation

The hardware model is implemented for prototype of transformer monitoring the results are obtained through mobile application and system output in Figure 7. The proto type model implemented is shown below
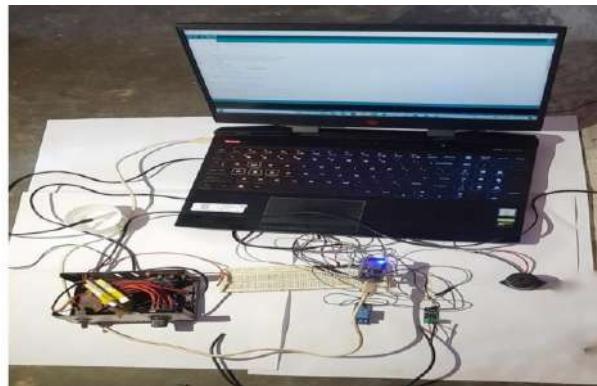
Figure 7. Hardware implementation

7. Conclusion

This paper has projected system arrangement which is commercial and replaced the errors that would occur by manual transformer monitoring set-up. This system provides a cloud-based storage and is accessible through a web application where the data is reachable distantly. There is image and hearing alert method to inform sub-station condition. In addition to the arrangement provides a path way to carry out necessary measure in case of urgent situation for the transformers. The projected arrangement saves cost very much and improving dependability.

References

[1] Santhosh, C., Aswin Kumer, S. V., Gopi Krishna, J., Vaishnavi, M., Sairam, P., & Kasulu, P. (2021). IoT based smart energy meter using GSM. Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2021.02.641

[2] Kaur, M., Mathew, L., Alokdeep, & Kumar, A. (2018). Implementation of Smart Metering based on Internet of Things. IOP Conference Series: Materials Science and Engineering, 331, 012015. https://doi.org/10.1088/1757-899X/331/1/012015

[3] Barman, B. K., Yadav, S. N., Kumar, S., & Gope, S. (2018). IOT Based Smart Energy Meter for Efficient Energy Utilization in Smart Grid. 2018 2nd International Conference on Power, Energy and Environment: Towards Smart Technology (ICEPE), 1–5. https://doi.org/10.1109/EPETSG.2018.8658501

[4] Avancini, D. B., Rodrigues, J. J. P. C., Rabêlo, R. A. L., Das, A. K., Kozlov, S., & Solic, P. (2021). A new IoT-based smart energy meter for smart grids. International Journal of Energy Research, 45(1), 189–202. https://doi.org/10.1002/er.5177

[5] Hlaing, W., Thepphaeng, S., Nontaboot, V., Tangsunantham, N., Sangsuwan, T., & Pira, C. (2017). Implementation of WiFi-based single phase smart meter for Internet of Things (IoT). 2017 International Electrical Engineering Congress (IEECON), 1–4. https://doi.org/10.1109/IEECON.2017.8075793

Improving Security Using Modified S-Box for AES Cryptographic Primitives

S. Sudhakar [a,1], A. Akashwar [b], M. Ajay Someshwar [b], T. Dhaneshguru [b], M. Prem Kumar

[a]*Assistant Professor, Department of Electronics and Communication Engineering*
[b]*UG Student, Department of Electronics and Communication Engineering*
Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering
College, Chennai, India

Abstract. The growing network traffic rate in wireless communication demands extended network capacity. Current crypto core methodologies are already reaching the maximum achievable network capacity limits. The combination of AES with other crypto cores and inventing new optimization models have emerged. In this paper, some of the prominent issues related to the existing AES core system, namely, lack of data rate, design complexity, reliability, and discriminative properties. In addition to that, this work also proposes a biometric key generation for AES core that constitutes simpler arithmetic such as substitution, modulo operation, and cyclic shifting for diffusion and confusion metrics which explore cipher transformation level. It is proved that in AES as compared to all other functions S-Box component directly influences the overall system performance both in terms of power consumption overhead, security measures, and path delay, etc.

Keywords. Cryptography, AES algorithm, Security, cipher

1. Introduction

In recent decades the emergence of social media and the invention of several digitalized transmissions – data security is becoming most important than ever and also a difficult task to accomplish. The need for each individual to identify themself digitally has spawned a wide variety of challenges, such as, for example, how to avoid fraud. Biometric data as fingerprint or iris scan is one way of identification, however, to use the data that is reliable for identification purposes the data must stay confidential, for that reason information security is important. The security measures inadequacies inherent in existing cryptosystems have driven the development of a new security model. With the development of several cipher models, the methodologies to explore the unique properties of each transformation model which can extract the information from cipher are also increased steadily. By using some combined approaches to provide solutions to the highly robust cryptosystem as well as other key issues are motivated by the research community.

---

[1] S.Sudhakar, Department of Electronics and Communication Engineering, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Chennai, Tamilnadu, India; E-mail: sudhakars106@gmail.com.

To ensure the cryptosystems are economically viable, all intermediate functions should be comprised with minimal resources and simplified computation which makes traditional security methodologies out of choice for them. The inherent data aggregation and information sharing properties render all existing pubic key cryptographic algorithms impractical. The major objective of a high-performance crypto core is to evaluate, develop, and analyze transformation models used in each stage of cipher conversion to achieve higher throughput efficiency with improved security. Among many crypto models, AES is the most promising method used for secured transmission in communication systems and has proven to be an optimal technique well suited for reliable data transmission over unsecured channels. It can able to support high bit rate data transmission and also robust to all kinds of attacks.

2. Related Works

This section includes the advantages of the existing AES core and its implications on security measures in detail. In general working principle and its parametric measures of any crypto, the core system is largely dependent on key length, the number of rounds, and various physical transformation functions involved during cipher conversion efficiency. In most cases, a trade-off is made among these measures. Implementation of AES core is related IoT applications are applicable only with some optimization models due to its parametric constraints...

In [1] introduced energy-efficient S-Box model for AES which is formulate using multi-stage Positive Polarity Reed-Muller form (PPRM) of composite fields. It includes hazard-transparent XOR gates that are located are replaced all other gates to block the hazards. In [2] developed FPGA technology mapping-based hardware optimization over AES core instead of logic optimization. As a result, both the linear and non-linear functional blocks of AES are considerably reduced, without causing any significant performance losses. In [3] optimized the AES core for lightweight applications and reduce hardware complexity effectiveness and energy consumption. To narrow down the computational burden various S-box architectures are analyzed based on FPGA design-specific hardware analysis for AES core. It includes pre-coded LUT-based S-box, the pure combinatorial gate-level logic implementation-based S-box using Galois field, and the path-optimized pipelined version of S-Box. Finally, optimal S-box using dynamic computation and Linear Feedback Shift Register (LFSR) random pattern generator-based S-Box is implemented for optimized AES core.

In [4] developed high-throughput bit sliced AES implementation based on novel data representation which exploits parallelization over many-core platforms. The building blocks used on each stage of AES utilized a bit sliced approach for parallelization. This improved parallelization offers a significant throughput rate. In this new AES model, the ShiftRows stage implicitly handled simple data rearrangement and neglects the complex computing process. Moreover, byte-wise operations are accomplished using the shift and swapping process.

Novel cipher model invented in [5] used self-inverse-permutation and Golden S-Boxes computations. By introducing a new mapping measure which comprises of self-inverse and arithmetic model ring of integers modulo 2n. the properties of this mapping function offer several key contributions to cipher structures by activating hard-core multipliers.
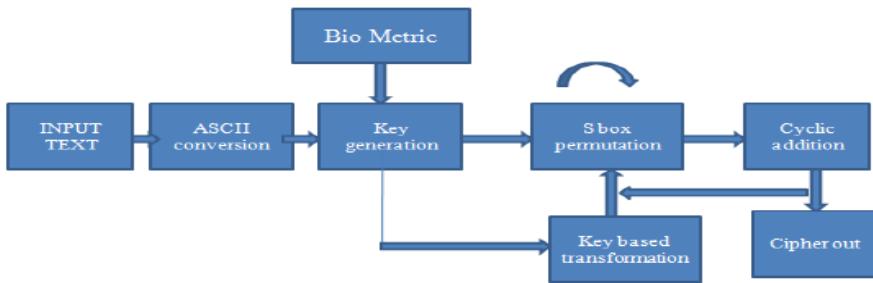
Figure 1. Proposed modified AES crypto model

3. Advanced Encryption Standard (AES)

Among various methodologies investigated for security, AES is considered as a prominent methodology that is useful for next-generation IoT and proved to be a potential alternative to exiting all other existing crypto core system [6]. In this real-time application 128 block size is preferred as AES standard due to its optimal tradeoff between security and computational complexity overhead as shown in Figure 1. During cipher conversion input text is arranged as a 4x4 matrix which is known as a state matrix or state array.

Add Round Key: This function involves key-based transformation and using bitwise Exclusive-OR operation between input cipher key and a state array.

Shift Rows: This is a physical transformation that operates on each row of a state array. Here only bytes of each state are shifted cyclically from left to right except the first row. The second, third, and fourth rows are shifted linearly.

Substitute transformation: As compared to all other functions S Box is nonlinear. During S Box computation byte-by-byte substitution is performed which generates a new byte value for each element in the state array. In general S Box is implemented in two ways namely Galois Field-based dynamic computation, Look-Up Table-based approach.

Mix   Columns: The   function   is   also   involved   physical   transformation performing matrix multiplication with some fixed compound values. Though AES is a technically advanced crypto core significant performance tradeoff is always occurs in many real-time applications due to its complex computations. Optimization is essential for AES to implement in IoT and embedded systems.

3.1 Performance security Measures

The attainable security of any cryptography algorithm largely depends on the key size and associated element of operations involved in the key generation process [7]. However, security enhancements through cipher key come with some significant computational time and hardware complexity overhead. Different algorithms used the different levels of complexity trade-off to meet desired security levels. Biclique Cryptanalysis widely used hash function for cryptanalysis measures over block cipher algorithms. In this type of cryptanalysis, the worst-case reference is used which is formulated from brute force and provides to a new benchmark.

3.2 Biometric Key Sequence Extraction Approaches

Biometric traits are always used to describe some unique discrimination model to explore the unique characteristics of different individuals. The generic biometric model extracts some invariant features from input biometric to incorporate the basic characteristics and all other relevant information useful for security measures. In real-time environments, the features are not stable in nature, and the method proposed should accommodate all nonlinear dynamics of image processing like interferences, rotational variations, and scale changes etc.

4. Experimental Results

4.1 Simulation Results

In order to validate the importance of the biometric key sequence generation process and to verify its impact in AES cryptosystem during cipher conversion process, AES cryptosystem is simulated using appropriate test inputs in various stages of data propagation as shown in Figure 2 and 3. The potential benefits of bio signal in key sequence generation and its efficiency over AES crypto design is also proved through simulation results
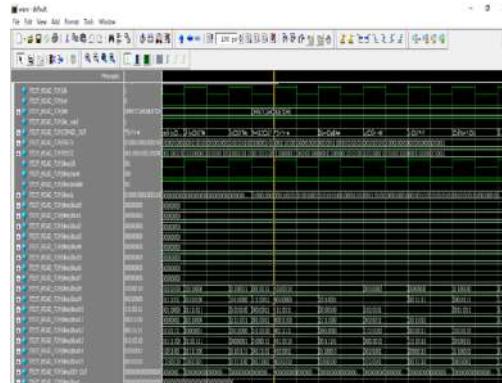
Figure 2. AES simulated cipher output for ASCII character input
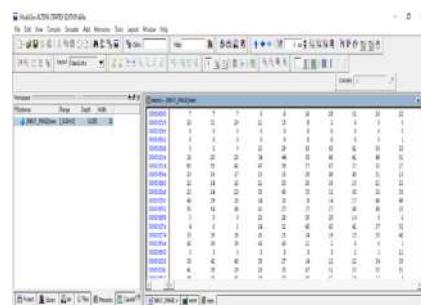
Figure 3. Bio metric digital image

4.2 Hardware synthesis results

In this chapter, we compare the performance metrics of proposed modified s box AES over conventional LUT based model and validated the metrics both in terms high performance and complexity trade off measures. The   proposed AES core is modeled using the Verilog HDL and synthesized using FPGA QUARTUS II EDA synthesizer for state-of-the-art comparison. The resultant S-Box is capable of achieving a flexible tradeoff with least possible design complexity and tolerable error protection in Figure 4. Moreover, by exploiting the benefits of dynamic computation side which can minimize memory space requirements and can able to support the path delay optimization using the sub pipelining model. In this pipelined composite s box can able to jointly optimize the computational complexity and energy from beneficiary GF enabled computation.

Table 1. Performance comparisons between modified dynamic nonlinear S box models using FPGA hardware synthesis results

S-Box model	Area (LE's used)	Fmax (MHz)
LUT model	216	228.21 MHz
Sub pipelined composite model	80	509.16 MHz
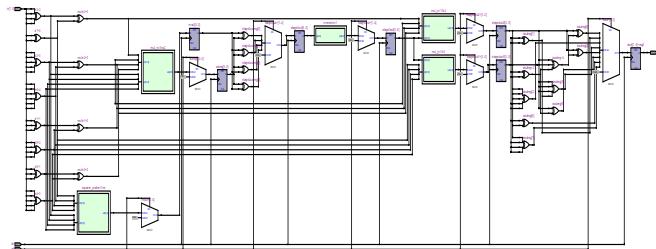
Figure 4. Dynamic non Linear S box RTL view

4.3 Performance comparison report

The FPGA hardware synthesizer tool has been used to measure the power utilization report and its experimental results are listed in Table 1. From the logical element's utilization summary, it is proved that the proposed modified S Box using composite field arithmetic model offers 6% area efficiency over conventional LUT based S- Box approach and achieves 14% hardware complexity reduction. The energy efficiency of optimized sub byte transformation is also proved to be the significant one as shown in Figure 5 through FPGA hardware synthesis results.

Figure 5. Power dissipation report.

5. Conclusion

Here in this work FPGA implementation of modified dynamic s box driven AES cryptographic algorithms with all forms of nonlinear transformation and analyzed its crypto core futures. It is also demonstrated that LUT based s box introduced for AES hardware open a back door to potential attacks. Here, we propose a nonlinear multiple instance S BOX along with biometric key sequence extraction model as a scan-protection scheme that provides security both at production time and over the course of the circuit's life. Compared to regular mode, this technique has no impact on the quality of the test or the model-based fault diagnosis. Here we proved that modified S BOX based AES will give better hardware complexity and power optimization with considerable delay enhancement.

References

[1] Good, T., & Benaissa, M. (2005). AES on FPGA from the Fastest to the Smallest (pp. 427–440). https://doi.org/10.1007/11545262_31

[2] Hafsa, A., Gafsi, M., Malek, J., & Machhout, M. (2021). FPGA Implementation of Improved Security Approach for Medical Image Encryption and Decryption. Scientific Programming, 2021, 1–20. https://doi.org/10.1155/2021/6610655

[3] Soltani, A., & Sharifian, S. (2015). An ultra-high throughput and fully pipelined implementation of AES algorithm on FPGA. Microprocessors and Microsystems, 39(7), 480–493. https://doi.org/10.1016/j.micpro.2015.07.005

[4] Swierczynski, P., Fyrbiak, M., Koppe, P., & Paar, C. (2015). FPGA Trojans Through Detecting and Weakening of Cryptographic Primitives. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 34(8), 1236–1249. https://doi.org/10.1109/TCAD.2015.2399455

[5] Yang, C.-H., & Chien, Y.-S. (2020). FPGA Implementation and Design of a Hybrid Chaos-AES Color Image Encryption Algorithm. Symmetry, 12(2), 189. https://doi.org/10.3390/sym12020189

[6] Hafsa, A., Sghaier, A., Machhout, M., & Malek, J. (2019). A New security Approach to Support the operations of ECC and AES Algorithms on FPGA. 2019 19th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), 95–100. https://doi.org/10.1109/STA.2019.8717302

[7] Wegener, F., De Meyer, L., & Moradi, A. (2020). Spin Me Right Round Rotational Symmetry for FPGA-Specific AES: Extended Version. Journal of Cryptology, 33(3), 1114–1155. https://doi.org/10.1007/s00145-019-09342-y

Assertion Driven Modified Booth Encoding and Post Computation Model for Speed MAC Applications

S.Sivasaravanababu [a,1], **T.R.Dineshkumar** [a], **Dr.G.Saravana Kumar** [a]

[a]*Assistant Professor, Department of Electronics and Communication Engineering, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, TN, India*

Abstract. The Multiply-Accumulate Unit (MAC) is the core computational block in many DSP and wireless application but comes with more complicated architectures. Moreover the MAC block also decides the energy consumption and the performance of the overall design; due to its lies in the maximal path delay critical propagation. Developing high performance and energy optimized MAC core is essential to optimized DSP core. In this work, a high speed and low power signed booth radix enabled MAC Unit is proposed with highly configurable assertion driven modified booth algorithm (AD-MBE). The proposed booth core is based on core optimized booth radix-4 with hierarchical partial product accumulation design and associated path delay optimization and computational complexity reduction. Here all booth generated partial products are added as post summation adder network which consists of carry select adder (CSA) & carry look ahead (CLA) sequentially which narrow down the energy and computational complexity. Here increasing the operating frequency is achieved by accumulating encoding bits of each of the input operand into assertion unit before generating end results instead of going through the entire partial product accumulation. The FPGA implementation of the proposed signed asserted booth radix-4 based MAC shows significant complexity reduction with improved system performance as compared to the conventional booth unit and conventional array multiplier.

Keywords. Multiply-Accumulate Unit (MAC), assertion driven modified booth algorithm (AD-MBE)

1. Introduction

The advent of High speed communication in 5G devices high performance arithmetic modelling is the key and heart of next generation applications to build unified accumulation and multiplication units that are optimized can be utilized in many applications such as finite impulse response (FIR) filtering , Fast Fourier transform (FFT) computation and wavelet transform etc. parallel prefix accumulation is widely preferred for many digital signal processing (DSP) and wireless communication devices for improved system performance. It is also used for formulating extended arithmetic units such as multiplication and division unit. To narrow down path propagation delay and design complexity overhead in adder unit many works introduced various prefix topologies. In most cases, the notable performance degradations occur with input bit

---

[1] S.Sivasaravanababu, Assistant Professor, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, TN, India; E-mail: sivasaravanababu@velhightech.com.

width of the adder unit. In general many DSP applications need to accommodate large number of multiply and accumulator (MAC) operation and its hardware accumulation are also increased accordingly [1].

2. Related Works

Many previous works focused only on FIR filter design optimization using various multiplication methods like booth, Vedic for high performance and canonical sign digit (CSD) formulation [2] of filter coefficient for low complexity. However, the major issues with these models are that, as the complexity reduced the performance rate is also significantly degraded; or vice versa. Therefore, unified model is required to narrow down this performance gap.

To accomplish this task, in recent years many works has been investigates the prefix accumulation and DA arithmetic for complexity reduction, high throughput and low power FIR design. Researchers also investigate the booth for various DSP applications since FIR design based on residue arithmetic offers both high speed as well as optimized computational complexity overhead The architectural choice of FPGA provides additional metrics in RNS systems due to its resource availability. Technology-dependent hardware optimizations for FPGA implementations of FIR filters are implemented in [3]. Here improved performance is archived by realizing 4:2 compressor and carry save adder (CSA) multiplication for different types of FIR filter architectures such as direct form, Transposed form and Hybrid form. The optimization also includes efficient post mapping using 6-input LUTs in addition to the hardware realizations. The hardware realizations process includes modification of design strategy and module instantiation within each boolean networks.

In [4], parallel prefix structures are used to suggest a novel signed-digit-to-canonical-signed-digit recoding. Some CSD models re-code from 2's complement binary numbers, while the presented CSD architectures convert from signed-digit numbers. Here each digit associate to the input is accompanied by its sign. The evaluation shows that the proposed CSD model with the conditional sum configuration outperforms current CSD models by 30%.

To solve this problem, the DA-driven reconfigurable block-based FIR filter architecture proposed in [5] used a modular DA scheme that could handle larger block sizes and longer filter lengths. Here the memory space requirements are not increase linearly with the filter order. In comparison to standard ones, the proposed approach provides 8 times the high-throughput reconfigurable FIR filters.

In [6] presented a detailed analyzes of various hardware efficient FIR design approaches. The decomposition of LUTs into small sub groups offers significant complexity reduction in DA LUT FIR structure. However, the degree of parallelism and FIR order continue to increase the complexity accumulation proportionally.
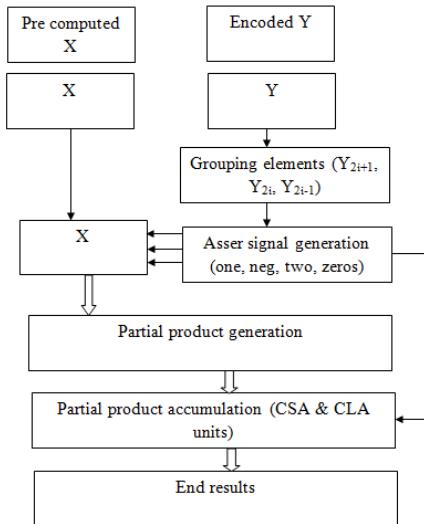
Figure 1. Proposed modified booth radix model

A * B = 123 * 10
01111011   *   10
011 110 11 0
Grouping
011   111   101   110
2X   0   -X   -X
PP   =   000010100 000000000 111110110 111110110 = 1230

3. Proposed Assertion Driven Booth Algorithm

Among various path delay reduction and hardware sharing-During partial product accumulation CSA adders were reused (hardware sharing). In case of zero signals asserted that associated PP generation & accumulation can be skipped (Digital Transition reduction). and also used CLA for last two rows of PP addition for reduced path delay overhead as shown Figure 1. Full of parameterized model – for reconfigurability which can configured dynamically. Two complements circuit insertion- to support both signed and unsigned multiplication.

- Since it can cut the number of partial items in half, the radix-4 booth algorithm is used to boost multiplier efficiency.
- In radix-4 , we consider 3 bits at a time as shown in Figure 2.
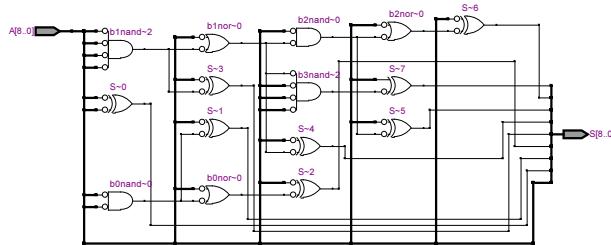
Figure 2. Schematic model of radix encoding model

3.1   Assertion methodology

This function involves transformation of booth encoded values into control signals which comprise of bitwise Exclusive-OR operation between input and a encoding state array as follows:

- Zeros – for avoiding PP generation and associated PP accumulation
- Twos – enable associated shifting operation   (2X is equal to X << 1) that can be done during PP accumulation.
- Negs -   to allows inversion and 2's comp units
- Ones – no changes required

3.2 Advantages of proposed BOOTH MAC System

High Performance: The absence of multiplier components during MAC arithmetic results in high speed computation.

Energy efficiency: To realise the MAC method, regardless of the order input operands, only a single accumulator portion is used, which minimises the digital transfer activities during MAC operations. Switching events that are 1-0 or 0-1 are directly proportional to power. This will simplify the overall architecture by regulating the arithmetic accumulation.

4.   Experimental Results

4.1 Simulation Results

To prove the value of assumption in the partial product generation process and its effect on the MAC machine during the multiplication process, suitable test inputs are used at different stages of data propagation, as shown in Figure 3. The potential benefits of optimal PP generation and its efficiency over conventional booth MAC design is also proved through simulation results
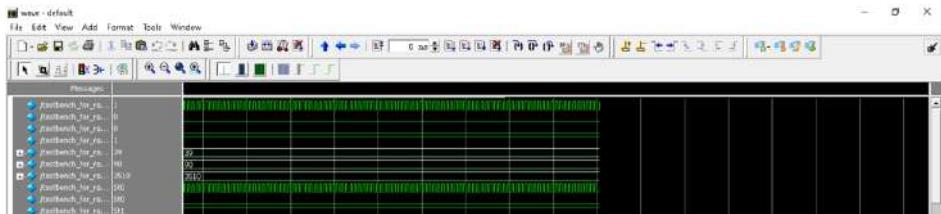
Figure 3. Booth simulation results with unsigned input

4.2 Hardware synthesis results

In this chapter, we compare the performance metrics of proposed modified booth over conventional encoding based booth radix-4 model and validated the metrics both in terms high performance and complexity trade off measures. For state-of-the-art contrast, the proposed booth radix-4 core is modelled using Verilog HDL and synthesised using FPGA QUARTUS II EDA synthesiser. The resultant MAC is capable of achieving a flexible tradeoff with least possible design complexity and tolerable energy efficiency .Moreover, by exploiting the benefits of PP computation which can minimize memory space requirements and can able to support the path delay optimization using the CSA and CLA model.   In this asserted booth radix-4 can able to jointly optimize the computational complexity and energy from beneficiary tree enabled PP computation.

Table 1. Performance comparisons between modified asserted booth MAC model using FPGA hardware synthesis results

MULTIPLIER MODEL	Design complexity(LEs)	Speed(MHz)	Total Power dissipation (mW)
Array multiplier	327	129.57 MHz	103.08mW
Booth radix-4	285	244.02MHz	101.73mW
Proposed booth radix-4	261	255.43 MHz	100.02mW

4.3 Performance comparison report

The FPGA hardware synthesizer tool has been used to measure the power utilization report and its experimental results are listed in Table 1. From the logical elements utilization summary it is proved that the proposed modified booth using asserted booth arithmetic model offers 20% area efficiency over conventional booth radix-4 approach and achieves 3% power consumption reduction. The energy efficiency of optimized PP transformation is also proved to be the significant one as shown in Figure 4 through FPGA hardware synthesis results.

Figure 4. Power dissipation report

5. Conclusion

Here in this work hardware implementation of assertion driven modified booth radix-4 algorithms with all forms of partial product transformation and analyzed its performance metrics. It is also demonstrated that assertion unit introduced for booth algorithm offers significant energy level optimization and open a platform for low applications. Here, we propose a hierarchical tree based PP addition along with assert signal extraction model as a path delay optimization scheme that provides least critical path both at encoding and end stage. Here we proved that modified assertion based booth will give better hardware complexity and power optimization with considerable delay enhancement.

References

[1] Liaskos, C., Nie, S., Tsioliaridou, A., Pitsillides, A., Ioannidis, S., & Akyildiz, I. (2018). A New Wireless Communication Paradigm through Software-Controlled Metasurfaces. IEEE Communications Magazine, 56(9), 162–169. https://doi.org/10.1109/MCOM.2018.1700659

[2] Pashaeifar, M., Kamal, M., Afzali-Kusha, A., & Pedram, M. (2018). Approximate Reverse Carry Propagate Adder for Energy-Efficient DSP Applications. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 26(11), 2530–2541. https://doi.org/10.1109/TVLSI.2018.2859939

[3] Zhou, Z., Huang, P., Xiang, Y. C., Shen, W. S., Zhao, Y. D., Feng, Y. L., Gao, B., Wu, H. Q., Qian, H., Liu, L. F., Zhang, X., Liu, X. Y., & Kang, J. F. (2018). A new hardware implementation approach of BNNs based on nonlinear 2T2R synaptic cell. 2018 IEEE International Electron Devices Meeting (IEDM), 20.7.1-20.7.4. https://doi.org/10.1109/IEDM.2018.8614642

[4] Saravana Kumar,G., Parthasarathy, V., Sivasaravana Babu, S., Sudhakar, S. An Energy Spectral Density Computational Scheme for Brain Computer Interface, International Journal of Applied Engineering, 10(8), 21139–21149.

[5] Parthasarathy, V., Saravana Kumar, G., Sujeet Blessing, C., Sivasaravana Babu, S., Sudhakar, S. (2014). Measurements of Pattern Identification Accuracy using Brain Computer Interface Technology for Neurological Disability Patients, Elsevier-SciVerse Science Direct, APCBEE Procedia,1–5.

[6] Parthasarathy, V., Saravana Kumar, G., Sivasaravana Babu, S., Grimm Christoph. (2015). Brain Computer Interface Based Robot Design, Journal of Theoretical and Applied Information Technology, 72(3), 315–321.

Highly Secured Dynamic Color QR Pattern Generation for Real Time Application

R.Sanjjey [a,1], S.Abisheak [a], T.R.Dineshkumar [a], M.Kirthan [a],
S.Sivasaravanababu[a]

[a] *Department of Electronics and Communication Engineering, Vel Tech High Tech Dr Rangarajan Dr Sakunthala Engineering College, Chennai, TN, India*

Abstract. This work advances the state-of-art secured WBAN system and QR pattern enabled authentication for privacy measures. An attempt was made to integrate all the above process to build high performance WBAN system. In this work, a comprehensive statistical framework is developed with randomized key generation and secured cipher transformation for secured sensor node communication. We create primary colour channels based on three different QR codes that are widely used for colour printing and complementary channels for capturing colour images. Last but not least, we produced a colour QR pattern.

Keywords. WBAN system, primary colour channels, QR codes

1. Introduction

The security measures inadequacies inherent in WBAN systems have driven the development of a new security model. On the other end applications of wireless sensor networks have been emerged steadily, ranging from industrial management to healthcare applications which leads key management and distribution poses significant challenges, especially in resource   constrained sensor networks. When it comes to wireless communication using sensor nodes security measures plays a vital role for reliable communication which has to be addressed [1]. Thus, Security is one of the essential things needs to be incorporated and challenging task needs to be accomplished with traditional cryptographic algorithms due to following reasons: i) Typical sensor nodes consist of a tiny computing device that forward information to the destination node. ii) Memory, energy and bandwidth constraints.   iii) Demands high throughput rate. Conventional block ciphers are not optimal for sensor networks due to its computation and hardware complexity overhead. The primary concern over many existing binary sequences extraction process for biometric key extraction is that they are complex in nature and poor resource efficiency which is incompatible with sensor networks. In recent years, ECG signal models have gained momentum in security applications for low complexity and complete randomness [2].

---

[1] *R.Sanjjey*, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Chennai. India; E-mail: sannjeyravichandran@gmail.com.

Though key generation from ECG signals are optimal for sensor networks since it doesn't require pseudo-random number generators (PRNG) and memory for seed storage it has the least invariance level over time period and significantly correlate with other classes which will affect uniqueness properties.

2. Specific Objectives

- To develop a novel tiny encryption modeling for secured sensor data communication.
- To develop QR pattern generation using color selection key to incorporate authentication requirement from WBAN applications.
- To develop key based QR pattern driven data protection and access control for WBAN system.
- To develop cloud storage mechanism for WBAN technology and security and privacy of patient and medical personals.
- To perform hierarchical complete processing steps involves in real time WBAN health care system.

3. Proposed System

The inadequacies inherent in existing WBAN system have driven the development of new light weight hardware architecture and to exploit the benefits of biometric characteristics of ECG biomedical signal for key generation task in digital crypto systems for high performance security system filter [3]. Moreover, FPGA devices have been used extensively for high throughput applications but they cannot full fill the several Gbps throughput requirement of next generation systems, and low power consumption with the invention of power compatible 5G devices. Such systems rely fully on arithmetic techniques used to carry out computation. In general bio cryptosystem framework includes a number of benefits: i) it uses the random binary sequence generated from input biometric, which solves key management issues with low computational cost. Most of the existing key generation methodologies are not completely random in nature to transform the input sequences into cipher and computationally intensive in nature. ii) the proposed bio-cryptosystem generate different binary sequences for every class and thus preserves discrimination; iii) this scheme is secured and more efficient as compared to other block Cipher; iv) completely randomized transformation nature of proposed bio-cryptosystems suits for both text and image encryption. This work is guided by the motivation of extending the merits of light weight crypto model to improve the performance of diffusion and confusion computations since highly simplified arithmetic computations only involved in light weight crypto system. This thesis is focused on the cipher key generation using ECG signal to deal with the challenging key management problems.
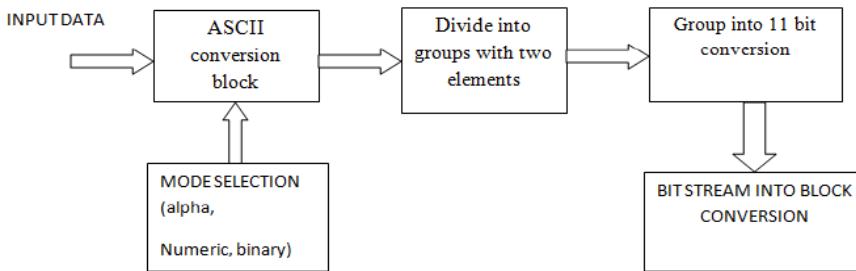
Figure 1. Bit stream to Block Conversion

4. Access Control Model

Authenticity in WBAN denotes the medical information transfer from access point to the storage space and rule sets to define each valid group. The basic parameters required for reliable medical data communication in WBAN and associated security measures are as follows:

Data authenticity: In multi path data propagation during wireless data communication some intended attacks are carry out using malicious nodes which is called bogus sensor; thus, the end terminals should validate the sensor nodes origin of information to ensure the data authenticity

Data confidentiality: The WBAN system is highly vulnerable to all kinds of passive attacks due to its open access nature and allows some loophole to reveal all personal information's about patient to external users. It is essential to overcome this limitation using crypto core based secured biomedical data transmission in all phases. Data integrity: To handle this WBAN system should verify the sensor details during data integration. Data availability: It comes under denial-of-service (Dos) related attacks over medical server where sensor information's are collected and stored. In case of access denials during medical services causes complete WBAN system failure. In most cases some unique detection methodologies are incorporated to detect the avoid this DOS attacks. IN order to narrow down the human intervention in automated patient monitoring system without compromising the reliability several hierarchical transformations are introduced between end-to-end data communication which starts from data collection using body sensor to the cloud database. It also includes data authentication and accessibility by medical personals. Here the proposed WBAN system includes some novel approaches in both access control and data security. Here some unique key driven QR pattern generation and associated color selection of each generated patterns all potential attacks are prevented.
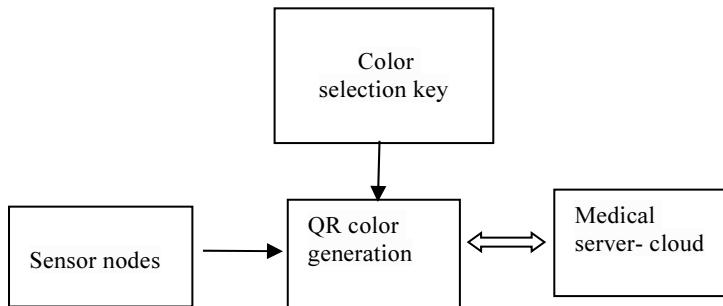
4.1 Color variant QR pattern for authentication

Due to its inherent error correction mechanism and associated data embedding metrics, QR patterns have been widely motivated for authenticating several commercial commodities including cloud data. It makes use of simplest computational process during QR code generation which allows better readability to all available QR code readers in smart phones. In particular unique color selection for each patterns allows them to accommodate several million users in this proposed WBAN healthcare system. Here sensor nodes are authenticated at the server side using QR pattern.

Figure 2. QR pattern authentication

QR pattern variants are used to manage WBAN access and protect data. Patients are recorded in cloud storage using a colour key created from their patient ID to convert their personal information into a QR pattern. During data transmission to the cloud, the sensor information from access points must be validated. The level of protection and resistance to malicious attacks is determined by colour variants. Color key randomization enhances major confusion metrics, and QR patterns' inherent error correction capabilities allow for the lowest possible false detection rate during validation.. Moreover, all these QR patterns are generated from some unique user defined details or template data and this QR pattern generation-based data authentication doesn't altered the original data since it makes use some mapping function or one way transformation.

5. Summary

The first part of this chapter discusses the proposed WBAN system's goals. The second section delves into the security methodologies employed in the efficient processing of WBAN data, followed by a thorough explanation of the characteristics. This work focused on the generation of color QR pattern for WBAN system cloud integration. In addition to that this chapter discussed in detailed about the performance metrics of WBAN system with the inclusion of QR pattern analyzes. The proposed methods show the adequate performance when compared to the existing methods.

6. Results and Conclusion
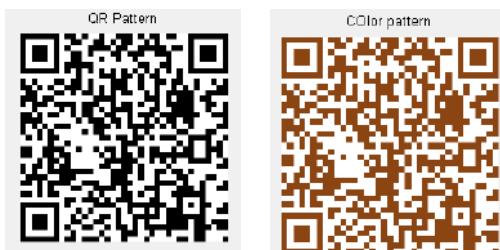
Figure 3. Input GUI Model

Here push buttons were created along with axes plot for dynamic user inputs and associated relevant results for each input dynamics. Both authentication and access control mechanism can be easily validated.

Figure 4.a. Patient details and generated QR pattern   **Figure 4.b.** Input Patient details server updation

After obtaining the user details from different individuals color QR codes are generated based on selected key components and forwarded it to medical server as a core template. With this method, instead of predetermining the user data as statistical values QR patterns are generated for all given inputs and construct as single compound templates which has all the basic properties of cancelable template.

6.1 Data authentication output

After obtaining the user details from different individuals color QR codes are generated based on selected key components and forwarded it to medical server as a core template. With this method, instead of predetermining the user data as statistical values QR patterns are generated for all given inputs and construct as single compound templates which has all the basic properties of cancelable template.

Figure 5.a. Access control model output

Figure 5.b. Modified QR pattern

6.2 Color key driven data authentication

The randomization of component key and pattern selection provides major benefits. Non-inevitability steps I QR patterns' built-in error correction metrics minimise false detection rates during validation (ii) due to compound template generation, variance in input trails won't be tolerated during validation because each template contains fine information from various types of multi modal system

7. Conclusion

Here we analysed the performance of color and pattern selection QR image for WBAN based applications of secured data transfer, data authentication and privacy measures. We also introduce the idea of toned – noisy – Rotated colour QR images, as well as a method for automatically decoding these QR codes. The graceful potential metrics of the QR pattern and its perceptual quality as a feature of embedding parameters in WBAN applications are demonstrated experimentally.

References

[1]   Zhang, C., Wang, J., Han, S., Yi, M., & Zhang, Z. (2006). Automatic Real-Time Barcode Localization in Complex Scenes. 2006 International Conference on Image Processing, 497–500. https://doi.org/10.1109/ICIP.2006.312435

[2]   Yang, H., Jiang, X., & Kot, A. C. (2010). Accurate localization of four extreme corners for barcode images captured by mobile phones. 2010 IEEE International Conference on Image Processing, 3897–3900. https://doi.org/10.1109/ICIP.2010.5651603

[3]   Huijuan Yang, Kot, A. C., & Xudong Jiang. (2012). Binarization of Low-Quality Barcode Images Captured by Mobile Phones Using Local Window of Adaptive Location and Size. IEEE Transactions on Image Processing, 21(1), 418–425. https://doi.org/10.1109/TIP.2011.2155074

25 Meter Rifle Shooting Training Kit

Mavillapalli Nikhitha [a,1], Shobana D [a], Anne Sylvia S [a], Batta Hemasree [a],
Priyadharshini P[a]
[a]*Department of Electrical and Electronics Engineering, Panimalar Institute of
Technology, Chennai, Tamilnadu,India*

Abstract. The goal of this project is to obtain all of the details about the shotgun so that we can correct any mistakes made during practice. With so many undetectable errors made by operators at work, this device aids in locating and adjusting the timing of adjustments in accordance with their errors. This procedure is carried out by detecting deviations from the shooters while shooting from a distance. Following each session, all data is verified, and steps are taken to improve the shooters game and correct their mistakes. This device improves the quality of the archer while also assisting in the improvement of the game. It also appears to be a list of minor blunders that one may deal with. The arduino IDE, which does all mathematical and mathematical operations, is in charge of this device. The light attached to this device indicates the start and stop positions of the shooter, allowing the shooter to be trained to start and stop fire.

Keywords. Gun Shooting, Arduino, Android, Application, Bluetooth.

1. Introduction

The purpose of this project is to obtain all of the details about Rifle Shooter in order to correct any mistakes made during the process [1]. There are several unmistakable mistakes made by the shooter in operation, and this device aids us in detecting them and changing the timing of their mistakes. The process is improved even further by detecting deviations from the shooters while shooting from a distance. The coach and the player (shooter) analyze this deviation. After each session, all data is verified, and steps are taken in accordance with the errors that occurred. The device improves the quality of the archer and, as a result, aids in the improvement of the game. The project has a large arduino Nano controller, arduino Mega, which is utilized for all statistical and mental units in all calculations. The sensors employed in this project are bluetooth modules, which are similar to input devices. But the system is for viewing telemetry data. The sensors are connected to read various posters and user status to detect mistakes made by them and to correct the mistake made by them.

Mavillapalli Nikhitha, Department of EEE, Panimalar Institute of Technology, Chennai, TN, India.
E-mail: mavillanikhitha723@gmail.com.

2. Existing System

Smart Detector that will be able to detect reeds above the knee level and, more importantly, helps to navigate around large objects alone with the help of RFID (radio frequency detection system). RFID technology is precious, precise, and compatible with Braille blocks [2]. The detector has a microcontroller and an RFID reader on the bottom that can scan a piece of RFID cards on the floor and assist the user in navigation. This can only provide enough user vibration for error resolution. After shooting people, User Target Detector programmes are created, but only a handful can provide powerful interaction to adapt to changes. Outside, it doesn't work as well. The Target Detector employs a precise positioning system, wireless communication, a portable computer, and a voice communication interface to test people and assist them in making a mistake in both familiar and unfamiliar environments. It only provides a valuable input of materials that do not provide bodybuilding processes. This is a significant disadvantage of the target detector [3-5].

3. Proposed System

Our project contains the Arduino Mega controller is used for all arithmetic and logic units. The Bluetooth, RT, and SD ard modules were employed as input devices in this project. The NRF module, display, audio amplifier, and speaker make up the output system. The sensors are connected to read various posters and user status in order to detect and correct mistakes made by them. Using the display, the gun manager selects the shooting mode with the android system and the controller leads to the display output, the audio output, and changes the brightness depending on the functions selected by the users.

4. Hardware Description

4.1 Arduino Nano

Arduino is based on computer hardware and in electronic platform it is open sourced and this is very easy for using in environment. This is designed for persons interested in making objects, artists, hobbyists or collaborative spaces. It senses the environment in absorbing the input from other sources and it can affects the environment in controlling the engines, light and objects in it. The microcontroller board is configured by language of Arduino system and Arduino development environment. The projects based on Arduino can create on their own ideas or interacting with software in computer as like in processing, flash. This Nano type Arduino is powered by fixing a mini-B USB, 5V external power supply in the pin 27. Then the source for power will be automatically selected from the power source of highest value.

4.2 Arduino Mega

This mega type of Arduino micro controller board is ATMega2560 based model. This type of Arduino holds 54 digital input and output pins in which 15 of which used

for PWM output, 16 analog input pins, 4 UART, 16 MHZ crystal oscillator, communication USB jack, reset button and ISCP header in it. This type of Arduino has all things which is needed in supporting the micro controller. This device can be used easily connected with the USB to the computer or it can be connected to power with an AC to DC adapter or can be stored with the help of battery. This Mega2560 Arduino has come to be an update to Arduino mega. This mega Arduino board is more consistent with more UNO shoes with previous boards Diecimila.

4.3 Bluetooth

HC-05 is a type of module which combines two-way multi functionality in our project. We can use this type of module to combine the two micro controllers to connect with any device also the functionality of Bluetooth with it such as laptop or mobile phones. Already many android applications are created to make this process much easier. This type of module will be able to communicate with USART at a rate of 9600 band and this makes the work easier to support USART with any type of micro controller. In this we can also make fixed values using the mode of command option. This module might be the good option when we look for a wireless module which transmit the data to a micro controller from computer or mobile phone. This module does not transfer multimedia application like images, songs and we can refer for that CSR8645 module. This HC-05 combines of two active modes namely, one in which data mode can be send and receive data from Bluetooth devices. Whereas, the other for AT command mode to change the default device settings. We can use this device by using the pin as mentioned in the description of the pin.

4.4 Speaker

Digital Speaker works on the 8002-speaker amplifier IC. This speaker is used as a buzzer or for Arduino with high playing quality. The volume of the speaker is managed with a potentiometer. This module has large benefits like intuitive interface, long life of working, compact text. This module can be combined with Arduino input and output expansion shield in related and sound creation.

4.5 Power Supply

As a result, as compared to other 12V batteries, this one is relatively light. The battery life will stay longer upto 3-5x than the SLA or VRLA battery equivalent. The light weight battery has 30% of SLA weight with 12v of system charging in it. It remains more strength through extraction when compared to lead. This can be end up in payment status left out being comprised. It is much safer than traditional lithium-ion batteries, sealed with a fabric. It is built on top of control and on top of release protection. The type of the product is of 12v 17Ah lead acid battery replacement. The various applications are ready to start with motor cycle, medical instruments, solar power, power supply in LED's ,12v DC power. This makes installation easy. It reduces costs and adds complexity. Battery life is more than 6 months for pre-configured transmitters and receivers. IrDA transceivers makes advantages in addition to reducing energy consumption.

5. Software Description

5.1 Arduino IDE

Integrated Development with Arduino The environment consists of a text editor for writing code, a message area, a text console, a toolbar with bottom for common functions and a series of menus. It connects to the Arduino and genuine hardware to upload programs and communicate with them. Program written using Arduino software (IDE) are called sketches. These sketches are written in the text editor and saved with file extension. No. The editor has features for cutting/pasting and searching/replacing text. The message area given feedback while saving and exporting and display errors as shown in Figure 1.
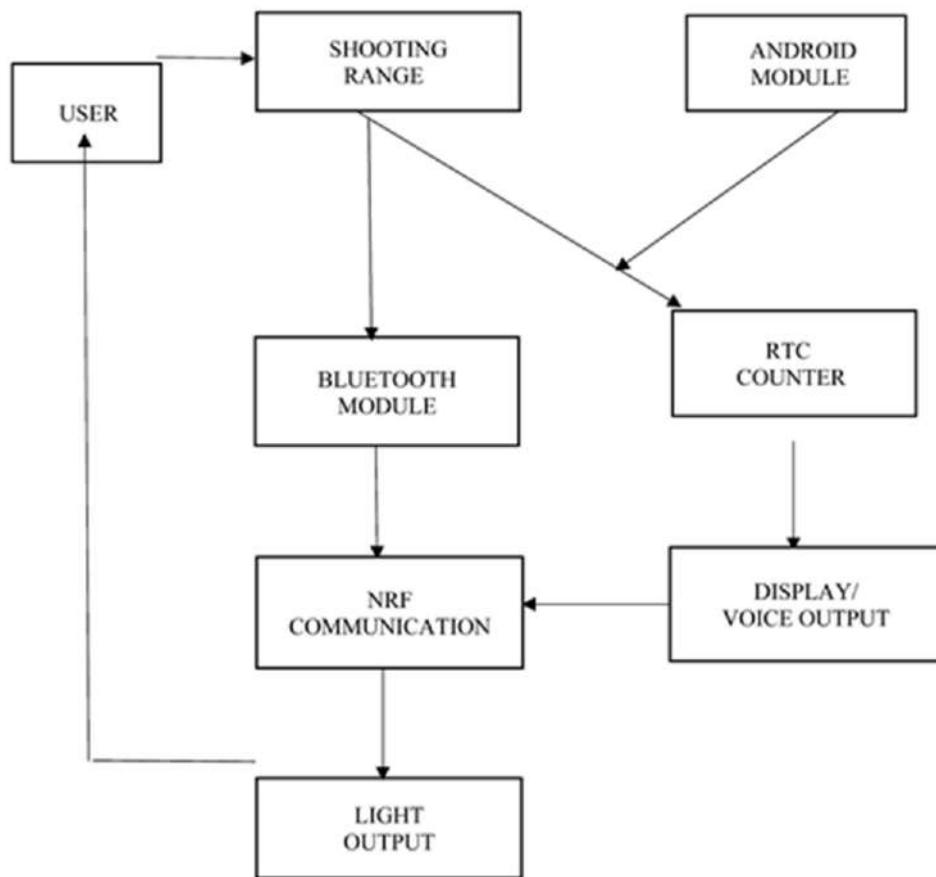
Figure 1. Block Diagram of Proposed System

5.2 Java

The Java platform is an appropriate computer platform. All platforms are supported, from servers to mobile phones to smart cards. These java technology combines the business activity to produce a seamless, protective platform and network in our business. This platform of java has benefits from a wide community of active workers to develop and support to deliver products and services designed for java technology and to create the platform by an open standard organization and based on community known as the java community process. We can see java technology in laptops, internet, mobile phones and trackside at formula of one grand prix races.

6. Hardware Results

The output system consists of an NRF module, display, audio amplifier, and speaker. The sensors are connected to read various posters as well as user positions in order to detect errors made by them and correct the errors made by them. The gunman decides how to shoot with android application through display. And the controller leads to display output, audio output, and light switching based on user-selected functions as shown in Figure 2 and Figure 3.

Figure 2. Sensor Display Controller   **Figure 3.** NRF Module

7. Future Scope

The system is utilised for a variety of purposes, including directing people in wheel chairs who are using IOT-based robotic operating systems. As IoT technology grows to be employed in almost all sectors, the associated production costs decrease. The idea of setting up this system is useful fry all games is possible legally. As the usage of IoT in the commercial sector continues to rise, production costs will be significantly reduced due to increased production, allowing IoT to be accepted in all games and into machine learning and archiving of sports and spontaneous training time.

8. Conclusion

This project includes a large Arduino Mega controller that is used for all arithmetic and logic units in all calculations. The Bluetooth, RT, and SD ard modules were employed as input devices in this project. The NRF module, display, audio amplifier,

and speaker make up the output system. The sensors are connected to read a variety of posters and user statuses. Using the display, the gun manager selects the shooting mode with the android system. And the controller connects to the display output, the audio output, and adjusts the brightness based on the functions selected by the users. The proposed method reduces reliance on the central database, which may necessitate certain more time-consuming tasks in order to compile more data. Because each marker is managed separately to generate time, this program is incredibly useful and takes very little time.

References

[1] Exposito, I., & Cuinas, I. (2013). Exploring the Limitations on RFID Technology in Traceability Systems at Beverage Factories. International Journal of Antennas and Propagation, 2013, 1–9. https://doi.org/10.1155/2013/916526

[2] Huang, Y., Cai, X., Kan, W., Qiu, S., Guo, X., Liu, C., & Liu, P. (2017). A flexible dual-mode proximity sensor based on cooperative sensing for robot skin applications. Review of Scientific Instruments, 88(8), 085005. https://doi.org/10.1063/1.4998995

[3] S., T., & S. Liu, J. W. (2013). Design and Implementation of RFID-Based Object Locators. In Radio Frequency Identification from System to Applications. InTech. https://doi.org/10.5772/53576

[4] Real, & Araujo. (2019). Navigation Systems for the Blind and Visually Impaired: Past Work, Challenges, and Open Problems. Sensors, 19(15), 3404. https://doi.org/10.3390/s19153404

[5] Bai, J., Lian, S., Liu, Z., Wang, K., & Liu, D. (2018). Virtual-Blind-Road Following-Based Wearable Navigation Device for Blind People. IEEE Transactions on Consumer Electronics, 64(1), 136–143. https://doi.org/10.1109/TCE.2018.2812498

Enhanced Data Privacy Using Vertical Fragmentation and Data Anonymization Techniques

R Sudha [a,1], G Pooja [a], V Revathy [a] and S Dilip Kumar [a]

[a] *Department of Computer Science and Engineering, Arasu Engineering College, Kumbakonam, Tamilnadu, India*

Abstract. The use of online net banking official sites has been rapidly increased now a days. In online transaction attackers need only little information to steal the private information of bank users and can do any kind of fraudulent activities. One of the major drawbacks of commercial losses in online banking is fraud detected by credit card fraud detection system, which has a significant impact on clients. Fraudulent transactions will be discovered after the transaction is completed in the existing novel privacy models. As a result, in this paper, three level server systems are implemented to partition the intermediate gateway with better security. User details, transaction details and account details are considered as sensitive attributes and stored in separate database. And also data suppression scheme to replace the string and numerical characters into special symbols to overcome the traditional cryptography schemes is implemented. The Quassi-Identifiers are hidden by using Anonymization algorithm so that the transactions can be done efficiently.

Keywords. Anonymization, Quasi-identifiers, Credit Card fraud Detection System, Traditional Cryptography, Three-Level Server Systems, Intermediate Gateway, Data Suppression Scheme.

1. Introduction

The upcoming methodologies of communications techniques, online payment transactions are increasing day by day. The major drawback for internet banking in current business is that fraudulent transactions appear frequently like authentic ones and it is not efficient to use the simple pattern matching techniques. We can implement vertical clustering algorithm to cluster the datasets into more than one level. Subsets of properties (that is, segments) structure the pieces [1]. Rows of the parts that relate to one another must be connected by an identifier that identifies tuple. A vertical fragmentation relates to a procedure that undergoes projection on the table. The recombination of data derived from the fragmented tuples is done to get original dataset. The joint operator is utilized for identifying tuple to link the columns in vertical clustering in order to link column to the fragments whereas the union operation is used on the rows of the datasets in horizontal clustering.

---

[1] R Sudha, Department of Computer Science and Engineering, Arasu Engineering College Kumbakonam, Tamilnadu, India; E-mail: suthathilo@gmail.com.

Given person-specific personal data, produce an arrival of the information with logical ensures that the people which are the subjects of the information can't be re-recognized while the information remain basically helpful. If the output information contains the information of each person then it said to have k-anonymity property and it cannot be differentiated from at least k-1 individuals whose information also appear in the output dataset.
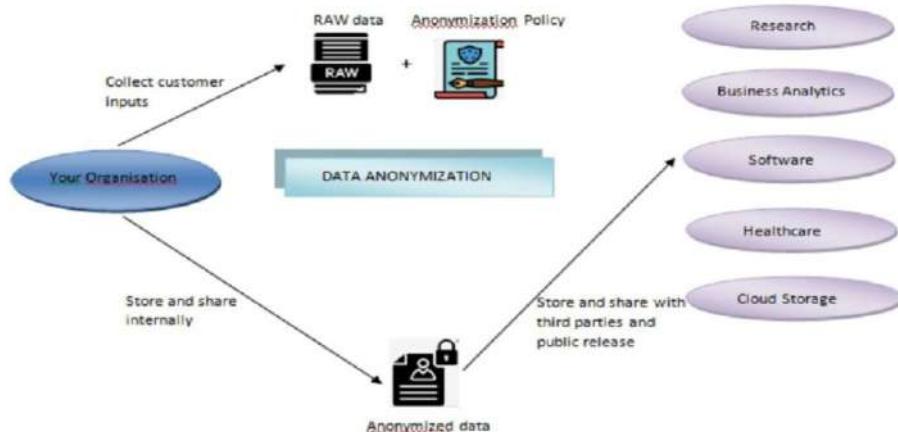
Figure 1. Data Anonymization

2. Existing models

In Existing paper, they have used traditional privacy preserving models, l-diversity and t-closeness. L-diversity make less the granularity in information portrayal utilizing strategies like generalization and suppression. If there were l no of well-versed values for the Quassi identifiers then that set of attributes is said to be l-diversity [2]. The distribution of the class which have equality is required by T-closeness which is moreover close to the attribute distributed in the whole table is shown in Figure 1. The necessitates of this method is the appropriation of a Quassi identifier in the identical class is near the conveyance of the sensitive attribute in the full table. These two strategies have certain drawbacks and prone to attacks called homogeneity attack and background knowledge attack.

The basic assumptions of this paper are the various attributes present in the database. The attributes which explicitly identify individuals in a database are called explicit attributes. To identify individuals the attributes combine with other tuples which are known as Non-sensitive and Sensitive QID's. The different methodology and projects for creating anonymised information giving k-anonymity security have been licensed.

3. System architecture
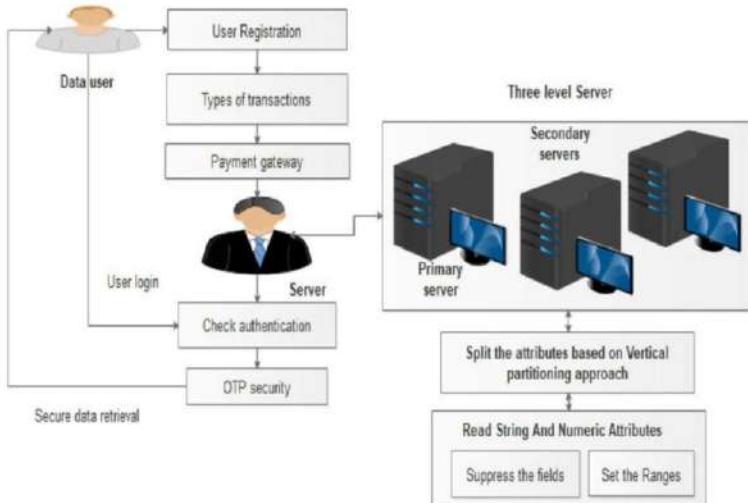
Figure 2. System architecture

4. Proposed Privacy Models

With the approach of correspondences strategies, web based business just as online installment exchanges are expanding step by step. In Figure 2, the serious issue for internet business today is that fake exchanges show up an ever increasing number of like real ones and straightforward example coordinating with strategies are not proficient to recognize misrepresentation [3]. Vertical clustering will do the projection operations on the table. Information from the pieces can be combined again to bring about the first informational index. For vertical clustering, the join administrator is utilized on the tuple identifier to interface the segments from the sections; in even fracture, the association administrator is utilized on the columns resulting from the pieces is shown in Figure 3. The different techniques and projects for producing anonymised information giving k-anonymity protection have been implemented.

4.1 Vertical Clustering Algorithm:

The Vertical Clustering Algorithm is a technique that fragments the records in column-wise manner and stores them in separate databases. The personal details, transaction details and account details in separate servers. There are two servers namely, primary server and a secondary server to store the fragmented record. The personal details are stored in primary server [4] . The Transaction details and account details are stored in a Secondary server. Vertical clustering is a strategy that segments the whole dataset into a few quantities of little data sets dependent on the segment, with the end goal that the fragmented data set doesn't have any copy data. There are basically two sorts of vertical information base specifically standardized and column portioning.

Name	DOB	Address	Last Transaction time	Last Transaction amount	Account number	Account type
Alice	07.08.1991	Aaaa	18.00	Rs.22,000	1468	Savings
Bob	06.05.1988	Bbbb	22.12	Rs.22,56,000	8359	Current
Carmalina	06.10.1993	Cccc	08.56	Rs.17,00,000	9531	Savings
Daisy	30.07.1997	dddd	09.17	Rs.12,56,269	3852	Current

Name	DOB	Address
Alice	07.08.1991	Aaaa
Bob	06.05.1988	Bbbb
Carmalina	06.10.1993	Cccc
Daisy	30.07.1997	dddd

Personal details

Last Transaction time	Last Transaction amount
18.00	Rs.22,000
22.12	Rs.22,56,000
08.56	Rs.17,00,000
09.17	Rs.12,56,269

Transaction details

Account number	Account type
1468	Savings
8359	Current
9531	Savings
3852	Current

Account details

Figure 3. Vertical clustering

4.2   K-anonymity Algorithm:

The k-anonymity models have been implemented for protecting individual identification. Ongoing investigations show that a more refined model is important to secure the relationship of people to sensitive Quassi data. In Figure 4, generally, protection is estimated by the data gain of a spectator. Prior to seeing the delivered table, the spectator has some earlier conviction about the delicate trait estimation of a person [5]. Our methodology is that we separate the data acquire into two sections: that about the entire populace in the delivered information and that about explicit people. The following expressions are very difficult for understanding the remaining of the project: Quasi-identifier (QI): a bunch of characteristics which can be utilized with particular external data to distinguish a particular person. T, T[QI]: T is the given dataset represented in a relational form T[QI] is the protrusion of T to the arrangement of characteristics contained in QI. Tk [QI].

In this module, we can implement K-Anonymity for secure the data privacy. Some secret data maintains a property called K-anonymity to hide the private data. The two main types of methods for maintaining k-anonymity for all the values of k. The Suppression technique, particular estimations of the credits are replaced by special characters '*'. Suppose all or may be a few estimations of a segment might be replaced by '*'. In Generalization technique the replacement with the most common category is done in this method, so that each and every value of attributes are replaced by with a broader category. We tracked down that a speculation testing approach gave the best authority over re-ID hazard and decreases the degree of data misfortune contrasted with pattern k-anonymity [6]. Certainly, The guarantee of achieving k-anonymity with the procedure of replacing every selected cell with a special character *, but this will result in leaving the database with no meaning. The expense of K-Anonymous answer for a data set is the quantity of *'s presented. A base expense k-anonymity arrangement chooses the least number of cells important to ensure k-anonymity.

Original data

Name	Last Transaction time	Last Transaction amount	Account number	Account type
Anbu	18.00	Rs.22,000	1468875759	Savings
Baskar	22.12	Rs.22,56,000	8359997888	Current
Chithra	08.56	Rs.17,00,000	9531987597	Savings
Dharshan	09.17	Rs.12,56,269	3852897597	Current

Name	Last Transaction time	Last Transaction amount	Account number	Account type
Anbu	18.00	Rs.22,000	14********	Savings
Baskar	22.12	Rs.22,56,000	83********	Current
Chithra	08.56	Rs.17,00,000	95********	Savings
Dharshan	09.17	Rs.12,56,269	38********	Current

Anonymized data

Figure 4. Data anonymization example

5. Conclusion

The main goal is to protect the personally identifiable information i.e. Quassi identifier in data privacy which is most wanted in online banking web application. Generally, if the information is linked directly or indirectly to a person or individual which they possess is called personally identifiable information. Thus, at the point when individual information are exposed to mining, the attribute esteems related with people are private and should be shielded from information theft. Instead of learning from the characteristics of a single individual, miners may learn from global models. In this paper, we can conclude that the proposed system provide improved security in cloud data. Vertical partitioning and K-Anonymity are two approaches that we can use. In data mining, K-Anonymity is a privacy-preserving technique for preventing the disclosure of private information. When anonymizing a database table, the procedure usually entails generalizing table entries, which results in the loss of relevant data.

References

[1] Chahar, H., Keshavamurthy, B. N., & Modi, C. (2017). Privacy-preserving distributed mining of association rules using Elliptic-curve cryptosystem and Shamir's secret sharing scheme. Sādhanā, 42(12), 1997–2007. https://doi.org/10.1007/s12046-017-0743-4

[2] Gunawan, D., & Mambo, M. (2018). Set-valued Data Anonymization Maintaining Data Utility and Data Property. Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication, 1–8. https://doi.org/10.1145/3164541.3164583

[3] Xiong, X., Chen, F., Huang, P., Tian, M., Hu, X., Chen, B., & Qin, J. (2018). Frequent Itemsets Mining With Differential Privacy Over Large-Scale Data. IEEE Access, 6, 28877–28889. https://doi.org/10.1109/ACCESS.2018.2839752

[4] Wang, J., Deng, C., & Li, X. (2018). Two Privacy-Preserving Approaches for Publishing Transactional Data Streams. IEEE Access, 6, 23648–23658. https://doi.org/10.1109/ACCESS.2018.2814622

[5] Wang, H., He, D., & Tang, S. (2016). Identity-Based Proxy-Oriented Data Uploading and Remote Data Integrity Checking in Public Cloud. IEEE Transactions on Information Forensics and Security, 11(6), 1165–1176. https://doi.org/10.1109/TIFS.2016.2520886

[6] Liang, K., Huang, X., Guo, F., & Liu, J. K. (2016). Privacy-Preserving and Regular Language Search Over Encrypted Cloud Data. IEEE Transactions on Information Forensics and Security, 11(10), 2365–2376. https://doi.org/10.1109/TIFS.2016.2581316

A Robot for Combat and Calamities with Encrypted WSN

S Abiraj [a,1], D Eeswar Samhithan [a], B Sanjai Kumar [a], Y Venkatesh [a]

ᵃ Department of Electronics and Communication Engineering Vel Tech High Tech Dr Rangarajan Dr Sakunthala Engineering College, Chennai, TN, India

Abstract. The needs of search and rescue teams in numerous ways reflect the requirements of the military. They both work in perilous situations, they need to discover ways to gather data while keeping personnel out of hurt, and they are both seeking out for individuals. So, this Robot Commands in manual mode using a Smartphone, Wireless cam for Real-time Broadcast, PIR and Metal detection sensors, GPS & GSM modules, ultrasonic sensor, Gas & Fire sensor, temperature and humidity sensor, Acid gun, RFID and Relay switch, In this wireless sensor networks (WSN) is an Encrypted Network and infrastructure-less wireless networks to Screening physical or natural conditions, such as vibrations, pressures, movement, or toxins and to agreeably transmit their information retrieved by using IOT Server.

Keywords: Passive Infra-Red (PIR), GPS, GSM, Radio frequency identification (RFID), wireless sensor networks (WSN), IOT.

1. Introduction

Recent sensational occasions such as the seismic tremors in Indonesia and japan or the numerous surges in Bangladesh had appeared that neighborhood respectful specialists and crisis administrations have inconveniences in satisfactorily overseeing emergency. Subsequently at any fiasco, to begin with 48 hours is pivotal for protecting victims. Fire, police department personnel work force and paramedics are being sent to play down the misfortune of resources, for diminishing misfortunes of victims and for getting exact data of the circumstance, a mechanical system can be utilized and can be adjusted concurring to desires of the circumstances.

2. Literature Survey

A low power wireless Miniature robot a seen from nowadays to very next day shows up cutting edge contraptions & weapons are utilized by equipped forces for decreasing such peril for their life and to conquer their foes. Within this progression of modern innovation, so that it majorly depends on the forefront weapons or state of the craftsmanship systems being utilized [1-3]. Front line assistive bot that can be characterized as a bot that ousts the mines in battle all on its purpose and could be utilized for surveillance purposes on the foes.

---

[1] S Abiraj, Department of Electronics and Communication Engineering, Vel Tech High Tech Dr Rangarajan Dr Sakunthala Engineering College, Chennai, TN, India, E-mail: abiraj.ece17@velhightech.com.

Those bots are really controlled remotely. The versatile secure robot comprises of four section that is specifically the Sensors, Micro-controller, Camera, Motor driver and Transmission unit. [4-6]

3. Proposed System

The Proposed system of our Bot consist of an operating system, in which we utilized Arduino UNO board, from now on Embedded C programming is utilized for commanding and to communicate with bot. The flow chart in Figure.1

3.1 Prototype Explanation

The Layout is appeared underneath within the Figure.2 portrays complete robotic prototype which outlines the Function of the whole prototype. In this Venture Microcontroller we utilized is Arduino Uno created by Arduino.cc.it is an open- source contraptions arrange basically based on Automatic voltage regulator (AVR) microcontroller Atmega328. Arduino Nano might moreover be a small, total, and breadboard-friendly board upheld the ATmega328. Passive infrared Ray (PIR) sensor is utilized to detect the development of alive victims within 6 to 7m region.

 An ultrasonic sensor may be a sensor that measures the space of a target address by transmitting waves that changes over the reflected sound into an electrical pulse. Metal Sensor identifies the metallic question which is display following to their dynamic side. This sensor works beneath the electrical foremost of inductance where a fluctuating current actuates a voltage (EMF) amid a target object. Gas sensors (too known as gas detectors) are electronic devices that distinguish and distinguish distinctive sorts of gasses. They are commonly utilized to distinguish poisonous or unstable gasses and degree gas concentration. A Humidity & Temperature sensor (or hygrometer) resources, measures and reports both dampness & examine the temperature. ESP32-CAM may be a low fetched advancement Wi-Fi cam. This permits the IP cam for making ventures for live video spilling with diverse resolution. The ESP8266 is a Wireless Module can be a self-contained SOC (system on chip) with arranges TCP/IP tradition stack that can provide any microcontroller get to your wireless network. GPS (global positioning system) is utilized for pursing the arrange at an occurrence. GPS get the information from the satellite and save within its memory, at whatever point the information required is been transmitted by micro controller to base station. It may be a scaled down GSM (Global System for Mobile Communications) modem to achieve nearly anything a standard smart phone can, text messages through SMS, it can get or make phone calls, interfacing to network through GPRS (General Packet Radio Service). EM18 is a RFID reader which is utilized to read RFID (Radio-frequency identification) labels of frequency 125 kHz. After reading labels, it transmits unique ID serially to the PC.NRF is a transceiver module which implies that it can both send and get the information. The motor driver is utilized as an enhancer, to open up the supply voltage to the motors. Motor Driver L298N Module is a high-power motor driver module for driving stepper and DC Motors. which can be controlled in both clockwise and anticlockwise heading. The IOT server is utilized by establishing the connection between the robot and the user which can be controlled in both clockwise and anticlockwise heading. The IOT server is utilized by establishing the connection between the robot and the user. The electrically operated Relay switch comprises of input terminals for a single or different control signals and consist of

working contact terminals. The robot is equipped with is Acid gun is actuated in case the suspicious human is unrecognized. A cellular smartphone phone is created with mobile OS which incorporates a few highlights of a PC operating system. This Smartphone has a numerous highlight such as, Internet, GPS and media player it acts like a personal digital assistant.
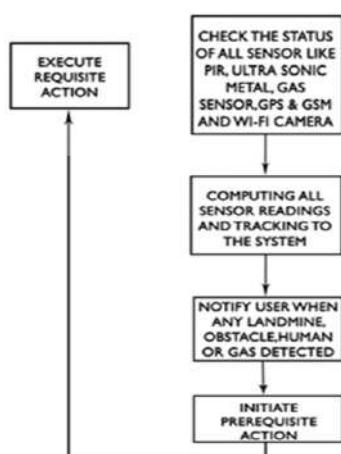
Figure 1. The Proposed system flow chart of the system
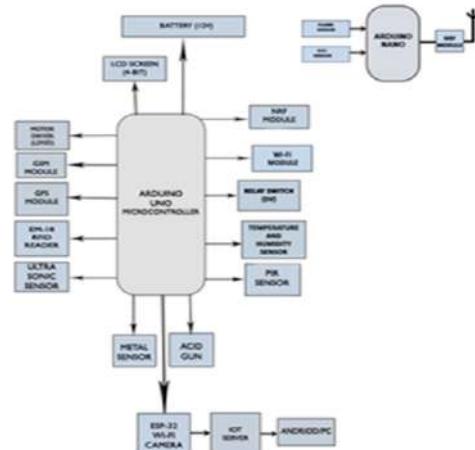
Figure 2. Layout of the proposed system

3.2 Software Description

The Flow Chart of Software description is shown in the Figure.3 and the following programming tools are linked with the prototype and programed systems 1. Arduino IDE   2. Embedded C. The Arduino IDE is a computer program which combines a program editing, Robotization tools, and debugging to build unique programmer for program headway. With a single user interface, Arduino IDE oversees all perspectives of embedded device advancement. The C Standards committee made Embedded C as a collection of language extensions for C programming to report issues and to provide medium-level get memory, and create a code creates format well for machines informational which needs negligible support to run-time time.
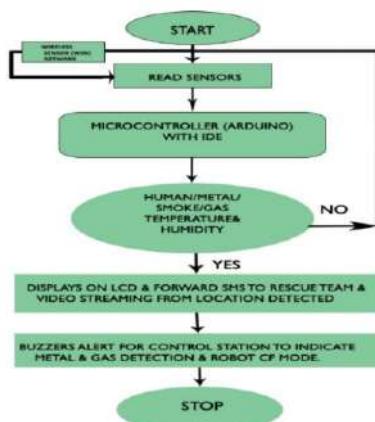
Figure 3. Software implementation flow chart
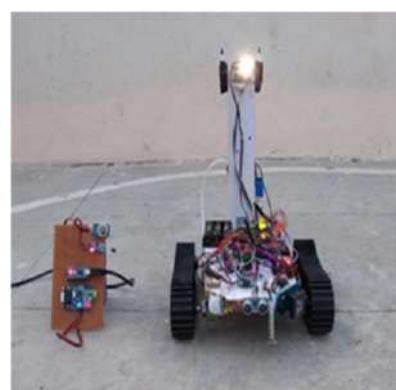
Figure 4. Final Prototype of Robot

4. Design Flow

The Design Flow of this Robot is in hand-operated mode and it is shown in Figure.5 In this Venture we utilized Portable device for informational is given to the controller which is associated through IP address, on the premise of given informational the bot wanders through like forward, in reverse, right, left, and stops in the event that any enlightening are not given it is appeared in Figure.6 We moreover utilized PIR sensor on the robot, at whatever point it faculties the movement inside the surmised area of 10m, the bot halt and informs user utilizing GSM that "Intruder detected". Wi-Fi cam will turn & commanded by the user to testify around the object. The Acid shooter enacted in the event that necessary and on the off chance the Victim is suspicious, the Victim can be encountered on location. we used the Metal Sensor to identify landmines & explosives buried underneath land. bot halt & area is followed by utilizing GPS. Coordinate values are followed & gotten by the user. So that, the landmines & Bombs is being dispensed. Ultrasonic Sensor, at whatever point a bot moving in a direction ultrasonic sensor transmits an ultrasonic pulse out at 40kHz which voyages through the air and in the event that there's a deterrent or obstacle, it'll recuperate through the sensor. In this Venture we used RFID tags used to read frequency 125 kHz. the tag comes in run of signal transmitted by the reader validate the card, in case an Invalid card is identified, at that point Caution SMS (short message service) is sent to user.

5. Results and Discussion

The robot is in manual mode by pressing the commanding button. A couple of detectors like gas and metal Sensors. which is executed for creating bot extra viable for distinguish harmful metals and gaseous and Information of this all is stored in IOT Server and its snapshot appeared below in Figure.7 Alert SMS is sent to the client on the off chance that any Victim or Metal is Detected is shown in Figure 8
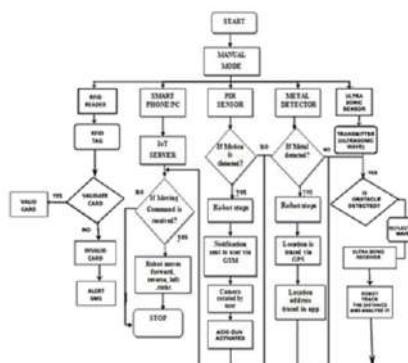
Figure 5. Flow chart of the robot operated in manual mode   **Figure 6.** Operational Controlling of the Robot
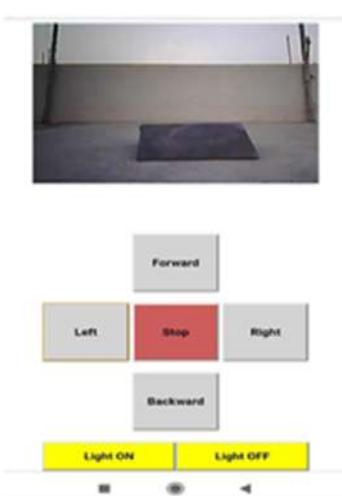
Figure 7. Snapshot of IoT Server
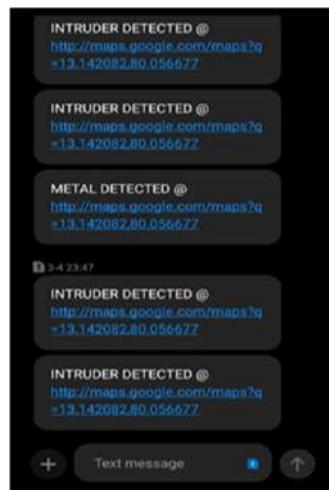
Figure 8. Snapshot of Alert SMS with Location.

6. Conclusion and Future Scope

Overall, this robot may be an exceedingly utilitarian gadget that decreases the workload on people in the midst of Disaster. Moreover, functions as a Surveillance vehicle to distinguish Armory items. In the midst of emergency & especially in Metropolitan fiascos, this framework will be exceedingly valuable. The misfortune can be detected quickly and offer assistance comes to the victims. The advance innovation to be created out to surmount confinements of the Bot in such a way failure of this Bot will wander within tropical regions. For measures & prototype of this Bot would be further diminished by utilizing inventive prototype and progressed fabric foe constructing Bot.

References

[1] Telkar, A. K., & Gadgay, B. (2020). IoT Based Smart Multi Application Surveillance Robot. 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 931–935. https://doi.org/10.1109/ICIRCA48905.2020.9183289

[2] Anand Vijay, K. M., Purustut, S., Suhas, R., Pavan, C., & Pradhan, P. (2018). A Live Human Being Detector in War Fields and Earthquake Location Using Robot with Camouflage Technology. 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2244–2249. https://doi.org/10.1109/RTEICT42901.2018.9012601

[3] Aniruddha Prabhu, B. P., & Hebbal, S. (2017). Small Unarmed Robot for Defense and Security:A Cost-Effective Approach Using Arduino Uno. 2017 2nd International Conference On Emerging Computation and Information Technologies (ICECIT), 1–6. https://doi.org/10.1109/ICECIT.2017.8453325

[4] Joshi, S. A., Aravalli, G., Vidyashree, A. K., Ranade, S., & Badami, S. S. (2017). Wireless controlled military combat robot system. 2017 2nd International Conference on Communication and Electronics Systems (ICCES), 712–715. https://doi.org/10.1109/CESYS.2017.8321173

[5] Murphy, R. R., Tadokoro, S., & Kleiner, A. (2016). Disaster Robotics. In Springer Handbook of Robotics (pp. 1577–1604). Springer International Publishing. https://doi.org/10.1007/978-3-319-32552-1_60 Aishwarya K Telkar and Prof.Baswaraj Gadgay, "IOT Based Smart Multi Application Surveillance Robot", Second International Conference on Inventive Research in Computing Applications (ICIRCA-2020), pp. 931, Sept 2020.

Diabetes Risk Forecasting Using Logistic Regression

Metharani N [a,1], Srividya R [a], Rekha G [b] and Ranjith Kumar V [c]

[a]*Department of CSE, Adhiparasakthi Engineering College, Melmaruvathur, India*
[b]*Department of CSE, St Peter's Institute of Higher Education & Research,*
Chennai, India
[c]*Department of Mechanical Engineering, Sri Sairam Engineering College,*
Chennai, India

Abstract. Diabetes can be a collection of metabolic problems and lots of human beings are affected. Diabetes Mellitus can be caused by a variety of factors including age, stoopedness, lack of activity, inherited diabetes, lifestyle, poor eating habits, hypertension, and so on. Diabetics are more likely to develop diseases like coronary illness, kidney contamination, eye sickness, stroke and other risks. Distributed computing and Internet of Things (IoT) are two instruments that assume a vital part in the present life with respect to numerous angles and purposes including medical care observing of patients and old society. Diabetes Healthcare Monitoring Services are vital these days on the grounds that and that to distant medical care observing in light of the fact that truly going to clinics and remaining in a line is exceptionally ineffectual adaptation of patient checking. Current practice in emergency clinic is to gather required data for diabetes conclusion through different tests and proper treatment is given dependent on analysis. Utilizing enormous data investigation can consider large datasets and discover covered up data, uncertain examples to find information from the data and expect the outcome as demand. Diabetics are caused because of a tremendous uphill in the blood partition containing glucose. There is an advancement conspire accessible using train test split and K overlay cross approval utilizing Scikit learn technique. Various ML algorithms consisting of SVM, RF, KNN, NB, Decision Tree and Logistic Regression are also used.

Keywords. Indian Pima Diabetes Dataset, Decision Tree, K Nearest Neighbor; Random Forest, Logistic Regression, Naïve Bayes

1. Introduction

This paper needs some exacting activities in regards to the control and counteraction of diabetes. Prior the facts previously demonstrated that around one in each ten individuals in US had diabetes. In any case, expectations have been made that by 2045 it might help up to one in each three people. This is a difficult issue we need to manage. The constant sickness of diabetes results right into it when there is an immense expansion in the blood glucose focus. This is a significant reason for different issues and infections, for example, kidney illnesses heart issues. Numerous unfortunate dietary patterns and absence of appropriate body practices likewise causes the diabetic

---

[1] Metharani N, Department of CSE, Adhiparasakthi Engineering College, Melmaruvathur, TN, India; E-mail: metharani53@gmail.com.

pre conduct. It has been expressed by the WHO that the absolute check of individuals experiencing diabetes has inconceivably expanded in the course of recent years.

Managing numerous diabetic datasets is obligatory to improve the current pace of patients experiencing diabetes and to take it to an unadulterated insignificant level by zeroing in on to lessening it by huge scale. Type 1, Type 2, and gestational diabetes are the most well-known forms of diabetes.

Type 1 diabetes: It is indeed a type of diabetes for which the body your body does not produce insulin if you have type 1 diabetes. It is very often evaluated in kids and teens, spite of the fact that it can show at whatever age.

Type 2 diabetes: You can create type 2 diabetes at whatever stage in life, in any event, during adolescence.

Gestational diabetes: Gestational diabetes creates in certain ladies when they are pregnant. More often than not, this sort of diabetes disappears after the child is conceived. Notwithstanding, on the off chance that you've had gestational diabetes, you have a more prominent possibility of creating type 2 diabetes sometime down the road. The patient risk level is classified using data mining techniques such as K nearest neighbor, Decision tree, Random forest, Logistic Regression and Naïve Bayes.

2. Literature Review

Defusal Faruque and Asaduzzaman, Iqbal H.Sarker mentioned that polygenic disease is one among the foremost common disorder of the material body it's caused due the metabolic disorder. They used various and important ML algorithms that are Support Vector machine, NB, KNN and DT to predict diabetes [1]. Sidong Wei, Xuejiao Zhao and Chunyan Miao presented that diabetes is mostly called as disorder in which glucose level in the body is high. In this paper they use popular methods such as SVM and a deep neural network for identify the disease and data processing [2]. Jian-xunChen, Shih-LiSu and Che-Ha Chang discussed about Ontology that generate a primary care planning to the medical professionals. The result of the research paper shows the model can be provided personalize diabetes mellitus care planning efficiently [3]. MM Alotaib, RSH.Istepanian, and A.Sungoor they are present a clever based mobile polygenic disease control system & tutoring model for the patients with diabetes. In this, a system is able to store the clinical information about the diabetes system, such an often blood sugar level and BP is measured and hypo glycaemia event [4]. Berina Alic and Lejila Gurbea,Almir Badnjevic they presented the overview of techniques in the machine learning in the diabetes classification and cardiovascular diseases using BNs and ANN [5]. ElliotB.Sloane, Nilmini Wickramasingle and Steve Goldberg they presented Wireless diabetes monitoring which is a cloud-based diabetes, it's a coaching platform for diabetes management and low cost, innovative, cloud-based diabetes support system . Minyechil Alehegn and Rahul Joshi had present about the ML technology that help to identify a dataset at the elementary so that rescue the life.By implementing NB and K-nn algorithms . Umatejaswi and P.Suresh Kumar had discussed about algorithms such as SVM, NB, DT for identify the mellitus make use of technique like data mining .

3. Methodology

In this paper, we have utilized our dataset for applying different machine learning algorithms for recognizing if an individual has coronary illness or not is shown in Figure 1. At that point, we will deal with the missing qualities in the dataset, visualize the dataset and notice the precision acquired by various AI calculations. AI calculations utilized are characterized underneath.

4. Result

4.1 Correlation Matrix

From this Figure 1. diagram, we can see that a few features are exceptionally related and some are definitely not.
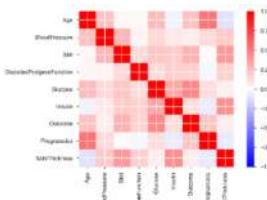
Figure 1. Correlation matrix

4.2 Histogram

A histogram is a statistical tool for the portrayal of the appropriation of the data set. It is an overall assessment of the likelihood conveyance of a persistent arrangement of variable information is shown in Figure 2. It is really a plot that answers all the inquiries with the fundamental recurrence appropriation of a bunch of nonstop and plausible information, it gives a feeling of the thickness of information.
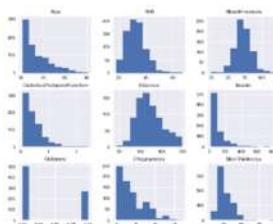
Figure 2. Histogram.

Outcome" is that the features in Figure 3.we have a tendency to be attending to predict, zero suggests that No diabetes, one means diabetes. Of these 768 knowledge points, 500 are tagged as zero and 268 as 1:
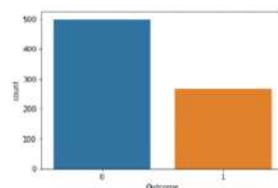
Figure 3. Count of Outcome variable

Classifying target variable between male and female and visualize the result is shown in Figure 4.

Box plot for target class with different features
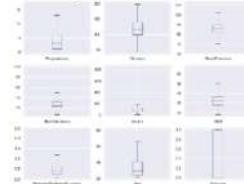
Figure 4. Target Class Box plot

Machine Learning Algorithms

4.3. Logistic Regression

The logistic regression, like all regression analyses, is a predictive analysis. To describe data and explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval, or ratio-level independent variables, logistic regression is used. In this paper, we achieved the accuracy of 73% by using this model.

4.4. Naive Bayes Classifier:

In short, the naive Bayes classifier expects that the presence of a specific capacity in one class has nothing to do with the presence of another capacity. Naive Bayes version is straightforward to construct and especially beneficial for extremely massive information sets. In this paper, we achieved the accuracy of 73% by using this model.

4.5. KNN Classifier:

The most straightforward AI calculation envisioned so far till now is the K-NN calculation because of its high exactness. The littlest reversible substances that make up the essential little units to work together and join to address a strong construction is this information that is thoroughly machine autonomous in all fields. The real factors and measurements those are altogether together consolidated with the end goal of reference and investigation is crude information objects. It is the fundamental methods for estimation and framework thinking for standardization purposes. The starting technique comes as the assortment of the information required and further articulating the given issue proclamation which characterizes the necessary information for framework examination. We should control the wonder of social event or probably joining the information and data from broadly accessible various sources. Presently our principle task comes as information preprocessing where three primary undertakings are played out those are organizing, cleaning and inspecting. In this paper, we achieved the accuracy of 78% by using this classifier.

4.6. Decision Tree Classifier

It learns to partition based on the value of an attribute. Recursive partitioning is a process of partitioning the tree in a recursive manner. This image is a framework. In this paper, we achieved the accuracy of 76% by using this model.

4.7. Random Forest Classifier

It is simple and easy to implement. A forest is comprised of trees. This classifier creates selection timber on randomly decided on records samples, receives prediction from every tree and selects the first-class answer with the aid of voting. The random wooded area composed of multiple selection timber. It creates a forest of trees. In this paper, we achieved the accuracy of 71% by using this model.

Table 1. Accuracy Values

Algorithms	Accuracy
Logistic Regression	73%
Naïve Bayes Classifier	73%
Decision Tree Classifier	76%
Random Forest Classifier	71%
KNN Classifier	78%

In Table 1, shows that KNN Classifier gives the best accuracy with 78% in assessment with the alternative system gaining knowledge of algorithms used in this paper.

5. Conclusion and Future Work

This document involves predicting the diabetes disease dataset with proper computing and implementation of machine learning algorithms. In this article, we will use five machine learning algorithms to make predictions.

Calculating Accuracy
Among all the machine learning algorithms used in this paper, the highest accuracy is achieved by K Nearest Neighbors Classifier with 87%. This article shows that the machine learning algorithms is accustomed predict the center sickness simply with totally different parameters and models. Machine learning is very useful in prediction, solving problems and other areas. Machine learning is an effective way to solve the problems in different areas too.

6. Future Enhancements

This may permit new calculation improvement to be performed off-site utilizing distributed computing programming, and afterward got back to the clinical setting as applications program interfaces (APIs) for PCs, cell phones and tablets.

References

[1]   Wei, S., Zhao, X., & Miao, C. (2018). A comprehensive exploration to the machine learning techniques for diabetes identification. 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), 291–295. https://doi.org/10.1109/WF-IoT.2018.8355130

[2] Rotheram-Borus, M. J., Tomlinson, M., Gwegwe, M., Comulada, W. S., Kaufman, N., & Keim, M. (n.d.). Diabetes buddies: peer support through a mobile phone buddy system. The Diabetes Educator, 38(3), 357–365. https://doi.org/10.1177/0145721712444617

[3] Alotaibi, M. M., Istepanian, R. S. H., Sungoor, A., & Philip, N. (2014). An intelligent mobile diabetes management and educational system for Saudi Arabia: System architecture. IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), 29–32. https://doi.org/10.1109/BHI.2014.6864296

[4] Alic, B., Gurbeta, L., & Badnjevic, A. (2017). Machine learning techniques for classification of diabetes and cardiovascular diseases. 2017 6th Mediterranean Conference on Embedded Computing (MECO), 1–4. https://doi.org/10.1109/MECO.2017.7977152

[5] Hsu, W. C., Lau, K. H. K., Huang, R., Ghiloni, S., Le, H., Gilroy, S., Abrahamson, M., & Moore, J. (2016). Utilization of a Cloud-Based Diabetes Management Program for Insulin Initiation and Titration Enables Collaborative Decision Making Between Healthcare Providers and Patients. Diabetes Technology & Therapeutics, 18(2), 59–67. https://doi.org/10.1089/dia.2015.0160

Sports Ground Audience Health Monitoring System [For COVID-19] Using Smart Wireless Sensor Network

Dharshini V [a,1], Aiswarya S [a], Ashwin R [a], Usha A [b]

[a]*UG Student, Department of Electronics and Communication Engineering*
[b]*Associate Professor, Department of Electronics and Communication Engineering*
Easwari Engineering College, Chennai, TN, India

Abstract. The recent spread of the COVID-19 has pushed us to a situation to reduce human contacts in crowded areas. The main objective of this work is to protect the audience from the spread of contagious and non-contagious diseases or infections by introducing automation in maintenance and management of sport stadiums. Taking some primary precautions such as social distancing, wearing masks, avoiding crowded areas, cleaning hands, and checking the temperature is very important. This work provides primary check-ups for the audience to pass through, to ensure that no symptomatic person is let inside the stadium. Social distancing is properly maintained using ultrasonic sensors placed inside the stadium, so that nobody sits over the sensor attached seats. To ensure the maintenance of audience data, tickets are being replaced with RFID tags and checking is done automatically through RFID readers. The information is stored and maintained automatically in the server through IoT.

Keywords. Automation, Check-up, Social distancing, Temperature

1. Introduction

In the last few years, we can see that healthcare has drawn a considerable amount of everyone's attention. Even during the time of widespread COVID-19 in our community, taking basic precautions are insisted to be followed. But in due time people are more likely to lead their routine lifestyle i.e., the one before COVID. Many public places were opened and there are gatherings too. So almost precautions must be taken in crowded areas to avoid the virus spread. This work tries to improve safety to prevent the spread of this virus. An automated security check-up in sports ground is introduced by which no symptomatic person is allowed inside the stadium. Here manual check-ups are avoided. Social distancing is followed by sensors implemented in-between seats. Temperature check is done to let persons inside and every individual data is stored through IoT. Various methodologies are been in usage for identifying the symptomatic person and for providing basic sanitization. Some of these methods involve humans for a check-up, where in some other places checking methods are automated.

---

[1] Dharshini V, Department of Electronics and Communication Engineering, Easwari Engineering College, TN, India; Email: dhachu.vpn@gmail.com.

Basic check-up like temperature is done with digital thermometers and digital infrared thermometers. Sanitization is mostly done by using a foot-operated sanitization stand with sanitizer in it. Both of these include manpower and there exists a constant contact of the checker with every individual which results in danger for both of them. George et al. [6] proposed this to find a solution for sanitization by using contactless sanitizers. This work connects an IR sensor and a motor with the sanitizer liquid bottle so that the sensor detects humans and sprays the sanitizer liquid.

The flow of liquid is controlled using a reducer connected to the pipe and the whole setup is connected with an RC timer delay. Here, three modes of control LEDs (White, Red, and Green) are used to provide a consumer convenient setup, where White indicates the setup is in working mode and battery is in usage, Red denotes the battery is in charging mode and Green specifies that the battery is in fully charged mode. It also saves costs and power.

Chiachung Chen et al. [3] proposed that instead of digital thermometers, infrared thermometers are more convenient, fast, and easy to use. Both tympanic and forehead thermometers are used to measure the body temperature. With the spread of the coronavirus, screening of infected people is primarily done with a temperature check. By this study, the performance of two types of thermometers- tympanic infrared thermometers and an industrial infrared thermometer was evaluated. Their results were proven to be with good precision. By keeping 36 °C as standard body temperature, temperature screening is done for individuals at forehead using infrared tympanic temperature and those who screen beyond the standard temperature are possibly considered ill. Guangli Long [7] proposed a solution to find body temperature in a fast and non-contact measurement using an infrared thermometer. It converts temperature from the infrared into a voltage signal. This system comprises an operational amplifier, filter circuit, and an ADC converter. An LCD is used to display the body temperature and a voice reporter is kept to give a voice-output of the measured temperature and time. The result displays the temperature, time of acquisition and, measurement error less than 0.5 °C. Md. Rezwanul Haque Khandokar et al. [4] proposed an automated way of ticketing using RFID tags in transportation industry. Uniform access is given for daily passengers and the data of passenger was updated automatically into the server each time they travel with the reader.

2. Materials and Method

The proposed automated monitoring system in stadium consists of two blocks, one is the entrance block and the other is the stadium block is shown in Figure 1. The sectional blocks comprise of the following components: Microcontroller, Power supply, LCD, IoT, Buzzer, RFID tag and RFID reader, Ultrasonic sensor, IR sensor, Temperature sensor, DC motor and Relay board. With these components, an automated system is created is shown in Figure 2 using Figure 1 as reference.
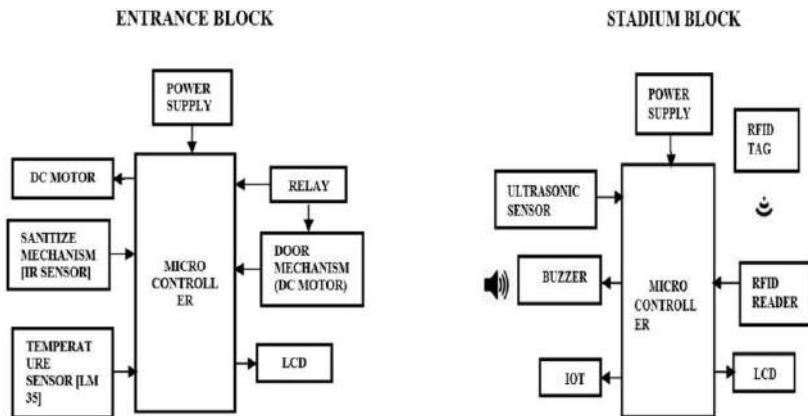
Figure 1. Block diagram of entrance block and stadium block
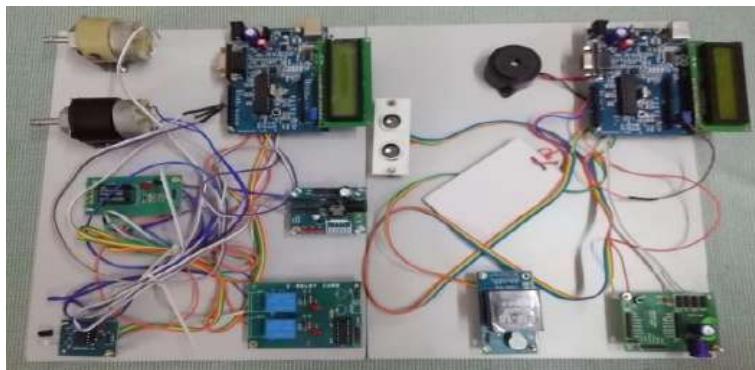
Figure 2. Hardware kit image

With respect to the block diagrams, the proposed system is mainly designed for monitoring the health condition of the audience in the sports stadium. The uppermost priority is to prevent COVID symptomatic persons from entering inside the stadium premises. This system comprises two microcontrollers and the entire block is split into two sections. In Figure 1, at entrance section the hands of each audience are sanitized using a touch-free sanitizing mechanism, i.e., using IR sensors placed at the entrance. Resistance Temperature Detectors (RTDs), a temperature sensor (LM35) is used for validating the audience's body temperature. Normal body temperature range is set over and persons exceeding that temperature range won't be allowed inside the stadium. After completion of this mandatory screening process, doors will be opened and access will be provided for the non-symptomatic audience i.e., persons with normal temperature. The audience is permitted to enter the allotted seats in the stadium section only after the verification of smart tickets using RFID tags and readers is shown in Figure 1. Individual tags are provided for the audience and they are read at the stadium entrance. The data like person's name, seat number, and ticket cost will be uploaded to the IoT server through a wireless network. Social distancing among the audience is monitored via ultrasonic sensors and intimated to the microcontroller with a buzzer sound. The main advantage of this system is that monitoring can be done remotely

without the spread of infection and it also reduces the time consumption for checking of each individual.

3. Results and Discussion

Once the power supply is given, the LCD connected with the microcontroller displays the current room temperature and the obstacles detected before the IR sensor as in shown in Figure 3. In entrance block, audience are first made to undergo basic sanitization which is done with the IR sensor is shown in Figure 2. In Figure 4,When the hands are detected before the sensor at a distance of 3cm and less the microcontroller displays a message with IR value as 1. It indicates that the first hand is detected and shows a message "show_ur_hand" and the sanitizer liquid a flow is shown in Figure 5.

Figure 3. LCD displaying initial temperature
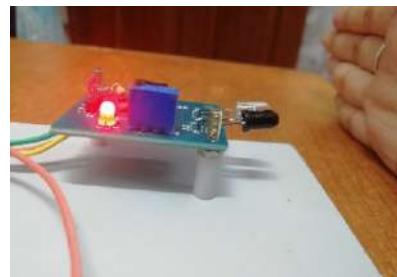
Figure 4. Hands placed over IR sensor

Figure 5. LCD display message

On completing the sanitization, the audience enters a temperature check-up. The normal temperature range (37°C) is set in the microcontroller. It's shown in Figure 6, that prototype work, temperature is measured using LM35 sensor with a heated object and if it exceeds the normal temperature range, the microcontroller displays an output with a warning "!!!TEMP_HIGH!" and people won't be allowed inside. If the temperature measured is below the normal range, access will be given to enter inside, and the LCD displays the temperature measured and "PLS^ENTER_INSIDE" message is shown in Figure 7. On displaying this message, a dc motor will run indicating the opening and closing of doors for that person. The time taken for measuring this temperature using LM35 sensor is less than 5-7 seconds.
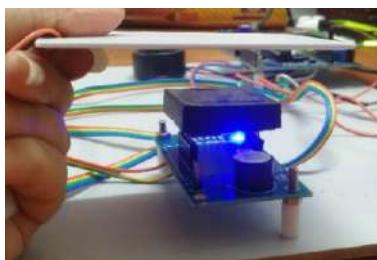
Figure 6. RFID tag above the reader

Figure 7. LCD display on obstacle
Over ultrasonic sensor

Figure 8. Data of individuals uploaded in the IoT server

All the individuals are allotted with alternate seat numbers so as to follow the basic social distancing inside the sports ground. If by mistake anyone tries to take over the non-allotted seats, a buzzer sound will arise, and the LCD displays a message saying "DON'T_SIT_HERE" is shown in Figure 8.

4. Conclusion

This work explores an automated check-up system in crowded areas, here in the sports-stadium. By implementing this in real life, it promotes a safer system than that involving manual check-up. It provides preliminary precaution support for the audience in the stadium by sending out audience with high temperature. The system also maintains the primary data of the audience in the IoT server and can be verified at any time. It makes them follow proper social distancing with the help of ultrasonic sensors. Thus, an automated electronic system for the prevention of virus spread while watching sports is successfully achieved. This work can be implemented in any enclosed crowded areas which also involve verification of each individual like theatres, schools etc. some additional data like entry and exit time, and measured temperature can also be made to store in the IoT server.

References

[1] Hammond, B., Ali, Y., Fendler, E., Dolan, M., & Donovan, S. (2000). Effect of hand sanitizer use on elementary school absenteeism. American Journal of Infection Control, 28(5), 340–346. https://doi.org/10.1067/mic.2000.107276

[2]   Baslyman, M., Rezaee, R., Amyot, D., Mouttham, A., Chreyh, R., & Geiger, G. (2014). Towards an RTLS-based Hand Hygiene Notification System. Procedia Computer Science, 37, 261–265. https://doi.org/10.1016/j.procs.2014.08.039

[3]   Chen, H.-Y., Chen, A., & Chen, C. (2020). Investigation of the Impact of Infrared Sensors on Core Body Temperature Monitoring by Comparing Measurement Sites. Sensors, 20(10), 2885. https://doi.org/10.3390/s20102885

[4]   Foisal Mahedi Hasan, M., Tangim, G., Kafiul Islam, M., Rezwanul Haque Khandokar, M., & Ul Alam, A. (2010). RFID-based ticketing for public transport system: Perspective megacity Dhaka. 2010 3rd International Conference on Computer Science and Information Technology, 459–462. https://doi.org/10.1109/ICCSIT.2010.5564067

[5]   Davis, M. A., Sheng, H., Newman, J., Hancock, D. D., & Hovde, C. J. (2006). Comparison of a waterless hand-hygiene preparation and soap-and-water hand washing to reduce coliforms on hands in animal exhibit settings. Epidemiology and Infection, 134(5), 1024–1028. https://doi.org/10.1017/S095026880600598X

[6]   Srihari, M. M. (2020). Self-Activating Sanitizer With Battery Imposed System For Cleansing Hands. 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 1102–1105. https://doi.org/10.1109/ICIRCA48905.2020.9183347

[7]   Long, G. (2016). DESIGN OF A NON-CONTACT INFRARED THERMOMETER. International Journal on Smart Sensing and Intelligent Systems, 9(2), 1110–1129. https://doi.org/10.21307/ijssis-2017-910

Solar-Powered Improved Full Bridge Resonant Inverter for High-Frequency Industrial Applications

Upasana Sapra [a,1], Aiswarya Sakthivel [b], X.Shobana Macdalin [b], S.Siva Ranjini [b]
[a] *Assistant Professor, SRM TRP Engineering College, Trichy, TN, India*
[b] *UG Student, SRM TRP Engineering College, Trichy, TN, India*

Abstract. Electrical energy consumption is alarmingly rising, but the availability of conventional sources is limited. To meet the increasing demand; the implementation of non-conventional sources is the need of the hour. Solar energy is the most sustainable alternative for power generation among non-conventional sources families. Resonant inverters are used in low-power high-frequency induction heating appliances. Full-bridge resonant inverters are most commonly used to convert solar received power into the suitable form required for high-frequency application device by providing maximum power to the load at resonant frequency The aim of the paper is to analyze the working of the resonant inverter by taking the input supply from the solar panel and converting the obtained dc input to ac input through the resonant inverter. This obtained output is supplied for the high-frequency industrial application which mainly includes Induction frequency heating Applications. Induction heating is one of the techniques used in casting foundry for the treatment of metals. It involves the heat treatment of the metals namely annealing, hardening tempering method. goes here.

Keywords. Solar Energy, Induction heating, resonant inverters

1. Introduction

Solar energy is a renewable energy resource where it takes the energy from the sun and is one of the powerful energy resources and also it is a truly renewable energy resource because it is available in all the areas of the world and it is available entirely for at most several billion years as according to the scientist's research and also we cannot run out of the solar energy. Solar energy also has many advantages where it is available for nearly half of the day and also during night time the energy can be stored and used accordingly, also it has an advantage of minimum maintenance costs apart from the initial costs for the setup of the solar plant. It also has the advantage of the usage of solar energy for diverse applications.

This paper is developed to improve the efficiency as well as to reduce the losses that is produced in the switching off and on of the semiconductor switching devices which in turn reduces the power losses in the circuit [1-2].

---

[1]Upasana Sapra, Assistant Professor, SRM TRP Engineering College, Trichy, TN, India; Email: upasana.s@trp.srmtrichy.edu.in.

Use of the conventional Pulse Width Modulation techniques in the inverters provides distorted current and voltage but the cost of higher switching losses occurs due to high switching frequencies. The efficiency of the inverter depends mainly on the power losses which is a sum of the conduction losses and switching losses where in the switching losses are mainly dependent on Supply voltage, Load current, Operating frequency, Rising and falling time of the switch voltage and current [2-3].

The switching losses mainly depend on the transistor current and voltage during the turn on and off of the process. The switching losses occur due to hard switching conditions which mean they are switched at either when the current or voltage is high. The switching losses were tested in a normal inverter has been calculated in the range of 500W in a normal VSI inverter [4].

The resonant inverter is mainly used for several applications as compared to the normal inverter because of the soft-switching characteristics and it has the advantage of reduced switching losses, small size, and less weight of the passive components [5], which are obtained by the ZVS and ZCS. The resonant inverter mainly consists the semiconductor switching devices namely MOSFET, IGBT, etc., and includes the resonant tank circuit which includes the LC components connected either in series or parallel connection depending on the type of the application. The frequency at which the resonance condition occurs for the resonant tank circuit will be taken as the switching frequency for the semiconductor switching devices in the resonant inverter. The diagram of the resonant inverter is shown in the Figure 1.

Figure 1. Block diagram of resonant Inverter

There are many types of resonant inverters in which series resonant converter has an advantage of good efficiency but a disadvantage of output voltage regulation in light load conditions, similarly parallel resonant inverter has good output voltage regulation but the efficiency is a bigger drawback. Among the hybrid converters connections namely LCC, LLC and CLL wherein LCC and CLL consists of ripples in the output voltage, when convertors are needed for high voltage applications the output voltage should be free from ripples [5]. The LLC topology is preferred because it is free of ripples and the efficiency obtained is higher compared to LCC and CLL. LLC resonant converter gains attention high value because it has an ability to operate at higher frequencies.

Induction heating is a method used for heating the metals through the electromagnetic induction method where the heat is generated in the metal through the eddy currents. The induction heating application consists of an electromagnet and electronic oscillator [6]. The continually varying magnetic field passes through the object which in turn generates the current inside the metals called the eddy currents [7].

2. Methodology

The input for resonant inverter is the solar energy which is obtained from the solar panel and it is DC voltage and this DC voltage may contain some harmonics and this DC voltage can be filtered using passive filtering components namely LC filter and using this filtering method the pure DC voltage is obtained without any harmonics. The Block diagram of Solar Powered Improved Full Bridge Resonant inverter is in Figure 2.

This DC voltage is then passed to the resonant inverter consisting of four MOSFET semiconductor switching devices and also a resonant tank circuit consisting of an LC resonant network where the frequency of resonance is taken as the frequency of operation of the semiconducting switching devices. The principle of operation in the resonant inverter is that the frequency of operation of semiconductor switching devices will be the same as that of the resonant frequency of the tank circuit consisting of an LC network.
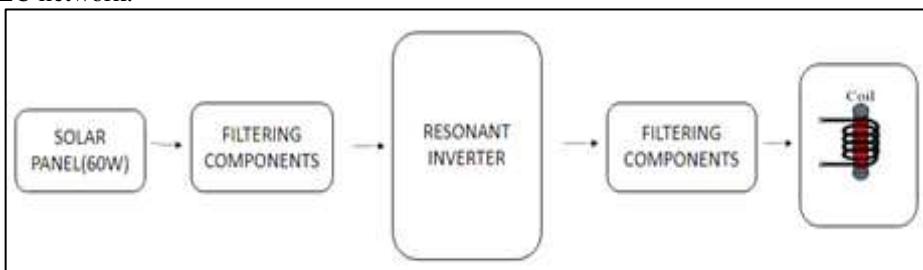
Figure 2. Block Diagram of proposed model

Let the resonant angular frequency be ω_o

$$\omega_o = 2\pi f_o,$$

Where f_o is the frequency of resonance.
Let the angular frequency of operation of semiconductor switching devices be ω.

$$\omega = \omega_n / \omega$$

Where ω_n- normalized operational frequency

$$\omega = 2\pi f$$

where f-switching frequency.

The output from the resonant inverter is AC which consists of several harmonics and it will not be a pure AC and this AC is then purified using passive filtering components mainly LC filter and the pure AC is obtained without any harmonics. The AC which is to be given load is not of the required voltage and then the output is passed to the high-frequency transformer and the required output that is to be given to the load is obtained and then the voltage is given to the load which would be the Induction Heating application which can be represented as a series combination of inductance and resistance which together is represented as induction coil for the induction heating application.

3. Simulation Diagram

The simulation diagram of the Solar Powered Improved Full Bridge Resonant inverter for High Frequency Industrial application is shown in the Figure 3. The simulation was carried out in MATLAB R2018a version.

The simulation diagram consists of the following parts mainly solar panel then filtering components which are connected to obtain the DC output. The DC voltage is given to the resonant inverter wherein it consists of four semiconducting switching devices and the resonant tank circuit and the load is connected which is mainly the inductive heating application which is represented as a series combination of inductance and resistance.

Table 1. Simulation Results

S.No.	Component	Value
1.	Solar panel	60W
2.	Inductor	42micro henry
3.	Capacitor	0.35micro farad
4.	Filter Capacitor	5.5micro farad
5.	Load resistor	190kohm
6.	Output Frequency	200Hz
7.	Output Voltage	48V

The simulation diagram results consists of input voltage i.e., the voltage from the dc voltage and the voltage output from the linear transformer and the output voltage and output current from the load and the y axis is taken as time in seconds. Table 1. Shows the ratings of components values used in simulation.
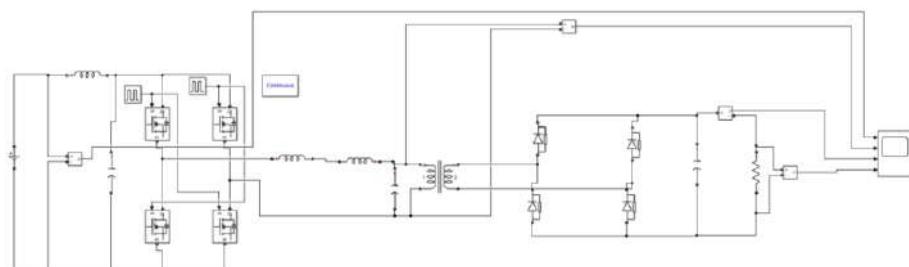
Figure 3. Simulation diagram

The results are obtained and the results of the simulation are shown in the Figure 4.
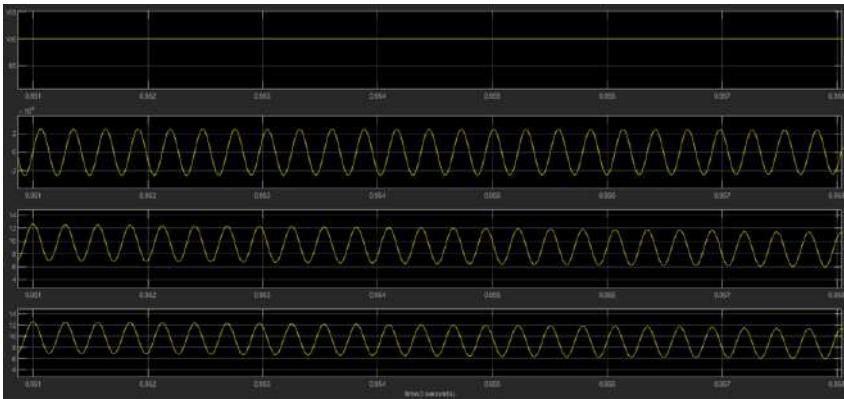
Figure 4. Simulation diagram results

4. Conclusion

The simulation output represents the input from the solar panel and the output current and voltage waveform. The results obtained from the simulation diagram in the MATLAB Simulink shows that the switching losses are less as compared to the usage of the normal inverter. At this value of input, the corresponding output has less switching losses due to the resonant phenomenon. The results also confirm that the ac wave obtained as an output and low harmonics content. The switching losses are reduced and the high frequency obtained so can be used for industrial heating purposes.

References

[1]   Sayed, K., Abo-Elyousr, F., Abdelbar, F. N., & El-zohri, H. (2018). Development of Series Resonant Inverters for Induction Heating Applications. European Journal of Engineering Research and Science, 3(12), 36–39. https://doi.org/10.24018/ejers.2018.3.12.975

[2]   Jin, J., Kim, M., Han, J., Kang, K., & Jung, J.-H. (2020). Input voltage selection method of half-bridge series resonant inverters for all-metal induction heating applications using high turn-numbered coils. Journal of Power Electronics, 20(6), 1629–1637. https://doi.org/10.1007/s43236-020-00147-y

[3]   Varecha, P., Makyš, P., Sumega, M., & Sovička, P. (2019). Power losses analysis in MOSFET 3-phase high current power inverter for automotive application area. Transportation Research Procedia, 40, 571–578. https://doi.org/10.1016/j.trpro.2019.07.082

[4]   C, B., & Rajesh Babu, R. S. (2018). Analysis of High Voltage High Power Resonant Converters. International Journal of Power Electronics and Drive Systems (IJPEDS), 9(1), 174. https://doi.org/10.11591/ijpeds.v9.i1.pp174-179

[5]   Mao, S., Popovic, J., Ramabhadran, R., & Ferreira, J. A. (2015). Comparative study of half-bridge LCC and LLC resonant DC-DC converters for ultra-wide output power range applications. 2015 17th European Conference on Power Electronics and Applications (EPE'15 ECCE-Europe), 1–10. https://doi.org/10.1109/EPE.2015.7311786

[6]   Haema, J., & Phudungthin, R. (2019). Full Bridge Resonant Inverter for Blade Induction Heating Application. Energy Procedia, 156, 391–395. https://doi.org/10.1016/j.egypro.2018.11.106

[7]   Khatua, M., Kumar, A., Yousefzadeh, V., Sepahvand, A., Doshi, M., Maksimovic, D., & Afridi, K. K. (2018). High-Performance Megahertz-Frequency Resonant DC-DC Converter for Automotive LED Driver Applications. 2018 IEEE Energy Conversion Congress and Exposition (ECCE), 2186–2192. https://doi.org/10.1109/ECCE.2018.8558088

Wireless Underground Sensor Network for Monitoring Various Fields Using Magnetic Induction

P. Manigandan [a,1], C. Balaji [b], M. Ramanan [b], S. Ragul [b]

[a]Assistant Professor, Department of EEE, SRM TRP Engineering College, Trichy, TN, India

[b]UG Student, Department of EEE, SRM TRP Engineering College, Trichy, TN, India

Abstract. The frequent accidents of mining safety caused severe losses and massive cost losses. The global mining sector urgently needs to improve operational efficiency and the overall safety of mines. This article suggests a WUSN based mining safety monitoring. The monitoring system collects temperature, moisture, soil vibration and gas values around the mine, and then transmits the data through wireless underground sensor network. Here Transmission based on magnetic induction (MI) is being suggested, in this approach soil is medium for communication so based on the soil conductivity the measured information are transferred. It is also an early warning system, which will help all miners in the mine to save their lives before a victim happens.

Keywords: Magnetic Induction, WSUN

1. Introduction

India is a big nation with abundant coal reserves. However, the actual healthy coal mine output level remains low, and disasters in coal mines have become more common in recent years, resulting in significant loss of property and life. Coal mine protection issues have increasingly risen to the forefront of the nation's and society's concerns. Because of the complexities of the mining setting and the wide range of coal mine job conditions [1], it is important to keep a close eye on the mine operating environment. Monitoring systems for traditional coal mines are typically wired systems for the network that take a significant role in ensuring mine of coal safety. As the mining areas and the depth of coal mines are expanding continuously, a lot of paths have become blind and many concealed hazards. Furthermore, laying wires, which is costly and time- consuming in the coal mines.

A coal mine Safety Monitoring System will be designed on the basis of an underground Wireless Sensor Network (WUSN), to improve production safety monitoring and reduce coal mining casualties [2]. Wireless sensor networks consist of a large number of small volume and low cost micro sensor nodes. It has self-organized wireless communication capabilities.

---

[1]P.Manigandan, Assistant Professor, Department of EEE, SRM TRP Engineering College; Trichy, TN, India; E-mail: manieee.recc24@gmail.com.

Underground mines are typically large labyrinths with long, narrow tunnels measuring a length of kilometers and width of several meters. Thousands of miners are required to operate under extreme conditions, and hundreds of miners die each year from mining accidents. Underground mining activities are now generally accepted as posing a high danger. In order to ensure mining safety and to manage various tasks, a monitoring and controlling system must be installed as a critical infrastructure. Underground coal mining often consists of random corridors and branch tunnels which make it extremely difficult to deploy a network skeleton.

In such cases it may take particular advantages for the automation of underground monitoring and control thanks to quick and flexible use of the wireless underground sensor network (WUSN) and any other sensing devices. Moreover, the multi-hop method of transmission can be adapted well to the tunnel structure and can thus provide a sufficient scalability to construct a mining system and is very suitable for extensive monitoring and control in the coal mineral industry. Mine detection methods have become more sophisticated as technology has advanced, but accidents in underground coal mines continue to occur. Coal mine Safety has increasingly become a major source of concern for society and the nation. The harsh climate and variable working conditions in coal mines are the primary causes of disasters. As a result, mine control systems must be implemented for the sake of protection. Traditional coal mining used to be a trend for wired network networks, which have played a major role in ensuring secure output in coal mines [3]. The continuous extension and deep expansion of the operating areas have turned into blind areas where numerous unnoticed hazards are concealed. Furthermore, costly cables cannot be laid there, that also takes time. Consequently, to ensure a good performance while protecting the environment, it is essential to have a wireless sensor network mine monitoring system installed in these mines.

2.   Related Work

For transmitting the calculated parameters, the current device uses a zigbee network. In Figure 1. Show this system the measures the parameters like temperature sensor and moisture sensor the measured sensor details are transferred based on zigbee network [4] . The received details are displayed on LCD. Traditional wireless signal transmission technology using EM waves can be used in many realistic cases only at very small distances due to high trajectory loss and vulnerability to soil moisture changes.
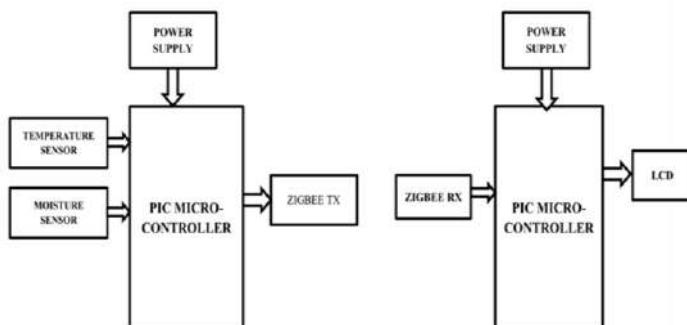
Figure 1. Related Work block diagram

In wireless signal propagation techniques with electromagnetic (EM) waves, only very short transmission can be implemented.

3. Methodology System

Continuous monitoring is needed, which necessitates the use of a reliable and precise sensor system. Various techniques are employed for the detection of these toxic gases, including the use of a gas sensor type semiconductor. These sensors can be used in coal mines, but they can also cause problems in the mining process. The sensor system was often damaged by accident. Another method is to employ a robot. These robots are effective, but they come at a high price. However, there is another way to obtain an efficient and low-cost sensor implantation solution [5] ; the prediction of explosion with the use of sensors and microcontrollers is the secret to manage coal mining accidents, as well as the generation of an alarm device before critical atmospheric levels are reached [6] in shown in Figure 2.The secret to preventing coal mine accidents is to anticipate outbursts using sensors and microcontrollers, as well as to set up a warning system before critical atmospheric levels are reached. Continuous monitoring is needed, which necessitates the use of a reliable and accurate sensing device. Proposed system consists of temperature sensor, Pressure sensor, vibration sensor and gas sensor for monitoring the parameter like temperature, Pressure condition, vibration level and gas. The measured parameters are transferred WUSN. At receiver side, it receives the measured parameters display on LCD and Buzzer is used for the purpose of emergency indication.
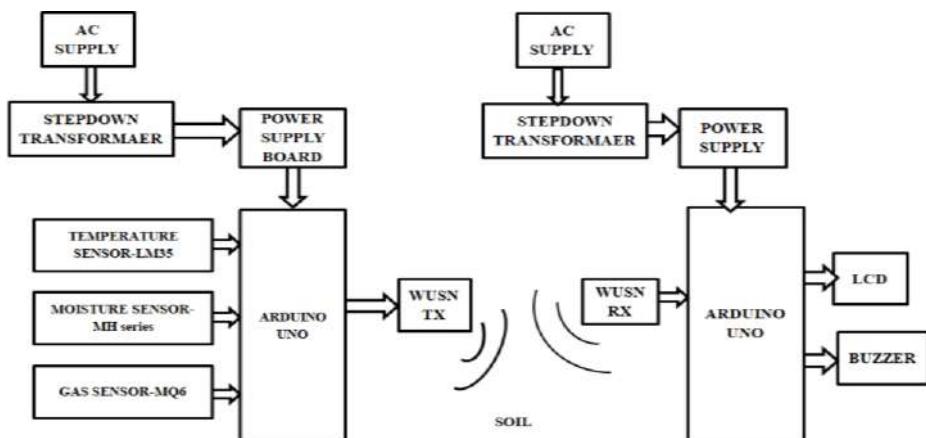
Figure 2. Proposed block diagram

In WUSN transmitter side temperature sensor, moisture sensor and gas sensor are interfaced with the Arduino Uno micro controller [7-9] . It picks up the sensor data that has been tested. Based on a magnetic induction approach, the measured sensor information were transmitted using the WSUN transmitter module. At receiver side the WUSN receiver module receives the information and it displayed on the LCD. If the measured sensor details goes high means buzzeris used to indicate It is highly efficient because it decreases the signal propagation vulnerability in properties of the soil, reduces the track loss and is easy to use Easy to install.

4.   WUNS Module

Sensor nodes are linked to the wireless underground sensor network, and are randomly distributed in the environment to monitor personnel locations or other interesting events. Here sensor nodes randomly distributed to occupy all corners of the environment as it is applied to grid based topology. For the network under consideration (underground sensor network), it is important to know the distance between pairs of nodes as it will allow us to give priority to either distance or energy consumed by nodes in the network [10] . WUSN devices are widely implemented under the ground and no wired connections are required. Each unit contains all sensors, memory, a processor, a radio, a power source and an antenna. Wireless underground sensor networks (WUSN) are made up of wirelessly linked underground sensor nodes that communicate with each other via the soil, and they have the potential to affect a wide range of new applications.

In this basic WUSN circuit as shown in Figure 3, the resistor R1 and the c1 form one adjustable arm of a simple resistance bridge network. While the two fixed resistors R2 and R3 form the other arm. Both these are connected to amplifier IC. The end of R3 connected to R4 which is connected to com port of controller.
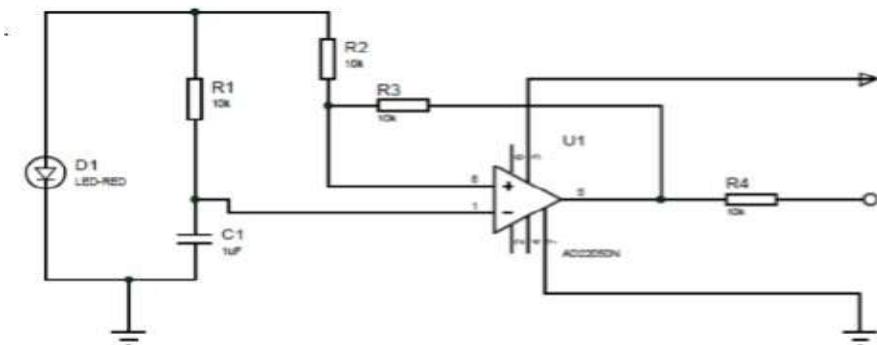
Figure 3. Circuit diagram of WUSN module.

5.   Proteus ISIS7 Simulator

As a fully functional procedural programming language, Simone Zanella developed Proteus (TEXT Easy to USE Processor) in 1998. Proteus integrates many functions derived from a variety of other languages, including C, BASIC, Assembly, and Clipper/dBase; it is particularly flexible when working with strings, with hundreds of dedicated functions, making it one of the most powerful text manipulation languages is shown in Figure 3. Proteus is named after Proteus, a Greek sea god who looked after Neptune's entourage and provided responses; he was known for his ability to turn into various forms [11] . This language's main aim is to translate data from one type to another. Proteus began as a multiplatform (DOS, Windows, UNIX) system utility for manipulating text and binary files as well as writing CGI scripts. In the Figure 4 the language was later extended to include hundreds of specialist network and serial communication functions, database interrogation, system service design, console applications, keyboard emulation, and ISAPI scripting (for IIS). Although a Linux version is still available, most of these features can only be found in the Windows

interpreter version. Proteus was created to be useful (simple to use, effective, and comprehensive), readable, and reliable.

6. Prototype

<div align="center">(a)   (b)</div>
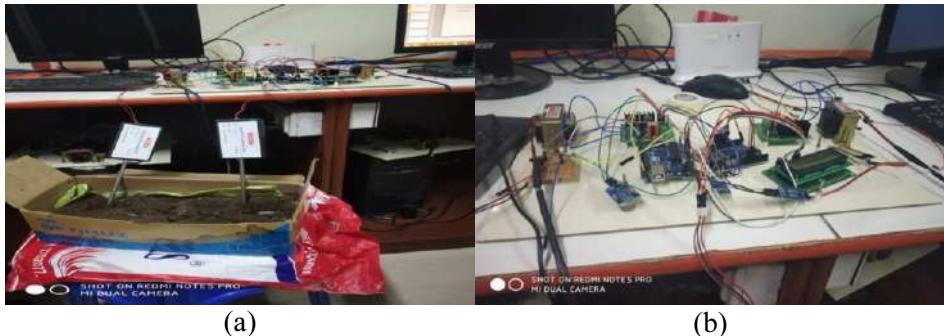
Figure 4. (a) Soil Testing system (**b**), Controlling and monitoring Hub

7. Conclusion

The coalmines safety monitoring system is therefore implemented through the use of the wireless network of underground sensors (WUSN). The advances are linked to different aspects of wireless communication and networking based on the magnetic induction approach. Moisture, Gas and temperature level is always sensed and if changes in those values from threshold level then it monitored by the monitoring center. From this proposed we can able to give immediate treatment for workers in case of emergency situation.

References

[1] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: a survey. Computer Networks, 38(4), 393–422. https://doi.org/10.1016/S1389-1286(01)00302-4
[2] Matin, M. A., & Islam, M. M. (2012). Overview of Wireless Sensor Network. In Wireless Sensor Networks - Technology and Protocols. InTech. https://doi.org/10.5772/49376
[3] Akyildiz, I. F., & Stuntebeck, E. P. (2006). Wireless underground sensor networks: Research challenges. Ad Hoc Networks, 4(6), 669–686. https://doi.org/10.1016/j.adhoc.2006.04.003
[4] Vuran, M. C., & Silva, A. R. (2010). Communication Through Soil in Wireless Underground Sensor Networks – Theory and Practice (pp. 309–347). https://doi.org/10.1007/978-3-642-01341-6_12
[5] Bansal, R. (2004). Near-field magnetic communication. IEEE Antennas and Propagation Magazine, 46(2), 114–115. https://doi.org/10.1109/MAP.2004.1305555
[6] Karalis, A., Joannopoulos, J. D., & Soljačić, M. (2008). Efficient wireless non-radiative mid-range energy transfer. Annals of Physics, 323(1), 34–48. https://doi.org/10.1016/j.aop.2007.04.017
[7] Zhi Sun, & Akyildiz, I. F. (2010). Magnetic Induction Communications for Wireless Underground Sensor Networks. IEEE Transactions on Antennas and Propagation, 58(7), 2426–2435. https://doi.org/10.1109/TAP.2010.2048858
[8] Sheth, A., Toyama, K., Tejaswi, K., Mehta, P., Parekh, C., Bansal, R., Merchant, S., Singh, T., Desai, U. B., & Thekkath, C. A. (2005). SenSlide. Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems - SenSys '05, 280. https://doi.org/10.1145/1098918.1098954
[9] Werner-Allen, G., Lorincz, K., Ruiz, M., Marcillo, O., Johnson, J., Lees, J., & Welsh, M. (2006). Deploying a wireless sensor network on an active volcano. IEEE Internet Computing, 10(2), 18–25. https://doi.org/10.1109/MIC.2006.26
[10] Akyildiz, I. F., & Vuran, M. C. (2010). Wireless Sensor Networks. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470515181

Issues of COVID 19 Screening with Machine Learning Algorithm and Data Sets Availability

G.N. Balaji [a,1], S.V. Suryanarayana [b], P. Vijayaragavan [c]
[a]Associate Professor, SRM TRP Engineering College, Trichy
[b]Associate Professor, CVR College of Engineering, Hyderabad
[c]Assistant Professor, SRM TRP Engineering College, Trichy

Abstract. There is a need to wear a mask during the coronavirus outbreak to efficiently deter the transmission of COVID-19 virus. In these instances, traditional facial screening technologies obsolete for monitoring of group entry at Airports, shopping malls, railway stations, etc. It is, therefore, vital to boost the efficiency of screening. This paper addresses the machine learning algorithm for contactless face screening systems in group participation, social interaction, school management, mall entry management, and market resumption scenarios in the case of COVID- 19. A method to screen entry with masks are developed using machine learning, which depends on various face specimens that were discussed here. The second fold discussion in this paper is that previously there are not many freely accessible masked face-databases. To this end, various forms of masked face data sets are identified, namely MFDD, Real MFRD, and Simulated MFRD. Such data sets became widely accessible to businesses and academics, based on which specific apps may be built on masked faces. The mathematical model, with the code was given. The availability and issues of the above data sets were discussed for the benefit of researchers**.**

Keywords: Machine Learning, COVID-19, Face Screening.

1. Introduction

Throughout the COVID-19 coronavirus outbreak, everyone carries a mask. Facial recognition technologies, majority critical means of recognition, have almost collapsed, causing an immense dilemma for authentication applications that depend on facial recognition, including crowd exit and entry, face authentication, face recognition-based entry at public stations, authentication based on face dependent remote payment and facial attendance. In specific, in the case of social safety checks such as train stations, gates focused on standard facial recognition technologies cannot efficiently identify masked faces; however, eliminating masks for passage identification would raise the possibility of virus contamination. Since the COVID-19 virus may be transmitted by touch, opening essentially based on authentication or fingerprints are not secure.

---

[1] G.N. Balaji, Associate Professor, SRM TRP Engineering College, Trichy, TN, India.
Email: balaji.gnb@gmail.com.

In order to overcome the above-listed challenges, it is important to enhance current facial recognition methods that rely heavily on all facial attributes, ensuring that identification authentication can be carried out efficiently in the case of inadequately visible faces. Here the data sets for doing research were addressed so that more algorithms can be proposed. The futuristic face detectors were fixed basis of machine learning, which depends on training a large dataset [1]-[5]. Hence, designing masked facial-recognition system includes a considerable of masked facial samples. Under the current, accessible masked face dataset is not available, and this study aims to address such datasets through a number of methods for the benefit of researchers. Face recognition accuracy improves as more facial features become usable as recognition data points. Face occlusions like these are common in real-world use cases, such as when someone is wearing sunglasses, a scarf, a low hat, or when their face isn't turned directly toward the camera. When any part of the face is hidden, a facial recognition algorithm must rely on whatever face landmarks it can see to make an accurate match.

It's a process that takes time to master, but a resilient algorithm can adjust when it sees a partially occluded face and still produce excellent results. Organizations may wish to be alerted when individuals are not wearing a mask, as face masks have become an effective tool in the fight against COVID-19. This kind of logic isn't found in any recognition systems. We can program our facial recognition system to give the workflows when masks are, or are not present.

2.   Identification of Datasets for Future Research

With reference to the latest common face masks, there have been two strongly linked and separate applications, respectively facial mask identification and masked face recognition tasks. The role of detecting a face mask must be to decide if an individual wears a mask as needed. Masked face recognition function requires to recognize the real identification of a masked individual. A growing function has specific data set specifications. The earlier only wants blurred face picture examples, whereas the other includes a dataset including several face images of that same subject with and without a mask. Reasonably, the databases used by the facial recognition function are much more challenging to create. In this respect, this paper identifies two tasks:

First task is to identify datasets for researchers and second one is to identify an algorithm for screening while entering into a mass gathering.

Our First task is to attempt to identify masked face detection and recognition databases; this paper addresses various kinds of masked face databases, comprising Masked Face Detection Dataset (MFDD), Simulated Masked Face Recognition Dataset (SMFRD) and Real-world Masked Face Recognition Dataset (RMFRD) [11]. The description of the MFDD, the RMFRD, and the SMFRD are listed below.

MFDD: The origins of the MFDD primarily consist of two parts: (a) Most of the samples come from similar research [6]; this dataset contains masked faces of 24,771 in count. In particular, to decide if an individual wears a mask since it is unlawful to wear a mask during most of the coronavirus epidemic.

MFRD: A python crawler device is used to search the front-facing photographs of prominent Figures and their accompanying veiled facial pictures from vast Internet infrastructure. Then, we automatically delete the incorrect representation of the face from the wrong communication. The method of filtering photographs requires a lot of manpower. In the same manner, we use semi-automatic label software, such as

LabelImg and LabelMe [7], to construct precise face regions. The collection contains 5,000 pictures of 525 individuals with masks and 90,000 photographs out of which 525 are without mask. To our understanding, this is the largest masked face data collection in the modern world reveals the facial expression pairs.

SMFRD: In the meantime, in order to develop the quantity and variety of masked facial detection databases, additional approaches have been used to place masks on current public large-scale facial databases. In order to show the utility of data processing, we have created a mask with Dliblibrary [8] based tools for automated mask- wearing. LFW [9] and Web face datasets are common face recognition datasets, including In addition, they created a masked face by virtual. In reality, artificial masked face datasets may be used alone. In reality, virtual masked face databases may be used together with unmasked initial equivalents reveals a collection of virtual facial pictures.

2.1. Machine Learning Algorithms:

Our second task is to identify a machine learning algorithm to screen or identify the persons who are with masks protection and who are not following the rules. So that while in mob gathering like shopping malls, airport entry, cine complex entry screening, we can identify with the system and warn them to leave from the entry and have the protection of masks and enter again.

3. Inception v3 model for identifying masked faces

In this method, we use the Inception v3 model to identify the pictures as masked and unmasked. Inception v3 is a machine learning image recognition algorithm which has demonstrated better than 78.1 percent precision in the ImageNet dataset. This has 42 layers and less constraints than other similar models, including AlexNet, VGG, and Inception v1. The concept is the result of several theories generated through the ages by several researchers. It's was built by Szegedy, et. al. This method can be adopted for identifying the people with masks and without masks so that the model can screen them at starting point and avoid the non-mask persons, to address the downside in Inception v2, which was that the auxiliary classifiers did not make a substantial impact until the end of the training cycle, when the precision was approaching saturation they often act as regularizes, particularly if they have Batch Norm or Dropout operations. The model is trained with MAFA dataset. The Figure. 1 shows brief glimpse of the Dataset and Figure. 2 show the Input to the Inception algorithm.

The layout itself comprises of symmetrical and including convolutions, asymmetrical building blocks, max pooling, concats, average pooling, dropouts, and completely linked layers. BatchNorm is commonly utilized in the software, and it is used to trigger inputs. Loss is being calculated through Softmax. The maximizing RMSProp is used together with the decay $\alpha= 0.9$, the intensity $\beta= 0.9$, and the $\epsilon= 1.0$. The following is the input and output using this method to screen the people while COVID 19 disaster. The output Figure. 3 shows the non- mask lady with "N," and its prediction took over 0.052 seconds.

```
# compilngthe classifier
classifier.compile(loss='binary_crossentropy',metrics=['accuracy'],
optimizer='RMSProp')
```

3.1. Results of screening test Conducted

Figure 1. Glimpse of Input Data Set

EXPLORING THE DATA

```
#EXPLORING THE DATA
sample_img_path = os.path.join(DIR_IMG_TRAIN,'M_1.jpg')
sample_img = plt.imread(sample_img_path)
plt.imshow(sample_img)
print ('Image have shape: {}'.format(sample_img.shape))
```

Image have shape: (91, 91, 3)

Figure 2. Input to Inception V3

```
img_path = DIR_IMG_TEST+'N_776.jpg'
display_img = image.load_img(img_path)
start = time.time()
score = predictFromPath (img_path)
end = time.time()
print("Prediction took {:.3f} seconds".format (end - start))
print("It's a {}! (with a score of {}) 0 -> Non Mask / 1 -> M
e))
plt.imshow(display_img)
```

Prediction took 0.052 seconds
It's a N! (with a score of 0.07839389145374298) 0 -> Non Mask
<matplotlib.image.AxesImage at 0x7f1bee9e10b8>
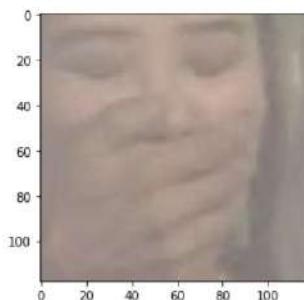
Figure 3. Inception V3 output 1

4. Identification of Issues

Face-based recognition can be loosely split into two design scenarios: unregulated and managed design conditions. The previous applies primarily to public video monitoring cases where the angle between the camera and the image, the posture, the occlusion, and the illumination are all unknown. For such situations, the quality of facial recognition is fairly poor. For a fact, the precision of using a face mask would be greatly decreased. Though, there are still a wide range of monitored implementation situations, such as compliance tests at job places, health screening at rail stations and facial recognition fees, etc.

Therefore, big-quality frontal facial photographs are quickly obtained, such that the job of masked facial identification is no longer such challenging. Of example, the idea is to eliminate mask intrusion and to assign greater emphasis to the valuable features of the uncovered face. The above issues should be consider while doing further research.

5. Application Status and accuracy

Perhaps owing to the abrupt outbreak of the COVID-19 virus, there are currently limited organizations that, implement facial recognition technologies to individuals wearing masks. Based on this study, the authors in registered 85 percent accuracy while individual's nose is exposed to 50 percent. Approximately 85 percent reliability of masked facial recognition is given by Hanvon Technology. The best result recorded over 90% is from MINI VISION Technology.

6. Conclusion

In the scenario of the COVID 19 disaster outbreak, we have identified an algorithm for screening people to enter into the mass gathering with masks and refuse the non-mask persons using deep learning inception v3 model. We have also addressed the available data sets for facial recognition so that more algorithms can be proposed for identifying or recognizing people even they wear masks for easy passage into shopping malls, airports, railway stations, cine complex, etc. According to the World Health Organization, two-thirds of COVID-19 patients have a dry cough rather than the usual wet cough associated with the common cold or allergies. Cough sounds have also been used to diagnose whooping cough, asthma, and pneumonia. As a future work we can extend this work to new diagnostic tests using machine learning algorithms "to determine what kind of frequencies the cough is made up of".

References

[1] Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive Angular Margin Loss for Deep Face Recognition. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4685–4694. https://doi.org/10.1109/CVPR.2019.00482

[2] Liu, B., Deng, W., Zhong, Y., Wang, M., Hu, J., Tao, X., & Huang, Y. (2019). Fair Loss: Margin-Aware Reinforcement Learning for Deep Face Recognition. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 10051–10060. https://doi.org/10.1109/ICCV.2019.01015

[3] Ul Rahman, J., Chen, Q., & Yang, Z. (2020). Additive Parameter for Deep Face Recognition. Communications in Mathematics and Statistics, 8(2), 203–217. https://doi.org/10.1007/s40304-019-00198-z

[4] Liu, W., Wen, Y., Yu, Z., & Yang, M. (2016). Large-Margin Softmax Loss for Convolutional Neural Networks. http://arxiv.org/abs/1612.02295

[5] Tran, A. T., Hassner, T., Masi, I., & Medioni, G. (2016). Regressing Robust and Discriminative 3D Morphable Models with a very Deep Neural Network. http://arxiv.org/abs/1612.04904

[6] Wang, H., Kang, B., & Kim, D. (2013). PFW: A Face Database in the Wild for Studying Face Identification and Verification in Uncontrolled Environment. 2013 2nd IAPR Asian Conference on Pattern Recognition, 356–360. https://doi.org/10.1109/ACPR.2013.53

[7] Yi, D., Lei, Z., Liao, S., & Li, S. Z. (2014). Learning Face Representation from Scratch. http://arxiv.org/abs/1411.7923

[8] Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., Yi, P., Jiang, K., Wang, N., Pei, Y., Chen, H., Miao, Y., Huang, Z., & Liang, J. (2020). Masked Face Recognition Dataset and Application. http://arxiv.org/abs/2003.09093

[9] Balaji, G. N., Subashini, T. S., Madhavi, P., Bhavani, C. H., & Manikandarajan, A. (2020). Computer-Aided Detection and Diagnosis of Diaphyseal Femur Fracture (pp. 549–559). https://doi.org/10.1007/978-981-13-9282-5_52

Machine Learning Using Big Data Link Stability Based Node Observation for IoT Security

R.Ganesh Babu [a,1] , S.Yuvaraj [a], A.VedanthSrivatson [b], T.Ramachandran [b], G.Vikram [b], N.Niffarudeen [b]

[a]*Associate Professor, Department of Electronics and Communication Engineering*
[a]*Assistant Professor, Department of Electronics and Communication Engineering*
[b]*UG Student, Department of Electronics and Communication Engineering*
SRM TRP Engineering College, Tiruchirappalli, TN, India

Abstract. IoT systems create a multi-hop organizational structure among mobile devices in required to send on data groups. The remarkable properties of gadgets frameworks cause communications to interconnect among competing handheld devices. Most physiological directing displays don't believe secure associations all through bundle communication to organize high communicate ability and genetic blocks that also prompts increased delay as well as bundle decreasing in mastermind. Only with continued growth and transformation of IoT networks, attacks on such IoT systems are increasing at an alarming rate. Our purpose will provide researchers with a research resource on latest research patterns in IoT security. As the primary driver of with us research problem concerning IoT security as well as machine learning. This analysis of the literature among the most research literature in IoT security recognized some very key current research which will generate organizational investigations. Only with fast emergence of different IoT threats, it is essential to develop frameworks that could integrate cutting-edge big data analytics and machine learning advanced technologies. Effectiveness are critical quality variables in shaping the best methods and algorithms for detecting IoT threats in real-time or close to real time.

Keywords. Machine Learning; Physical Routing Protocol; Big Data; Traffic Density, Node Observation; Link Stability, IoT Security.

1. Introduction

Previous experts consider Connectivity which enable unavoidable availability among devices and do not rely on costly framework system. Communications between devices and previous establishment discharge a wide range of skilled applications besides explorers to motorists. The implementations provide security as well as reassurance to drivers by allowing them to collaborate and communicate with each other in order to avoid any mishaps, for example, a road transformed parking garage, innocuous barriers, speed violation, access, temperature records, sight as well as sound congresses, and etc.

---

[1] R.Ganesh Babu, Associate Professor, Department of Electronics and Communication Engineering, SRM TRP Engineering College, Tiruchirappalli, TN, India; Email: ganeshbaburajendran@gmail.com.

Regardless of the fact it is a subclass of compact specially designated systems, a network device does have few distinguishing characteristics that distinguish from other improvised systems [1] . The much more important distinctions are the elevated simplicity schedule; rapidly constantly evolving configuration results in higher framework fragmentation and dysconnectivity in sorting out. Amazing optimization techniques, in any scenario, are also not completely sporadic, and development of sensor hubs is constrained by ways as well as, in the most portions, evident.

The Internet of Things (IoT) is just a network with billions of connected devices which can send or receive information over the network. Nowadays, these systems could be located in a variety of settings, including homes, workplaces, transportation, medicare, telecommunications, agriculture, and so on. IoT systems are rapidly expanding, making a significant difference with us everyday lives and assisting industries such as healthcare as well as transport infrastructure in making decisions. According to the research results of Business Analyst's 2020 IoT document, the IoT size is forecast to grow by more than $2.5 trillion per year by 2028. This contains an increase in the number of IoT gadgets from 9 billion in 2019 to 44 billion besides 2028. IoT has decided to bring substantial benefits to everyone lives, social system, and industrial sectors over its years; that being said, its technology was using has yet to mature sufficiently to provide utilized to carry as well as communication. As the number of devices connected grows, adversaries have more opportunities to gain direct exposure to all of them and are using them to release huge threats.

Securing IoT gadgets is becoming increasingly difficult for both producers and consumers. Several of the significant security challenges recognized mostly by researchers include weak, fallback, or internet data storage without even a passcode. IoT gadgets are frequently shipped to switch, simple, or no passcode. Hackers can very easily exploit such vulnerability by gaining access to these devices. Such vulnerability puts the consumers' privacy at risk and allows hackers to use IoT devices to launch large-scale attacks like DDoS. It confirmed that even an unclassified health history of over 6 million people in the United States, or even huge sizes globally, is accessible online. This data is kept online from over 197 servers, because anyone with an internet browser can availability it by operating effective governance. According the Sophistry document [8], the ten largest credit and quickly guessable passcode being used with IoT gadgets attacks are as follows are shown in Figure 1.

Only with help of machine learning (ML) and thoughtful threats in IoT gadgets, it is essential to identify a defense actions and understand important limitations within the security measures once more for interchange in diverse networks[9]. This procedure is challenging since an IoT device with limited abilities frequently has complexity correctly forecasting the whole infrastructure as well as inflict damage level[10].

2. Machine learning in IoT Security

Throughout this we will discuss the inspiration for using ML throughout IoT security inside the context of existing information security used within IoT networks. We begin by shedding light also on particular qualities of IoT networks and is relevant with security, but instead we start debating the security threats that impede IoT implementation, and also the holes in current security products. After with the intention of establish the incentive for utilize ML to deal with security deal with in IoT gadgets[1].
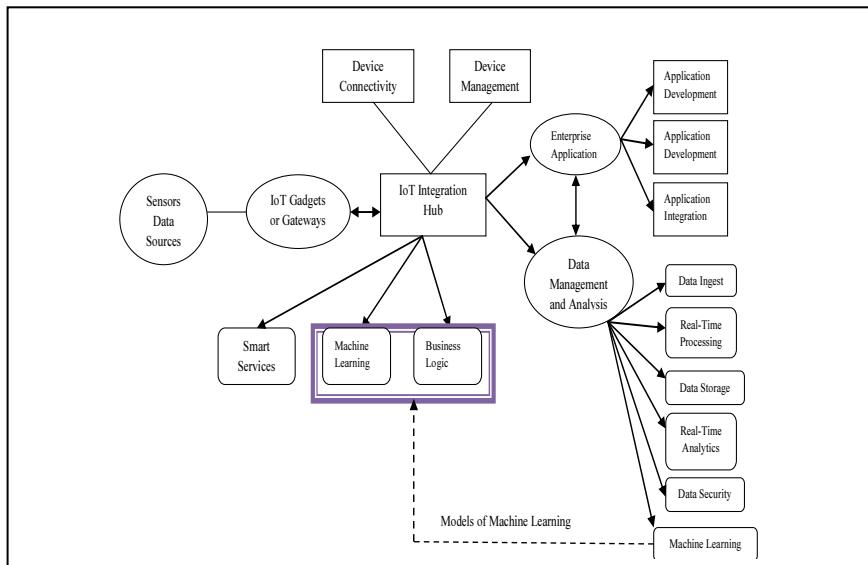
Figure 1. Machine Learning with IoT Security using Big Data Analysis

3. Security Challenges in IoT Deployment

Data protections are two of the most popular considerations with in commercialization of IoT applications and services. The host Server is an enticing playground besides security threats starting from pure bots to organizational excellently data breaches which have harmed multiple sectors such as wellbeing as well as corporate [2]. The restrictions of IoT gadgets, as well as the environment of the organization, present unique challenges for the security of the both gadgets and applications. To date, security requirements in the IoT system are being thoroughly studied from different perspectives, including secure communication, information security, confidentiality, architecture and design security, information security, threat intelligence, and etc [3]. It concentrated on the differences and similarities between IoT as well as conventional IT gadgets in terms of security. They as well concentrated on privacy concerns. The main motivators for arguing about commonalities and contrasts are technology, equipment, networks, and implementations.

4. Machine Learning and IoT Security

Throughout this segment, we will look at machine learning techniques and how they can be used in IoT systems.

$$Dis\tan ce = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

(1)

There are four types of ML algorithms: monitored, unmonitored, semi-supervised, as well as validation classification techniques.

$$M_s = M_1 + M_2$$

(2)

Where

$$M_1 = Radius_1^2 . a_2 - \frac{Radius_1^2 . \sin(2a_1)}{2}$$

(3)

$$B_2 = Radius_1^2 . \alpha_2 - \frac{Radius_1^2 . \sin(2\alpha_1)}{2}$$

(4)

The line connecting the source as well as impartial center points $_{90^0 = (\alpha_1 = 45^0)}$ is a point bisector; throughout this method, the half-region is used M_s .

$$A = Radius_1^2 \left[\frac{\pi - 2}{4} \right] + Raduis_2^2 \left[\alpha_2 - \frac{\sin(2\alpha_1)}{2} \right]$$

(5)

Supervised Classification: It is occurs while potential measures have been delineated to be attained from a number of objects [4]. For such a method of training, its data is classified initially, then trained with the classification model.

$$Dis_{wv} = \max \left[\log \left[\frac{CD_{s,d}}{CD_{j,d}} \right], 0.1 \right]$$

(6)

Unmonitored Learning: It is the climate only significantly contributes without any desired outcomes. It's doesn't require classification model and can analyze similarities among unstructured data as well as classify it into various groups.

$$Dir_{wv} = \left[LQ \left(\overline{D}_n, \overrightarrow{D_{pt}} \right) \right]$$

(7)

Reinforcement Learning (RL): No desired objectives are described in Reinforcement Learning (RL), as well as the representative discovers from responses after observing the environment.

$$TD_{wv} = \left[1 - D_c \right] + \left[\min \left[\frac{N_{avg}}{N_{con}}, 1 \right] \right]$$

(8)

It's also very critical to select the appropriate objective functions as its agent's future direction is determined by the reward systems accumulated [5]. With in opposite situations, RL technologies are frequently used:

- When there is a lack of critical data and previous examples besides training phase.
- The precise correct and incorrect values again for particular circumstance are unknown a priori.

- The ultimate goal is recognized, and the climate could be discerned in order to increase both short-term as well as long-term benefits.

5. Function of Basic ML Techniques

Monitored and unmonitored learning techniques are primarily used to solve predictive analytics problems, whereas RL is used to solve comparative analysis as well as decision-making problems [6]. The nature of the data available influences this classification and the decision of ML Classification algorithms. Classification model is used when its category of input data as well as the target output (labels) are recognized [8] . In this case, the system has only been given training to information of the user to expected output. Learning algorithms are two classifier techniques; to regression dealing with constant outputs as well as classification dealing with separable outputs is show in Figure 2. Support Vector Regression (SVR), regression methods, and memory modules are examples of frequently utilized logistic regression.
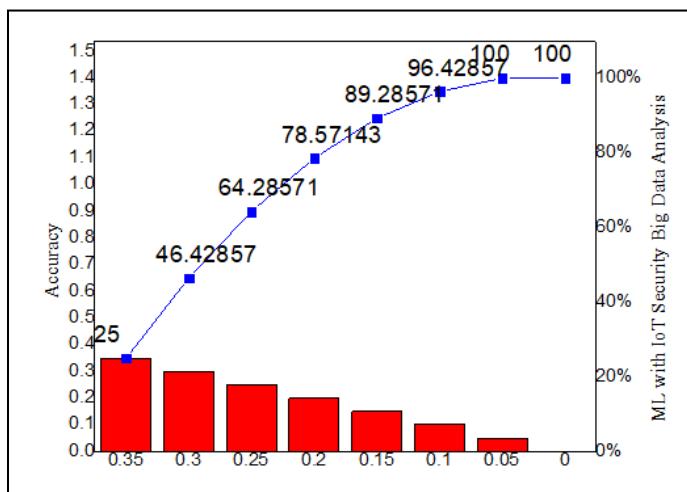
Figure 2. Big Data Analysis for Machine Learning with IoT Security

Classification, from the other side, works to discontinuous target value [7-9]. Classification algorithms that are commonly include using K-nearest neighbour, regression analysis, as well as Support Vector Machine (SVM). Several other algorithms, including such neural networks, are used for classification as well as reversion [10]. Supervised classification techniques can be used to classify the device when the deliverables weren't very well but the system must uncover the framework inside the original data. Clustering, that also group's items following established number of clusters including such K-means cluster analysis, is show in Figure 3.

6. Conclusion

Despite the fact that with us concentrate in this systematic review (SLR) had been on a quite short period of time, the comparatively lot of research articles published indicated this is a popular area of research. This area of research combines several fast-

increasing fields: IoT, protection, machine learning, as well as data analytics are all buzzwords these days. We chose mentioned research information to check our SLR in order to assist researchers in identifying current trends inside the chosen research area. We encapsulated specific topics including such growing in popularity.
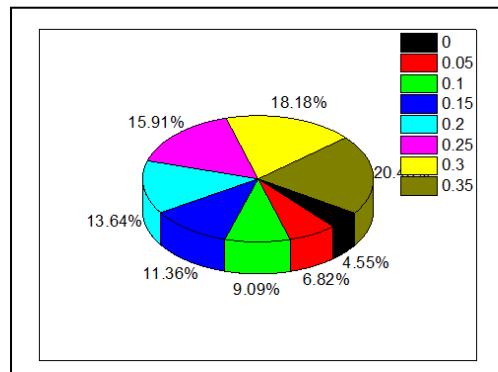
Figure 3. Performance Analysis for Machine Learning Techniques using Big Data link stability in Node Observation for IoT Security

References

[1]   Hossain, E., Khan, I., Un-Noor, F., Sikander, S. S., & Sunny, M. S. H. (2019). Application of Big Data and Machine   Learning in Smart Grid, and Associated Security Concerns: A Review. IEEE Access, 7, 13960–13988. https://doi.org/10.1109/ACCESS.2019.2894819

[2]   Karthika, P., Ganesh Babu, R., & Karthik, P. A. (2020). Fog Computing using Interoperability and IoT Security Issues in Health Care (pp. 97–105). https://doi.org/10.1007/978-981-15-2329-8_10

[3]   Babu, R. G., Nathan, V. V., Bino, J., Amali, C., & Ganesh, S. (2021). IoT Security Enhancement with Automated Identification Device using IOT SENTINEL. 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 518–523. https://doi.org/10.1109/Confluence51648.2021.9377165

[4]   C., C., R., G. B., M., S., M., A., R., B., & Balaji, M. (2020). Machine Learning Based Condition Recognition System for Bikers. 2020 7th International Conference on Smart Structures and Systems (ICSSS), 1–6. https://doi.org/10.1109/ICSSS49621.2020.9202245

[5]   Babu, R. G., Nedumaran, A., & Sisay, A. (2019). Machine Learning in IoT Security Performance Analysis of Outage Probability of link selection for Cognitive Networks. 2019 Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 15–19. https://doi.org/10.1109/I-SMAC47947.2019.9032669

[6]   Karthika, P., Babu, R. G., & Nedumaran, A. (2019). Machine Learning Security Allocation in IoT. 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 474–478. https://doi.org/10.1109/ICCS45141.2019.9065886

[7]   Ali, E. S., Hasan, M. K., Hassan, R., Saeed, R. A., Hassan, M. B., Islam, S., Nafi, N. S., & Bevinakoppa, S. (2021). Machine Learning Technologies for Secure Vehicular Communication in Internet of Vehicles: Recent Advances and Applications. Security and Communication Networks, 2021, 1–23. https://doi.org/10.1155/2021/8868355

[8]   Soe, Y. N., Feng, Y., Santosa, P. I., Hartanto, R., & Sakurai, K. (2020). Machine Learning-Based IoT-Botnet Attack Detection with Sequential Architecture. Sensors, 20(16), 4372. https://doi.org/10.3390/s20164372

[9]   Syafrudin, M., Alfian, G., Fitriyani, N., & Rhee, J. (2018). Performance Analysis of IoT-Based Sensor, Big Data Processing, and Machine Learning Model for Real-Time Monitoring System in Automotive Manufacturing. Sensors, 18(9), 2946. https://doi.org/10.3390/s18092946

[10] Rehman, A., Haseeb, K., Saba, T., Lloret, J., & Tariq, U. (2021). Secured Big Data Analytics for Decision-Oriented Medical System Using Internet of Things. Electronics, 10(11), 1273. https://doi.org/10.3390/electronics10111273

Enhancement of Power Generation in Highway Using Wind Energy Conversion System Integrated with PV

Kiruba.K [a,1], Deepika.D [b], Jaitha.G [b], Madhuja.S [b]

[a]*Assistant Professor, SRM TRP Engineering College, Trichy, TN, India*
[b]*UG Student, SRM TRP Engineering College, Trichy, TN, India*

Abstract. The rising global population and economic growth, combined with rapid urbanization, will result in a significant increase in energy demand. To solve this problem in the coming years, the world will need significantly more resources, primarily cleanly produced electricity. On the other hand, electricity demand is rising at twice the rate of overall energy consumption, and is expected to more than double by 2040. So, in order to meet the energy demands, the proposed approach includes a concept of a new Vertical Axis Wind Turbine (VAWT) design that generates power from moving vehicles and further integrated with PV for increased power generation. Seasonal variations can be accommodated by the related hybrid scheme. Using a charge controller, the produced power can be stabilized to a 12V output. The generated energy can be stored in batteries or supplied to the grid, acting as an energy storage device for society. The power that has been stored can be used in the future or during non-windy seasons.

Keywords. VAWT, Solar Energy, PV, Charge Controllers

1. Introduction

Renewable energy, also known as non-conventional sources, is generated from naturally replenished sources or processes on a consistent way. Wind and solar energy can now be harnessed effortlessly and maintained in more sophisticated and less expensive ways, and renewable energy sources are becoming a more important source of electricity. WECSs are commonly used in distributed generation systems, micro grids, and smart grids today, as well as in stand-alone systems for supplying electricity to isolated loads. Appropriate energy management processes must be used in both of these applications in order to optimize wind turbine energy output and pass wind-generated energy to consumers with high performance. Solar PV systems transform sunlight into electricity using cells. One or two layers of a semi conducting material, typically silicon, make up a PV cell. As light shines on a cell, it generates an electric field that causes electricity to flow through the layers. The movement of energy is proportional to the strength of the sun. PV cells are measured in terms of how much energy they produce. Meanwhile the hybrid power system is entirely reliant on intermittent renewable energy sources; the output voltage fluctuates, causing harm to devices that need a constant supply [1]. Since solar and wind power are intrinsically

---

intermittent and volatile, increasing their penetration in existing power systems may pose significant technical challenges, especially for poor grids or stand-alone systems lacking adequate storage capacity [2]. The average wind velocity in potential locations, such as coastal areas, is still now below the required level for effective operation of wind turbine [3]. The VAWT is built in such a way that is efficient to catch wind from all directions; with a power output of 28W for a speed of 6.1m/s; the VAWT's efficiency can be improved by changing the size and shape of the blade; the theoretical and experimental results vary since the theoretical measurement assumes that the wind reaches all eight turbine blades, which is not realistic [4]. This idea of producing a significant amount of energy without burning fossil fuels, which is the largest source of CO_2 emissions, and thereby helps to minimize reliance on fossil fuels and fossil fuel transportation [5].

2. Objective

The main aim of designing a wind turbine on a highway incorporated with PV is to contribute to the national trend of energy production in a realistic way. Traditionally, wind turbines have been used in rural areas. To generate electricity, the turbines will use the wind draught generated by vehicles on the highways. By integrating a PV system along with Wind Energy Systems the overall power output of the system can be enhanced.

3. Proposed Design

Our proposed model is shown in Figure. 1, has a wind turbine constructed in vertical manner so that it can effectively generate electricity from wind during vehicle movement. These turbines can be erected on highways with a high amount of fast-moving traffic. The produced electricity will be stored in batteries.
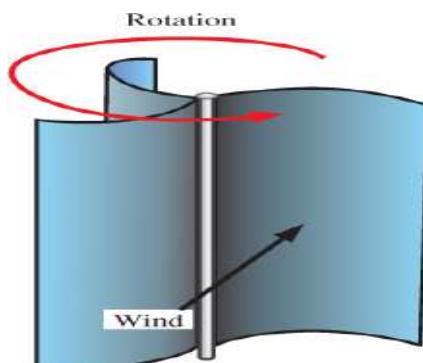
Figure 1. Vertical Axis Wind Turbine

VAWT has slower blade speeds since the blades are closer to the axis of rotation. Hence it is capable of producing electrical energy at very low wind speeds. So in order to improve the output the system design includes a solar PV combined with Wind [6] is shown in Figure. 2 .

Figure 2. Solar Pane

4. Design Methodology

In the Proposed system as shown in Figure. 3, Vertical Axis Wind Turbine (VAWT) integrated with Photovoltaic cell can be installed in highway median strip to produce power during high demand. The wind power obtained from rotation of blades is coupled with permanent magnet DC generator to generate DC output. The output of the designed turbine system is limited to low value since it has slower blade speeds. Hence it is efficient of producing electrical energy at low wind speeds. Hence in order to achieve high output, a PV cell is integrated with Vertical Axis Wind Turbine (VAWT) [7]. In order to prevent the battery from being overcharged, a charge controller can be used to regulate its output. The primary ratings of the charge controller can be of 15-A/200-W unit and by the mean way to speed up the process of solar charging the most common and efficient MPPT algorithm is used. The algorithm examines the output power obtained from panel and compares them with battery potential to attain maximum current into the battery. DC-DC converters are commonly used to efficiently generate a regulated voltage from a source that may or may not be well managed to a variable load. A LUO converter is used here, which steps up the voltage to produce a voltage higher than input voltage.
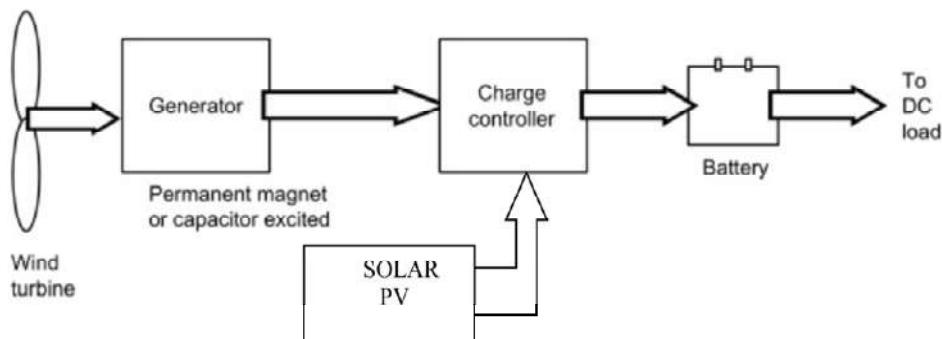
Figure 3. Block diagram of Proposed Methodology

A) Vertical Axis Wind Turbine:

The main aim of using VAWT instead of HAWT is the turbines constructed in a vertical manner don't require yaw control mechanism, which is nothing but the rotation of nacelle according to the direction of wind for power generation [6-10]. Hence the vertically constructed turbines are free from detecting wind and alignment mechanisms.

B) Generator

A permanent magnet DC generator is used here with double winding structure. The generators are mainly used here to extract the output from rotation of wind blades.

C) Solar PV

The most commonly used renewable source for transforming light energy into electrical energy is Solar. It is attained by using solar cells that exhibits photovoltaic effect. In our proposed methodology single solar cells are combined together to form segments such as panels capable of generating a voltage of 0.5 to 0.6 volts. Thereby for efficient amount of energy production additional number of solar modules can be used. Also the solar panels integrated along with wind turbine in our proposed design also contributes for enhancing the power, since the output generated from wind energy systems is not sufficient.

D) Charge Controller

The main purpose of using charge controllers is to prevent battery getting damaged from over charging. It is used to regulate the output parameters obtained from solar systems formerly entering into the battery. The process of comparing the output parameters of solar and battery is effectively achieved by using Maximum Power Point Tracking Algorithm. In addition converters can also be used to achieve maximum output.

5. Results and Discussions

The proposed system as shown in Figure. 4, is simulated in SIMULINK to obtain a constant voltage of 12V from charge controller. The output obtained from solar PV is accelerated by comparing its output with MPPT techniques. As shown in figure. 5, The output ripples are eliminated and the output voltage is effectively increased by using a LUO converter, which is a form of buck boost converter. The hardware model is realized as shown in Figure. 6.
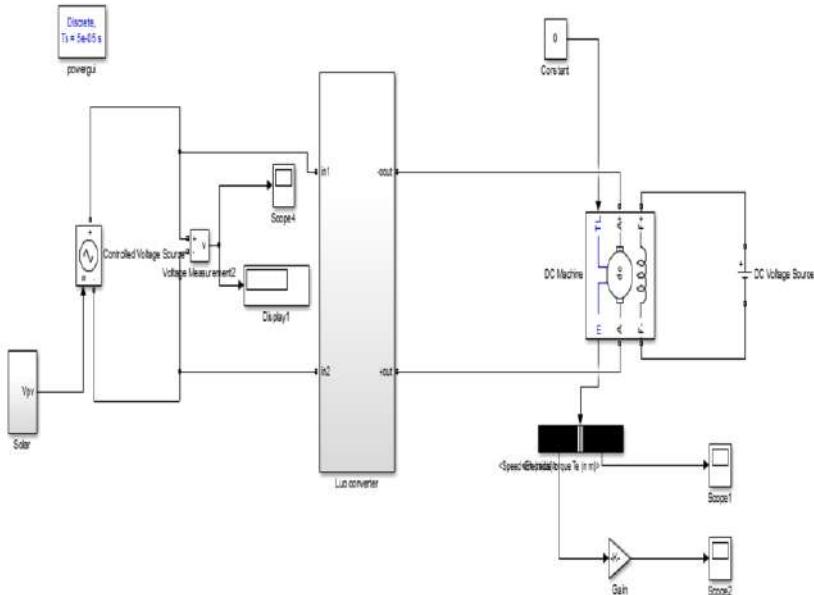
Figure 4. Simulink model of proposed system
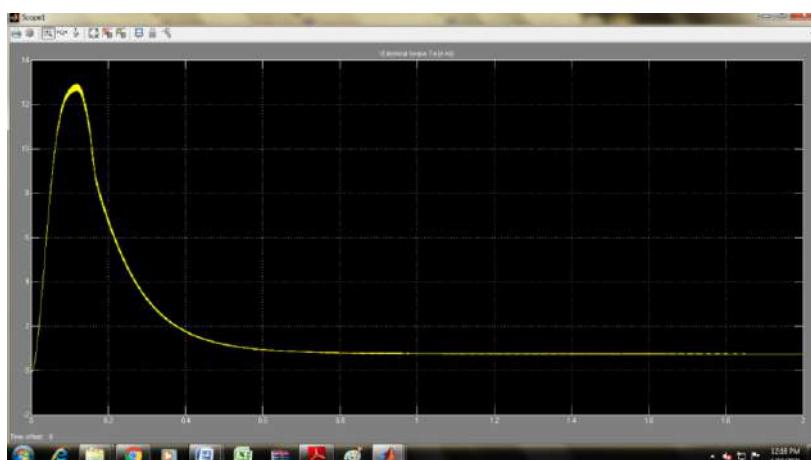
Figure 5. Simulation output for charge controller

Figure 6. Hardware model of proposed system

6. Conclusion

On highways, the proposed vertical axis wind mill model integrated with solar PV will be a good source of renewable energy. Wind energy produced by moving vehicles on highways can be used to produce electrical energy, and further it can be enhanced by combining this output with PV output. The obtained output can be stored in a battery and used for street lighting, traffic signals, and road studs, among other things. This design philosophy is intended to be environmentally friendly and long-lasting. If these turbines can be mounted on long, high-speed transmission lines, it is possible to produce a significant amount of electrical energy, which can be used in a variety of ways that resolve the energy crisis.

References

[1] Mohod, S. B., Parihar, V. R., & Nimkar, S. D. (2018). Hybrid power system with integration of wind, battery and solar PV system. In IEEE International Conference on Power, Control, Signals and Instrumentation Engineering, ICPCSI 2017 (pp. 2332–2337). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICPCSI.2017.8392134

[2] Badwawi, R. A., Abusara, M., & Mallick, T. (2015). A Review of Hybrid Solar PV and Wind Energy System. *Smart Science*, *3*(3), 127–138. https://doi.org/10.1080/23080477.2015.11665647

[3] Ehsan, M. M., Ovy, E. G., Chowdhury, H. A., & Ferdous, S. M. (2012). A Proposal of implementation of ducted wind turbine integrated with solar system for reliable power generation in Bangladesh. International Journal of Renewable Energy Research, 2(3), 397–403. https://doi.org/10.20508/ijrer.36008

[4] Kulkarni, S., Kulkarni, S. A., & Birajdar, M. R. (2016). Vertical Axis Wind Turbine for Highway Application. *Imperial Journal of Interdisciplinary Research (IJIR)*, *2*(10), 1–5.

[5] Sharma, P., & Harinarayana, I. (2013). Solar energy generation potential along national highways. *International Journal of Energy and Environmental Engineering*, *4*(1), 1–13. https://doi.org/10.1186/2251-6832-4-16

[6] Bhati, V.S., Berwal, R., & Kasniya, B. (2018). Modelling of Hybrid Solar / Wind Energy System for Rural , Remote and Hilly Areas in Rajasthan (India).

[7] Brijeshkumar N. Trivedi and Prof. Piyush Patel "Simulation and Analysis of Wind-Solar Hybrid Energy Conversation System", International Journal for Technological Research in Engineering ISSN (Online): 2347 – 4718 (February – 2018).

[8] Hybrid power generation system using wind energy and solar energy. (2016). *International Journal of Advance Engineering and Research Development*, *3*(07). https://doi.org/10.21090/ijaerd.030719

[9] N. Venkata Subbaiah and S.Deva Kumar "Power generation by using highway vertical axis wind mill", 2017 IJCRT, Volume 5, Issue 4, December 2017.

[10] Hannoon, N., Vijayakumar, V., Vengatesan, K., & Hidayat, N. (2019). Stability Assessment on Doubly Fed Induction Generator (DFIG) Wind Turbine Micro Grid Power System. Journal of Computational and Theoretical Nanoscience, 16(2), 778-785.

[11] Kumar, A., & Singh, S. K. (2015). Wind Power Generation On Highway. *IJIRST-International Journal for Innovative Research in Science & Technology*|, *1*(12), 426–431. Retrieved from www.ijirst.org

Smart Ticketing System for Public Transport Vehicles

J. Subramaniyan [a, 1], P.Senthil Kumar [b], S. Sripriya [b]
C. Jegan[b], M. Jenny [b]
[a]Assistant Professor, Department of EEE, SRM TRP Engineering College, Trichy TN, India
[b]Assistant Professor, Department of EEE, Roever Engineering College, Perambalur, TN, India

Abstract. Recent Growth in technologies has taken drastic improvements in all fields especially public welfare. Soon transport systems with demanding technologies like frequency Identification Devices (RFID), GSM, and face recognition will gain the spotlight. The RFID concept is applied in a public transport identification card which is a reliable system, automatically detects the passenger, and the camera will recognize the passenger's face and debit the fare following the distance traveled. A feedback message is forwarded to the corresponding person's mobile, as a sign of good security. The security system is equipped with a GSM modem.IR sensor will count the persons entering and exiting the bus.

Keywords. GSM, RFID, IR Sensor

1.   Introduction

Today, everything within the world is sensible and digitalized. Many advances are made within the transportation sector too. However, public conveyance buses in India have always been a neighborhood where such new advances have turned their faces out. Work Intelligent vehicle for conveyance is one among the research areas. Here global positioning system plays a crucial role to seek out positions. In certain urban areas, there could also be some errors to seek out the situation so an alternate approach may be a visual geometry [1-2]. In developing countries such as India, public transportation could be a significant source of revenue.

The ticket issuer faces various problems in issuing the tickets. Provided automatic issuing, the fare will be deducted corresponding to the distance traveled [3]. RFID technology is evolving in recent years. The proposed methodology includes a Contactless Fare Media Technology (CFMT) and an automatic money debiting system, as well as novel features, to bring India's public transportation bus system up to global standards [4]. The current ticketing system was riddled with flaws, malicious arguments among the general public, and corruption. It also intends to reduce fare-related fraud and revenue loss by utilizing open standard, secure transaction technology.

---

[1] J. Subramaniyan, Assistant Professor, Department of EEE, SRM TRP Engineering College, Trichy TN, India; E-mail: subramaniyan.j@trp.srmtrichy.edu.in.

Nowadays the general public transportation just like the metro is well advanced. To improve the efficacy, existing public transportation is to be mutated towards intelligent transportation within the market [5].

The offline-based issuing system for receiving the fare is difficult to validate the acquisition of a ticket by every passenger. After the arrival of the destination, the hard copy of the ticket is useless. The number of unsold tickets per day is extremely high. In an era of advanced techno-communications, India needs to specialize in inculcating an automatic system for confiscating carfare.

The RFID-based ticketing system aims to provide a comfortable, stress-free, and straightforward mode of transportation while also facilitating the scaling of manpower. During this automated system, we replace the traditional ticket system with a sensible system that stores all of the user's data [6].

This is often almost like the ATM card. If people notify the continuation of a specific bus, the person can receive the tickets using the RFID technique, and by showing the ticket ahead of the bus, the door opens automatically and closes after some predetermined seconds. We will know the number of passengers and free spaces on the bus by using an LCD.

A ticket issuing system is primarily proposed to overcome the difficulty of exchanging tickets between two individuals in a traditional ticket system. Ticket friendly machines keep the fine print about the events provided by promoters so that customers can buy tickets in their accounts using smart cards [7].

Time and money are valuable when we try to find the simplest way to avoid problems. When travelling by bus without carrying cash, the proposed method requires only one identification card. Once the passenger inserts the open-end credit into the ticket friend machine, the RFID reader within the open-end credit contains pay mode terms, which check for the account's balance. Following the receipt of RFID data, the information about transportation facilities is transmitted via an enormous display [8].

The evaluation of human faces in images is referred to as face detection. It disregards other objects such as buildings, trees, and bodies. Face perception is a hot topic in the computer vision community right now [9]. In applications such as video surveillance, human-computer interface, face recognition, and image management, face localization and detection is usually the first step. Identifying a face could be a consideration for facial recognition or expression analysis. The Haar-Classifier is used for face tracing in this paper [10].

During this current period, numerous incidents such as robbery, theft, and unwelcome entry occur on an unanticipated basis. As a result, security is important in this lifestyle. People are always busy with their day-to-day tasks, but they also want to ensure the safety of their prized possessions.

They forget to search for important things like keys, wallets, credit cards, and so on. If they don't have these, they won't be able to get into their house or anywhere else. A key, a security password, an RFID card, or an ID card are all required to gain access to a conventional security system. These security devices, on the other hand, have vulnerabilities, such as being forgotten or stolen by unauthorized people [11-13].

As a result, there is a need to create applications that guarantees a higher degree of protection, which could be used as a model. Our brain can only think in pictures, not words, which is one of its defining characteristics. You might forget where you put your car key, but you'll never forget to bring a smile with you. Everyone has a distinct appearance that God has bestowed upon them. Since it can represent a person's feelings, the face is the most important part of our bodies. Physiological characteristics

(face, fingerprint, finger geometry, hand geometry, palm, iris, ear, and so on) and behavioral parameters are the two forms of biometrics (signature, and keystroke attributes).

When you're sick, your behavioral parameters can change. A face detection and tracking device, in comparison to other biometrics, can be an easy and non-intrusive method. Face detection (1:1) and face recognition (1:2) will be the two categories of the device (1:2). (N:1). Face detection necessitates the separation of face and non-face regions, while face recognition necessitates the matching of a single face image with multiple images from the input image.

2.   Methodology

The traditional issue of the ticket system for debiting the travel fare is a vital source of loss in revenue. It is difficult to check the reliability of a ticket by every passenger. This paper explores an automated card-driven system using RFID for bus journeys in India. PTS is currently experiencing severe malfunction and security issues. The main issue is that there are several misunderstandings among passengers regarding fares. Besides, a security crisis arises in PTS.

2.1   Existing system

Each bus is generally controlled by a conductor. The conductor takes money from each passenger and issues them a ticket. Initially, tickets are printed on paper or tokens. Nowadays, handloom machines are used to print tickets. The passenger must carry the ticket until the train arrives at its destination, the conductor must ensure that everyone has a ticket, the ticketing time is relatively long, and paper is required to print the ticket. As an example, suppose a passenger wishes to travel by bus. He should bring the cash with them. The conductor then collects the money and hands out the ticket. This must be repeated by all passengers. This results in more time and waste, as well as a drain on human resources and energy.

2.2   Proposed system

Nowadays the public transportation system like the metro is well advanced. The proposed system is shown in Figure 1. In this system, RFID comprises two parts, RFID Tagger, and Reader. RFID Tag contains data related to the user. RFID reader deduce the above information's from the passenger RFID Tag However face recognition is done by a camera that is implemented by haar cascade techniques. IR sensor counts the persons entering the bus and then the bus door will be opened with the help of a relay circuit. The distance will be analyzed using a proximity sensor. Whenever the passenger gets down from the bus IR sensor count will be decremented and the reader detects the tag again and then the door will be opened. The amount will be debited from the account automatically based on distance travelled.
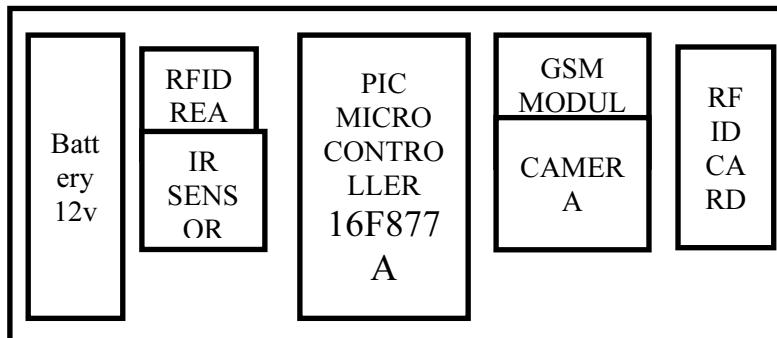
Figure 1. Proposed Block Diagram

3.   Result discussion

An RFID-based fare collection system deployed as shown in Figure 2, could soon eliminate operating costs.
Authorities shall trace the vehicle in real-time which in turn reduce delays and thereby providing extra resources (buses or trains) to corresponding routes
The reader can communicate with the tag via radio waves. With the RFID smart card in his/her pocket, an individual can be detected and debited the appropriate amount.

Figure 2. Smart Ticketing System

4.   Conclusion

The working model focus on a hybrid energy generation technique where it combines solar system and vertical axis wind turbine system, The Generated power from all the renewable sources needs concentration on a better state of extraction of power without any distortion. We have proposed a Modified DC to DC converter topology for providing a ripple-free output. Future work on this extension may include detailed analysis of the selection of renewable power on season and availability along with a suitable controller.

References

[1]   Ravada, B. R., Tummuru, N. R., & Ande, B. N. L. (2021). Photovoltaic-wind and hybrid energy storage integrated multi-source converter configuration for DC microgrid applications. IEEE Transactions on Sustainable Energy, 12(1), 83–91. https://doi.org/10.1109/TSTE.2020.2983985

[2]   Sathe, A. D., & Deshmukh, V. D. (2017). Advance vehicle-road interaction and vehicle monitoring system using smart phone applications. In *Proceedings of 2016 Online International Conference on Green Engineering and Technologies, IC-GET 2016*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/GET.2016.7916825

[3]   Herdiyanto, D. W., Endroyono, & Pratomo, I. (2017). Passenger authentication and payment system using RFID based on-board unit for Surabaya mass rapid transportation. In *Proceeding - 2016 International Seminar on Intelligent Technology and Its Application, ISITIA 2016: Recent Trends in Intelligent Computational Technologies for Sustainable Energy* (pp. 305–310). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ISITIA.2016.7828677

[4]   Hargunani, K., Kengar, P., Lokhande, M., Gawade, R., & More, S. K. (2018). Integrated Bus System Using QR Code. In *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICCUBEA.2018.8697591

[5]   Akter, S., Islam, T., Olanrewaju, R. F., & Binyamin, A. A. (2019). A Cloud-Based Bus Tracking System Based on Internet-of-Things Technology. In 2019 7th International Conference on Mechatronics Engineering, ICOM 2019. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICOM47790.2019.8952037

[6]   Harini, B. K., Parkavi, A., Supriya, M., Kruthika, B. C., & Navya, K. M. (2020). Increasing Efficient Usage of Real-Time Public Transportation Using IOT, Cloud and Customized Mobile App. SN Computer Science, 1(3). https://doi.org/10.1007/s42979-020-00161-8

[7]   J.Subramaniyan, S.Ajitha, A.Jothi, S.K.Nitharsini, Secured Disposal System, International Journal Of Information And Computing Science, **6** (2019), 639.

[8]   J.Subramaniyan, V.Muthukumarasamy, V.Vengatesan, G.Ramprakash, Money Withdrawal from ATM By Using OTP with Secure Bio Metric JASC: Journal of Applied Science and Computations, 5(2018).

[9]   Ganesh, S., Vengatesan, V., Richard Jimreeves, J., & Ramasubramanian, B. (2020). Simultaneous network reconfiguration and pmu placement in the radial distribution system. *Advances in Mathematics: Scientific Journal*, 9(10), 8143–8151. https://doi.org/10.37418/amsj.9.10.44

[10]   Ganesh, S., Ram Prakash, G., & Michline Rupa, J. A. (2021). Optimal placement of green power generation in the radial distribution system using harmony search algorithm. In *Advances in Intelligent Systems and Computing* (Vol. 1167, pp. 245–252). Springer. https://doi.org/10.1007/978-981-15-5285-4_23

[11]   Selvaraj, G., Thangaiyan, V., & Rajangam, K. (2021). Numerical method for islanding the location of ground fault in the material based distribution system. *Materials Today: Proceedings*, *45*, 634–639. https://doi.org/10.1016/j.matpr.2020.02.723

[12]   Ganesh, S., & Kanimozhi, R. (2018). Multi objective approach for capacitor placement based on load balancing index in radial distribution system. *Journal of Computational and Theoretical Nanoscience*, *15*(1), 282–288. https://doi.org/10.1166/jctn.2018.7086

[13]   Ganesh, S., & Kanimozhi, R. (2018). Meta-heuristic technique for network reconfiguration in distribution system with photovoltaic and D-STATCOM. IET Generation, Transmission and Distribution, 12(20), 4524–4535. https://doi.org/10.1049/iet-gtd.2018.5629

Reduction of Voltage Ripple in DC Link of Wind Energy Conservation System Using Modified DC-DC Boost Converter

L. HUBERT TONY RAJ [a, 1], R. SIVAKUMAR [b], R. AKASH [b] and
M. ANANDHA CHAKRAVARTHI [b]

[a]*Assistant Professor, Department of EEE, SRM TRP Engineering College, TN, India*
[b]*UG Student, Department of EEE, SRM TRP Engineering College, TN, India*

Abstract. Renewable energy provisions must be extracted in a more resourceful way, with a power converter added to the mix. If the supply-demand curve rises with the seasons, it becomes clear that renewable energy sources are used to provide clean energy. This clean energy cannot be used on load directly due to fluctuating conditions, to solve this problem a modified DC to DC converter with a ripple-free output is introduced. The Vertical Axis Wind Turbine (VAWT) and Solar PV were combined to achieve a constant DC output in a hybrid renewable energy conversion system. For renewable energy applications, a redesigned converter with ripple-free output is used. The simulation is made under MATLAB/SIMULINK and experimental parameters were measured using a nominal prototype.

Keywords. VAWT, Solar PV, Hybrid system and Ripple

1. Introduction

Renewable electricity sources including photovoltaic (PV) and wind power account for the majority of overall renewable energy production in India. Since both origins are sporadic in nature, for a stable operation, successful solutions are needed [1]. PV and wind have the intrinsic advantage of being complementary in design. As a result, combining PV and wind energy sources will help to mitigate long-term intermittency to some degree.

A hybrid energy system combines multiple energy sources (two or more) with effective energy conversion technologies. linked to provide electricity to the local load/grid There is no single standard or framework so it falls under the distributed generation umbrella [2]. It gains from the lower line and transformer losses, as well as lower environmental impact.

However, as a result of this, the magnitude of short-term power variations would be greater, putting more strain on battery storage and potentially increasing battery scaling. Furthermore, the lifespan of the battery is harmed [3]. As a result, the issue of battery lifespan depletion and improved sizing can be solved by integrating battery and supercapacitor as a combined energy storage system but this includes a static system for improvement.

---

[1] L. Hubert Tony Raj, Assistant Professor, Department of EEE, SRM TRP Engineering College; Trichy, India, E-mail: ltonyraj@gmail.com.

Furthermore, since PV, wind, and storage units typically have lower voltage ratings, voltage boosting is needed. As a result, the main study areas are appropriate converter configurations and control mechanisms for the application of Hybrid Energy Storage and green power sources [4].

The most popular renewable energy sources are wind and solar energy. Small-scale off-grid (microgrid) systems are deployed in rural locations. rather than constructing a transmission line to transport electricity from power plants to consumers. A microgrid system is a compact system that mostly uses solar and wind energy [5-7]. Because of the solar and wind systems' instabilities, intermittency, and high costs, non-renewable energy is becoming more common.

To ensure a consistent and reliable power supply, sources and energy storage have been added. The method is referred to as renewable energy sources are combined with other energy sources. Renewable energy networks are hybrid (HRESs) [8]. HRESs may be classified as on-grid or off-grid based on their connection to the power grid [9]. HRESs has several benefits, including the potential to use a variety of energy sources to serve a population, increasing the penetration of renewable energy sources and reducing the use of fossil fuels, and stability [10].

This paper explains the combination of both Vertical axis wind turbine system along with the Solar PV which combines to supply a nominal Resistive Load. The configuration of the hybrid module is driven with an efficient Dc to Dc converter to supply the load [11]. Here the system is under open-loop condition. The forthcoming stages of explanation fall on the Methodology, Simulation Result discussion, and Hardware implementation.

2.   Methodology
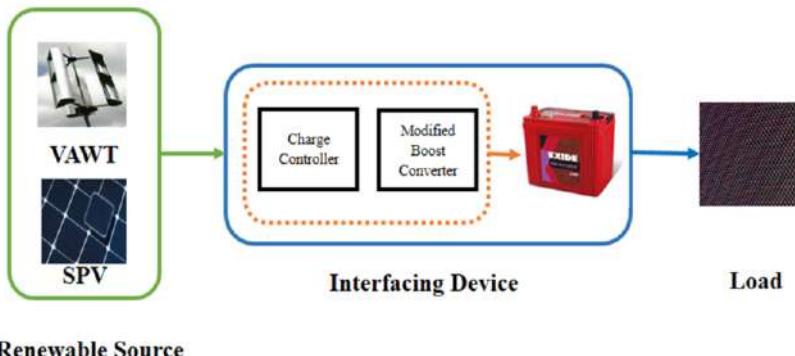
Figure 1. Block Diagram of Hybrid Renewable Source with Interfacing System and Load
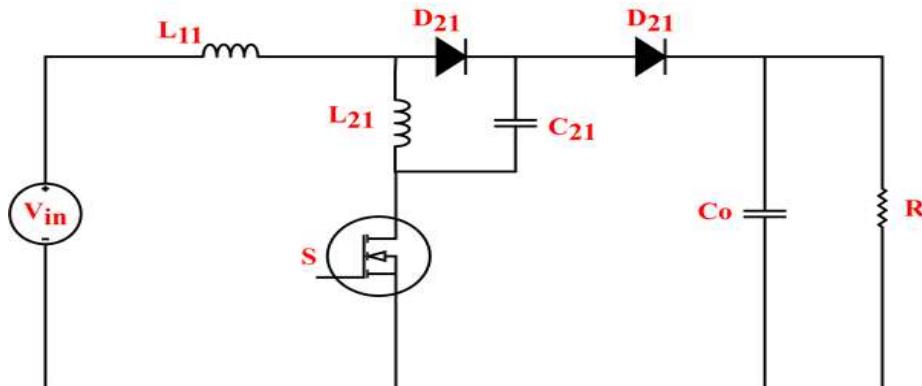
Figure 2. Circuit Diagram of the Modified Converter

The proposed modification on the converter is to interface the attained less input renewable input to a constant load Figure 1 shows the exact experimental block diagram of the proposed system.

To drive a load that is continuous in demand a constant source with some boosting technology is required. This technology needs to be in a specific rile free output regarding the effect of the ripple on the load influence the life and operation standard will be reduced in stages after some time

Figure 2, describes the circuit model of the proposed system, The operating stages of the modified circuit are as follows

During Ton D21 is forward biased and D12 is reverse biased L21 and C21 are connected in parallel, whereas C0 is supplied to load

During Toff D21 is reverse biased and D12 is forward biased L21 and C21 are connected in series, whereas C0 is supplied to load

Output Equation:

$$Vo = \frac{2-D}{1-D} \, Vin \text{ --------(1)}$$

$$Gain = \frac{2-D}{1-D} \text{ --------(2)}$$

The Input Energy attained from the system

$$IEO = \int_0^{T1} V(s)\, i(s)\, dt = Vs \; Is \; T1 \text{ --------(3)}$$

Stored Energy

$$El = \frac{1}{2} \, Li2 \text{ ------- (4)}$$

$$Ec = \frac{1}{2} \, Cv2 \text{ --------(5)}$$

3.   Results and Discussion

The modified circuit with the combined input sources was studied using A MATLAB Simulink Tool. The design procedures under modified DC to DC converter explain the Current in the Input and the Inductor in Figure 3. The input current from the source is not exceeding 0.14A maximum and the inductor current is 0.001 A at the stage of operation

The Gate pulse of the MOSFET switch is including 20 K Hz of frequency cycle for triggering the MOSFET and the voltage across the switch is shown in Figure 4.

The Diode Voltage of Vd1 and Vd2 are perfectly inverse this is represented in the below stimulated Figure 5. Later the finite operating output stages attained from the Modified converter have the current of 2 mA in range together with the output voltage of 60 V as an average output voltage and current, this is clearly explained in Figure 6.
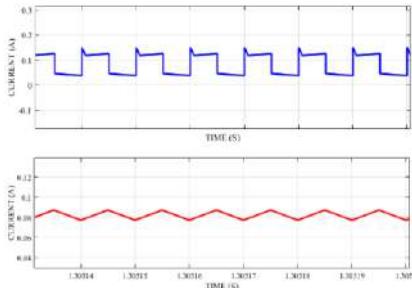
Figure 3. Input current and current at the inductor
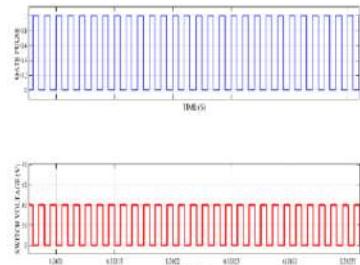
Figure 4. Gate Pulse of MOSFET and Switch Voltage
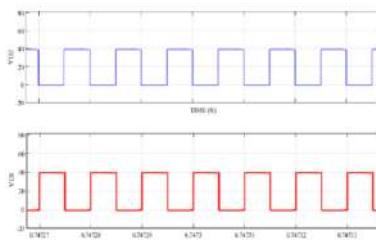
Figure 5. Diode Vd1 and Vd2 voltage
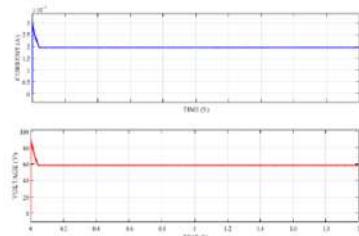
Figure 6. Output Current and Voltage

4.   Hardware results and Discussion

Figure 7. Experimental Setup of the proposed LED Display projecction with SPV and VAWT.

The above-represented Figure 7 is for the Hardware prototype which has a hybrid structure under open-loop conditions. The controlling input from the hybrid combination is fed to the battery using a successive charge controller and a modified

boost converter. Table 1 represents the hardware components used for designing the overall setup of the hybrid module under common usage. This combination will be a precise design for minimum green energy extraction.

Table 1. Components and ratings for the successive design of prototype

SI. No	List of Hardware Components	Rating
1.	Solar Panel	100 watts
2.	LED Display	Amps – 5 to 6 Amps
3.	Aluminum Sheet	1.25 mm
4.	Shaft	35 inch
5.	Gear	1: 4 ratio
6.	Ball Bearing	52mm Diameter
7.	Dynamo	12 volts
8.	Charge Controller	10-30 AMPS
9.	Boost Convertor	12 – 46 V
10.	Rechargeable Battery	60 Ah

a   b

Figure 8. **(a)** Modified Converter **(b)** Charge Controller

The hardware is made for at VAWT that rotated under 5 m/s of minimum wind speed design. Dynamo power output is not well sufficient to drive an LED this made an interest to support the
dynamo which provides a seasonal power supply to the club with the Solar PV system under an open-loop condition to rectify the demand. This Combination was a success to describe the efficient extraction under the conversion stages a Modified DC to DC converter is proposed. The Segmented output is boosted via this proposed converter the system is also supported with a charge controller to decide the overflow on the battery which reduces the life span of the same.
The LED display with a decorative display needs a constant supply this is achieved with the battery to let a constant power set for the LED segment. The proposed LED segment is a DOT Matrix Type of Wall that Displays the Department of EEE SRM TRPEC as shown in Figure 8.

5.   Conclusion

The working model focus on a hybrid energy generation technique where it combines solar system and vertical axis wind turbine system, The Generated power from all the

renewable sources needs concentration on a better state of extraction of power without any distortion. We have proposed a Modified DC to DC converter topology for providing a ripple-free output. Future work on this extension may include detailed analysis of the selection of renewable power on season and availability along with a suitable controller.

References

[1] Ravada, B. R., Tummuru, N. R., & Ande, B. N. L. (2021). Photovoltaic-wind and hybrid energy storage integrated multi-source converter configuration for DC microgrid applications. IEEE Transactions on Sustainable Energy, 12(1), 83–91. https://doi.org/10.1109/TSTE.2020.2983985

[2] Kushwaha, R., & Singh, B. (2021). Bridgeless Isolated Zeta-Luo Converter-Based EV Charger with PF Preregulation. *IEEE Transactions on Industry Applications*, *57*(1), 628–636. https://doi.org/10.1109/TIA.2020.3036019

[3] Elangovan, P., & Mohanty, N. K. (2016). PI controlled active front end super-lift converter with ripple free DC link for three phase induction motor drives. *Journal of Power Electronics*, *16*(1), 190–204. https://doi.org/10.6113/JPE.2016.16.1.190

[4] Kushwaha, R., & Singh, B. (2020). Design and Development of Modified BL Luo Converter for PQ Improvement in EV Charger. In IEEE Transactions on Industry Applications (Vol. 56, pp. 3976–3984). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/TIA.2020.2988197

[5] Lu, Z., Li, C., Zhu, A., Luo, H., Li, C., Li, W., & He, X. (2021). Medium Voltage Soft-Switching DC/DC Converter with Series-Connected SiC MOSFETs. IEEE Transactions on Power Electronics, 36(2), 1451–1462. https://doi.org/10.1109/TPEL.2020.3007225

[6] Sindhuja, S., & Sripriya, S. (2013). Continuous Conduction Mode of Bridgeless SEPIC Power Factor Correction rectifier. In *Proceedings of International Conference on Computation of Power, Energy, Information and Communication, ICCPEIC 2013* (pp. 28–30). IEEE Computer Society. https://doi.org/10.1109/ICCPEIC.2013.6778516

[7] Kushwaha, R., & Singh, B. (2019). A Modified Luo Converter-Based Electric Vehicle Battery Charger with Power Quality Improvement. *IEEE Transactions on Transportation Electrification*, *5*(4), 1087–1096. https://doi.org/10.1109/TTE.2019.2952089

[8] Kaliamoorthy, M., Rajasekaran, V., Raj, I. G. C., & Raj, L. H. T. (2014). Generalised hybrid switching topology for a single-phase modular multilevel inverter. *IET Power Electronics*, *7*(10), 2472–2485. https://doi.org/10.1049/iet-pel.2013.0847

[9] Fang Lin luo Senior Member IEEE , Hong ye Member IEEE , Muhammad H. Rashid Fellow IEEE "Super-Lift Luo-Converters'' University of West Florida Pensacola, Florida 325 14, USA.Downloaded on March 02,2021 at 06:12:58 UTC from IEEE Xplore.

[10] Mylsamy, K., Vairamani, R., Irudayaraj, G. C. R., & Lawrence, H. T. R. (2014). Experimental validation of a cascaded single phase H-bridge inverter with a simplified switching algorithm. *Journal of Power Electronics*, *14*(3), 507–518. https://doi.org/10.6113/JPE.2014.14.3.507

[11] Ganesh Babu R., Karunakaran A., Manikandan G., Kalimuthu Kumar S., Selvameena R. (2021) IoT in Smart Automation and Robotics with Streaming Analytical Challenges. In: Singh K.K., Nayyar A., Tanwar S., Abouhawwash M. (eds) Emergence of Cyber Physical System and IoT in Smart Automation and Robotics. Advances in Science, Technology & Innovation (IEREK Interdisciplinary Series for Sustainable Development). Springer, Cham. https://doi.org/10.1007/978-3-030-66222-6_7

917

Smart Reader Glass for Blind and Visually Impaired People

S. ANBARASI[a 1], S. KRISHNAVENI[a], R. ARUNA[b]
K. KARPAGASARAVANAKUMAR[c]

[a]Associate Professor, P.S.R. Engineering College, Sivakasi, TN, India
[b]Assistant Professor, P.S.R. Engineering College, Sivakasi, TN, India
[c]UG Student, P.S.R. Engineering College, Sivakasi, TN, India

Abstract. Visually impaired people fail to read the text with existing technology. The proposed project targeted to design a spectacle with a camera by which the blind visually impaired people can read whatever they want to read based on contemporary OCR (optical character recognition) technique and text-to-speech (TTS) engines. This proposed smart reader will read any kind of documents like books, magazines and mobiles. People can access this novel technology with blindness and limited vision. The earlier version of the proposed project was developed successfully with mobile reader which had certain drawbacks such as high cost due to the need of android mobile, not user friendly and improper focusing. To overcome these disadvantages, a spectacle type reader with camera is proposed in this project, which will be cost effective and more efficient.

Keywords: Micro camera, Python, Raspberry pi, Smart reader, Text-to-speech

1. Introduction

Vision is the beautiful gift that God has given to all living creature. Whereas, the people with blindness and deprived of vision can match with the person with normal vision by gathering information using their hearing capability only. Most of the visually impaired people use Braille system for reading documents and books [1]. However, the Braille system has many disadvantages as follows [2]

- Converting all the materials in to Braille is very expensive
- The Braille materials can't be read by people with normal vision
- The error in Braille system can't be corrected easily
- Reading document and books requires more practice and time and this system is not applicable for all books, documents and many articles.
- It would be very hard to make every visually challenged person to adopt Braille method.

Even though lots of tools are available in online to convert text to voice there are certain limitations such as they are not wearable and user friendly, not in compact size and couldn't be used commonly to read all materials like books, computers and mobiles. To overcome the above-mentioned problems our project aimed to design assistive smart glasses in wearable design format for the blind and visually impaired people. This smart

---

[1] S. ANBARASI, Associate Professor, P.S.R. Engineering College, Sivakasi, TN, India,
Email: tsoundarapandiananbu@gmail.com

spectacle with micro camera setup proposed in this project is designed to support reading printable version of any books, documents, mobile texts by converting text to audio, which can be heard by microphones or speakers. This portable and economical smart spectacle is programmed with raspberry pi module and the image processing technique helps in recognizing and extracting the text from the image [4, 5]. Finally, the extracted text is converted into speech and can be heard by blind and visually impaired people. The final hardware model is tested with two test samples, the first one is with book page and another one is mobile document. The designed smart glass converted both test samples into a right audio format. This project is very affordable to all category of people and will be more useful.

2. Objective

In recent days, various learning tools are being developed for blind and visually impaired people using artificial intelligent techniques [3-5]. However, most of them are not cost effective. In addition, each reading aids can be used to read a specific application. The reading assist used in Personal Computer (PC) is not suitable to read texts in book [6]. Furthermore, not all the reading aids are user friendly, affordable and compact size. Though various types of reading assistants are present they could not be used to read multiple texts which is presented in the printed materials or books, computer screens, and newspaper and mobile screens. In addition, the existing techniques are not configured for different fonts, colors, scales and multiple orientations. Hence, to overcome the above-mentioned disadvantages our project aimed to develop smart reading spectacles for blind and visually impaired people. In our project, the micro camera is fitted on the reader's spectacle, which captured the text image to be read and converted this text image to text again and further it is converted into audio using text to speech engine. The main objective of this project is to satisfy the needs of visually impaired people in reading and browsing articles in documents, computers or in mobiles. This spectacle can also be suggested for normal people during their travelling time. Since, people getting distraction and tired of reading large number of pages during travel time. This proposed project can be offered in compact size and affordable prize due to its simple design.

3. Description of Smart Reading Model

The block diagram representation of Raspberry pi designed smart glass model is shown in Figure 1. The text image, which has to be converted into audio, is placed under the micro camera located in spectacle type smart reader. A separate portable rechargeable power bank arrangement is provided for raspberry pi and camera setup. In raspberry, pi an assistive system offered text localization algorithm with some accuracy. On startup, the system verifies its features, matching, alignment, tracking, motion analysis, deformation estimation and 3D reconstruction. Finally, the good quality of image is then converted into text and with the help of text-to-speech (TTS) engines; the audio can be heard through headphones.
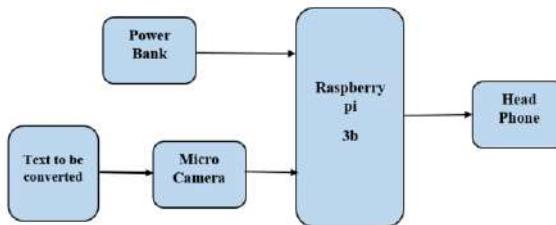
Figure 1. Block diagram representation of smart reader glass

4. Improving Quality with Image Processing

The main challenges faced in computational reading analysis with smart mechanism are different categories of materials, different colors and different backgrounds. Hence, appropriate approaches might be used for processing the images to remove noise, correcting edges and illuminations. In this proposed project the Raspberry Pi embedded with integrated peripheral devices such as Bluetooth and USB, which can be used for wireless transformation of data output.

The digital image processing allows to extract clear and required information from the enhanced image is also used in this proposed project in which the image is converted to an array of small integers, called pixels, representing a physical quality such as scene radiance, stored all devices and their connectivity is correct. Further, computer and other digital hardware modules carried out the segmentation and extraction process.

5. Working Methodology

The capture button is placed on the spectacle handle. When it is clicked, the document has to be read will be captured with the camera which is linked to ARM microcontroller via USB. The selected image is further processed with Optical Character Recognition (OCR) technology in which the internal process is carried out and the labels are separated using CV library [7]. Hereby the scanned image is converted in to printed text. In this proposed project, the TESSERACT library is suggested which is used to convert text-to-speech. Finally, the printed text is converted into audio using text to speech engine and it will be pronounced. This voice can be heard through the microphones [8].

The image to text conversion differs for each document. For written messages, some special methods such as machine learning might be used for clear transformation of image to printed text. However, this project mainly aimed to convert documents in printed textbooks, computers and mobiles and the machine learning technology is suggested in near future while using to read hand written materials. The circuit diagram for the proposed smart reader glass was drawn using …………. Software and depicted in Figure 2.
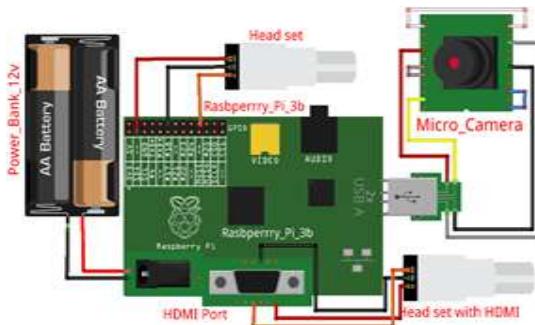
Figure 2. Circuit Design of Smart Reader Glass

6. Flow of Assessment

The methodology of proposed smart reader glass for blind and visually impaired is described well in section 5. As the image processing plays a vital role in this project it is essential to describe the process flow for ease of understanding of the readers. clearly explains the concepts of conversion of the scanned image into an audible text.

6.1 Image Capturing/ Scanning:
A micro camera is fitted on the glass of the spectacle, which is used to capture the documents to be read after the captured button is clicked.

6.2 Pre-Assessment:
The main processes in Pre-Assessment are skew correction, linearization and noise removal. The primary step in pre assessment is Skew detection, which is applied to the scanned text image to convert it into digital format. The second step in pre assessment is linearization, which supports to linearize the opto-electronical conversion function (OECF) with three methods such as Full Image Range, Modulation Transfer Function (MTF) or Full Data Range. The final step in pre-assessment is noise removal in which the undesirable by-product such as random variation of brightness or color information in images, and is usually an aspect of electronic noise is removed to eliminate obscures in original information.

6.3 Feature extraction:
The feature is stated as a piece of information about content of image. Feature Extraction aims to fine tune specific structures in the image such as points and edges.

6.4 Recognition:
Image recognition is a technique for capturing, processing examining and sympathizing images. It is an ability to notice fine features in a digital image.

6.5 Text to Speech:
The good quality of text image is finally converted into audio with the help of text-to-speech (TTS) engines, which can be heard through headphones.

7. Hardware Implementation

The proposed smart reader glass system includes the hardware components, which include Raspberry Pi, battery package, rectifier unit, micro camera and microphone as shown in Figure 3.

The Raspberry Pi is a small computer that plugs into the monitor and connects to a

keyboard. Raspberry Pi 3 is an active compact sized electronics board which is more powerful with 10x faster processing speed than first generation model. In addition it is equipped with wireless Bluetooth facilities with HDMI or VGA port. Raspberry Pi 3 is powered up with battery setup connected through micro USB cable. The image-processing unit, the software program modules and python can be installed in the Raspberry Pi.
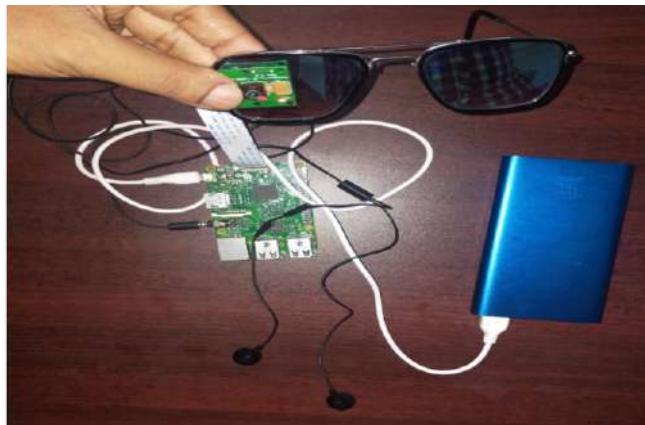
Figure 3. Hardware module equipped with spectacle and headphone

The Micro camera is fitted on the spectacle as shown in Figure 4. The input power supply from battery is connected through the USB port and a regulated output voltage of +5V with output current of 100 mA fed to the Raspberry Pi and the audio output can be accessed through the headphones.

Figure 4. Hardware module with Raspberry Pi

8. Summary and Conclusion

The proposed project - Smart reader glass for blind and visually impaired people mainly aimed to assist the blind and visually impaired people to read the documents in the form of text book, mobile screen or computer screen. Even though some of the glasses are available in the market they are too costly and not affordable to common people. Our proposed project is designed with simple components which can be placed in shirt pocket. Hence this product is user friendly and will be available at very low price and

compact size. The audio output can be heard using normal microphones used for mobile phone. The rigid assembling of hardware components ensure the lifetime of the proposed smart glass. Really, this smart reading spectacles will be an optimal choice for visually impaired peoples.

References

[1] Mohamed, S. A. E., Hassanin, A. S., & Othman, M. T. Ben. (2014). Educational System for the Holy Quran and Its Sciences for Blind and Handicapped People Based on Google Speech API. Journal of Software Engineering and Applications, 07(03), 150–161. https://doi.org/10.4236/jsea.2014.73017

[2] Sabab, S. A., & Ashmafee, M. H. (2016). Blind Reader: An intelligent assistant for blind. 2016 19th International Conference on Computer and Information Technology (ICCIT), 229–234. https://doi.org/10.1109/ICCITECHN.2016.7860200

[3] Sarkar, R., Das, S., & Roy, S. (2013). SPARSHA: A Low Cost Refreshable Braille for Deaf-Blind People for Communication with Deaf-Blind and Non-disabled Persons (pp. 465–475). https://doi.org/10.1007/978-3-642-36071-8_37

[4] Sharma, A., Devi, S., & Verma, J. K. (2020). Braille Book Reader using Raspberry Pi. 2020 International Conference on Computational Performance Evaluation (ComPE), 841–843. https://doi.org/10.1109/ComPE49325.2020.9200110

[5] K, S. G., & T, S. (2017). Real Time Implementation of Optical Character Recognition Based TTS System using Raspberry pi. International Journal of Advanced Research in Computer Science and Software Engineering, 7(7), 149. https://doi.org/10.23956/ijarcsse/V7I7/0117

[6] Ali Hassan, E., & Tang, T. B. (2016). Smart Glasses for the Visually Impaired People (pp. 579–582). https://doi.org/10.1007/978-3-319-41267-2_82

[7] Alkhalaf, K. (2014). OCR-Based Electronic Documentation Management System. International Journal of Innovation, Management and Technology, 5(6). https://doi.org/10.7763/IJIMT.2014.V5.560

[8] Bhui, N., Prasad, D., Sinha, A., & Kuila, P. (2021). Design of an Automatic Reader for the Visually Impaired Using Raspberry Pi (pp. 175–188). https://doi.org/10.1007/978-981-15-7533-4_14

Demand Side Management for Commercial Area Using Teaching Learning Based Optimization

Dhanshri Narayane [a,1], Amarjeet S Pandey [a], D B Pardeshi [a], Renuka Rasal [a]

[a] *Department of Electrical Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra, India*

Abstract: In Smart Grid Demand side management (DSM) plays a crucial role which permits customers to form educated selections concerning their energy consumption. It allows the strength to companies lessen the height load call for and reshape the burden profile. Most of the present demand aspect management ways utilized in ancient energy management system is with specific techniques and algorithms. In addition, the present ways handle solely a restricted range of governable a lot of restricted varieties of loads. This paper covers a requirement aspect management strategy supported load shifting technique for demand aspect management of future sensible grids with an outsized range of devices of many sorts. The day-in advance load shifting technique is proposed and mathematically formulated as a minimization problem. Teaching Learning Based Optimization (TLBO) is an efficient optimization is proposed. Considering Smart Grid with commercial customer, Simulations has been carried out. The respective results emphasis that the considered demand side management strategy attains substantial savings, whereas suppresses the mark of load demand of the smart grid. The outcome is by improve in sustainability of the smart grid, in addition to reduced standard operational value and carbon emission levels. The proposed algorithms can be easily applied to various optimization problems.

Keywords: Commercial, Demand Side Management, Smart Grid, TLBO, Simulation, Demand Response.

1. Introduction

Modern electrical power systems are continuously fulfilling the demand of power consumption with a satisfactory level of reliability, environmental friendliness and quality. Though, meeting these challenges has become more and more difficult due to the increase in electricity demands caused by population and industrial growth. In addition to this difficulty, the infrastructure investments required to guide the growing worldwide power demand for might be large due to the fact growing older strength machine additives will need to get replaced.

Accordingly, the same old practices geared toward balancing energy supply and demand must to be tested carefully. Power price varies in accordance with real-time electricity needs. Especially, increases charges boom as demand for rises and vice versa.

---

[1] Dhanshri Narayane, Department of Electrical Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra. Email: narayanedhanshrielect@sanjivani.org.in.

The administration of power systems is classified into supply-side management and demand-side management (DSM). Both approaches are useful for reducing possibilities, to boost network loading capacity and shorten peak loads. Smart Grid plays vital role in Demand Side Management that supervises the load on customer side which targets on efficiency and sustainability. Modern conversation technology has expedited the productive implementation of smart grid (SG), that aims at plans of demand aspect management (DSM), consisting of demand response (DR). The major ambition of sensible grid is to fulfill the forthcoming power demands. these days sensible grid is one in all the goodly trends within the wattage business and has achieved acceptance in several application departments.

The consumer's demand for energy is altered in DSM through monetary incentives. Demand Side Management processes the choice; planning, associate degreed implementation of measures designed to own an importance on electricity demand on the customer-side of electrical meter, that the utility be able to either management directly or stirred up indirectly. Due to supply-demand gap of electrical Energy is continuously increasing thanks to such technologies which will support to meet the target in nearby future. Underneath such situations, Demand side management is applied to manage the load demand and cut down the energy consumption. The most acquainted motivation for DSM within the power sector is that the most value effectiveness to cut back the electricity demand instead of increasing the ability offer generation or transmission. Application of DSM benefits the customers to adopt electricity with more efficiency. The load profile of the consumer can be flattened that improves the efficiency of various end-uses.

2. Demand Side Management

Deregulation in the world's energy market is experiencing vast changes, based on the real time data of electrical consumption new avenue has been created. To ensure a balanced and reliable market the commitment for balance between electrical offer and demand is critical. To try and do this there are 2 ways that, initial is that the Supply-side Management that adds the accessibility once demand is high and therefore the alternative is DSM that curtails the system need as per the availability.

In offer Demand Side Management, it's impracticable to fulfil the demand because it takes a while for the unit to begin and meet the demand like a shot. So, to fulfil this growing demand the demand facet management will be enforced and additionally helps to economic balance. In DSM designing, implementation and watching of utility activities are provided to manage end users which will end in sizable changes within the utility's load form, i.e., modification within the time pattern and value of a utility's load.

For many years, Demand Response (DR) or Demand-Side Management (DSM) is being thought-about because the "Holy Grail" of competent power generation. For the classical orienting downside DSM is essentially recognized because the fruitful answer, those facts this power system will be designed to carry peak however not the typical demand.

Generous quantity of the system capability is wasted once demand fluctuates considerably over a period. the most effective aim of DSM is thus to interrupt the burden interval by victimization "splintering the crests" and "filling the cribs" or, in numerous words: to transfer as Associate in Nursing awful ton of the bendy demand as

possible far from peak time into intervals of lower pursuit. A dependable and resourceful approach would make rational for existing generation substructure without ill effect on load side. The system improves with weather parameter.

The conventional thanks to perform DSM is via rate incentives, i.e., by means that of lowering tariffs from time to time whereas the aggregative necessitate is expected to be at a lower place common, therefore on inspire the stop person to shift versatile lots nearer to those periods. thanks to generation barriers and absence of automation, such laws have up to currently been terribly coarse-grained and regularly static (e.g., "day-charge" vs. "night-charge"), as a minimum within the customer market the client marketplace.

3. Teaching Learning Based Optimization

The TLBO technique is predicated on the impact of the sway of a tutor on the product of aspirants in a very category. Here, output is taken into account in terms of output. The tutor is usually thought of as an extremely student World Health Organization (WHO) shares the data with the beginners. the standard of a tutor affects the result of aspirants. It is evident that a good tutor or mentor put hard efforts to improve the results. Moreover, beginners learn with collaboration more, which additionally improve the output.

Basic

Assume 2 completely different tutors, *T1* and *T2*, tutoring a theme with same curriculum to an equivalent benefit extent aspirant in 2 completely unlike categories.

Fig. 1 reveals the nature of marks scored by aspirant. Curves 1 and 2 represent the output by *T1 & T2* tutors. With observation skewness nature has been showed.
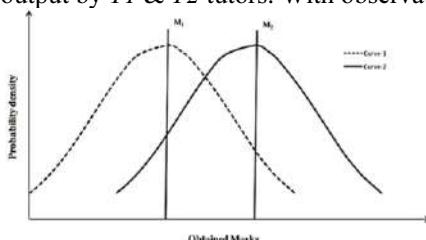
| **Figure 1.** Distribution of grades obtained by the aspirants taught by of two different tutors | **Figure 2**. Model for obtained marks distribution for a group of aspirants. |
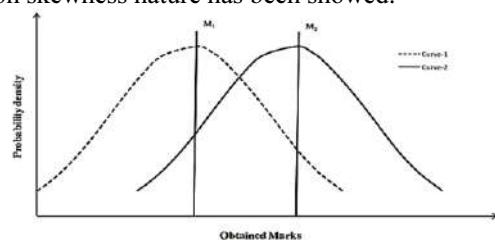
It has been observed from Figure. 1 that curve-2 has more improved output than curve-1 and so it proves that tutor *T2* is better than tutor T1 in the form of coaching. The point which differs the output is mean (Curve-2 has *M2* and curve- 1 has *M1*), i.e., worthy tutor makes more impact when collaborative teaching approach is defined.

From figure. 2 represent that A-curve having mean M_A. Tutor is taken into account as a learned human and best candidate is mirrored like tutor, and this can be shown by tantalum the tutor efforts to unfold data among the aspirants which is able to successively increase the data level of the total group and facilitate aspirants to urge smart marks or grades. Therefore, a tutor will improve the mean of the category per his or her potential.

In tutor to move mean towards their own level in step with his or her potential, thereby increasing the aspirants' level to a brand-new *M2* mean. *T1* Tutor can place most effort into educating his or her students; however, students can gain data in step with the standard of coaching delivered by an educator and also the quality of aspirant's gift within the category. The standard of the aspirants is arbitrated from the average of all. Tutor *T1* work hard for improvement in the standard of the aspirants from *M1* to *M2*, at that stage the aspirants need a replacement tutor   . Hence, the choice will be tutor *T2*.

Like different nature-inspired algorithms, TLBO is additionally populations primarily build technique that uses a population of answers to proceed to the world solution. As per TLBO algorithm is concerned, the population is taken into account as a bunch of aspirants or a category of aspirants.

In improvement procedures, the population be made up of varied vogue parameter. In TLBO, altogether totally different vogue variables square measure similar to different cases proposed and among which 'fitness' act as novel technique. The tutor is taken into consideration as a result of the most effective resolution as per the condition. Reveals the nature of marks scored by aspirant. Curves 1 and 2 represent the output by *T1* & *T2* tutors. With observation skewness nature has been showed.

The tutor is taken into account because the best answer obtained up to now. TLBO flow chart is as shown in Figure 3.
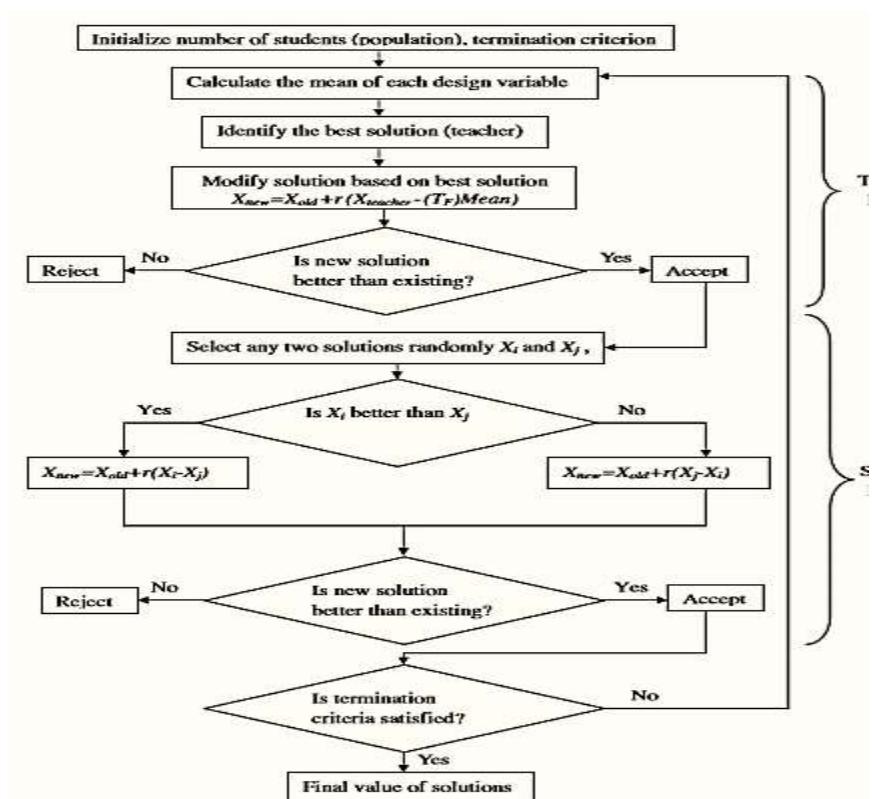
Figure 3. TLBO Flowchart

Tutor Phase

The tutor hard efforts result into improvement of M_1 mean (i.e. T_A) (to his or her level. However, a lot of its unfeasible and a coach can move the mean of the class house money supply to the opposite value cash in hand that's best than money supply relying on his or her potential of the class. This follows a random technique relying on varied factors.

Suppose M_j be the mean and T_i be the tutor at any i iteration. As per eqn.1, T_i selected as M_{new} and also the difference between the prevailing mean and new mean.

$$Difference\_Mean_j = r_i(M_{new} - T_F M_j) \qquad (1)$$

where, represent teach factor clarify the mean change, and is the random number in the range [0, 1]

Value of is either one or a pair of that may be a heuristic step or it's determined indiscriminately with equal likelihood as shown in eq. (2)

$$T_F = round[1+rand(0,1)\{2-1\}] \qquad (2)$$

The coaching issue is generated randomly as per eqn. 1-2, within which one corresponds to data level increment and a represent pass on of data. The in between values indicate quantity of transfer level of data. Basis on aspirant ability the transfer level of data is define. As per equation 2, one or two reckoning data is recommended. Consideration of 1-2 any price of T_F for *Difference_Mean*, the present resolution is updated as per equation (3)

$$X_{new,i} = X_{old,i} + Difference\_Mean_i \qquad (3)$$

Aspirants Phase

As per eqn. (3), the aspirant enhances their data by mutually interaction. As per system demand the interaction prevails cluster interaction for better output among aspirants. Additional data help for new learning.

Mathematically, it can be expressed as,

For i = 1: Pn

Randomly select another aspirant *Xj*, such that

$i \neq j$

if $f(X_i) < f(X_j)$

$X_{new,i} = X_{old,i} + r_i(X_i - X_j)$

else

X_{new}

$i = X_{old,i} + r_i (X_j - X_i)$

end if

end for

Accept X_{new} if it gives a better function value.

4. Problem Formulation

In this analysis work we tend to square measure reshaping the particular load curve just about objective load curve by minimizing peak demand supported the target the target as below.

A. *Objective function*

$load_i$ be the Initialization of load or load curve with forecasting. The advantage of this curve is depending upon the system the curve parameter will change. Implementation of DSM occurs when $load_f$ will be modified.

$$\text{Minimize: } f_1: max(load_f)$$

Consider *max load_f* as load curve peak which will decrease within system.
For Optimization of final loaf curve function $load_f$ is created.

B. *Constraints*

Considering subsequent constrains for minimization problem in which area unit will be positive in all time instant for number of devices of any sort.

$$X_{kh} > 0 \qquad\qquad \forall\, k,h$$

5. Results and Discussion

The parameter values of TLBO improvement are listed in Table 1. In TLBO aspirant count (20) will be proportion to size of population, Subject count of 3 will be proportion to style parameter or variable, improvement ways of iterative variety will be 35.

Table 1. Parameter values for TLBO with DSM Technique

Parameter Value	
Parameter	Value
Ending Criterion	35
Subjects Count	3
Aspirants Count	20
Rand	0 to 1

The hourly forecasted load for the commercial area is shown in Table 2. The hourly Wholesale cost of electricity in cents/kWh. The time take daily based in hours and accordingly wholesale price is mentioned with respect to time. As shown in Table 3. Hourly forecasted load is given with respect to hours.

Table 2. Forecasted Load Demand and Wholesale Energy Costs of Commercial Load

Hour	Wholesale Price	Hourly Forecasted Load	Hour	Wholesale Price	Hourly Forecasted Load	Hour	Wholesale Price	Hourly Forecasted Load
1	12	923.5	9	16.42	1558.4	17	8.65	404
2	9.19	1154.4	10	9.83	1673.9	18	8.11	375.2
3	12.27	1443	11	8.63	1818.2	19	8.25	375.2

4	20.69	1558.4	12	8.87	1500.7	20	8.1	404
5	26.82	1673.9	13	8.35	1298.7	21	8.14	432.9
6	27.35	1673.9	14	16.44	1096.7	22	8.13	432.9
7	13.81	1673.9	15	16.19	923.5	23	8.34	432.9
8	17.31	1587.3	16	8.87	577.2	24	9.35	663.8

Figure 4. Shows Forecasted Load Hourly and Objective curve Load Hourly without DSM. Hourly forecasted load is at peak . The aim is to minimize hourly forecasted load to hourly objective load curve. This result obtained is without implementing DSM.
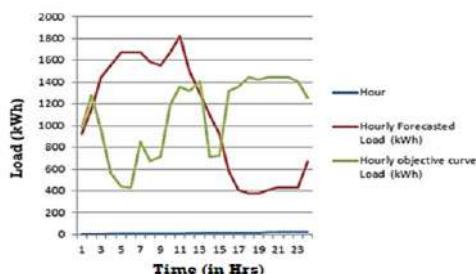
Figure 4. Forecasted Load Hourly and Objective curve Load Hourly without DSM
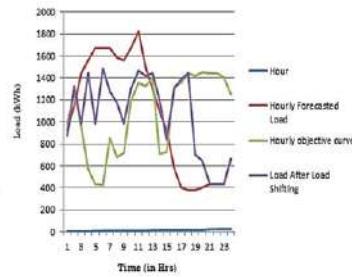
Figure 5. Forecasted Load Hourly and Objective curve Load Hourly with DSM

After implementing DSM, hourly forecasted load curve is altered as shown in Figure 5. There is a decrement of 353 kW considering difference of first & second row.

Table 3. DSM Strategy

Parameter Name	Value
Peak Load Without DSM	1818.2
Peak Load With DSM	1465.2
Peak Reduction	353
Percentage Reduction	19.41480585

6. Conclusion

Due to Demand Side Management the whole sensible smart grid, DSM is gaining a great deal. It decreases the excess power requirement during peak hours and also decreases the consumer utility bill. One of the unique features for future smart grid has been successfully implemented in this paper. For different sorts of commercial load TLBO methodology has an outsized variety significance for shifting of load. TLBO method can be used for residential and industrial consumers too.

It has been mathematically developed as a minimization downside. Mathematical approach could be considered as a minimization downside. TLBO algorithm has start

off with minimization of peak demand and monetary benefits in utility consumption. The experimentation has done on commercial area load for smart grid.

References

[1]. Thillainathan Logenthiran, Dipti Srinivasan, and Tan Zong Shun, "Demand side management in smart grid using heuristic optimization," ‖ IEEE transactions on smart grid, Vol. 3 Issue 3 pp. 1244–1252, 2012.

[2]. Ishan Gupta, G.N. Anandini, Megha Gupta "An Hour wise device scheduling approach for Demand Side Management in Smart Grid using Particle Swarm Optimization," IEEE Transaction, pp. 4799-5141, 2016.

[3]. Dhanshri A. Narayane, Prof. Kishor C. Muley, "Statistical Study of Microgrid Based on Object Oriented Usability Indices," IJRECE Vol. 7 Issue 1 ISSN: 2348-2281 pp. 2350-2353, 2019.

[4]. Oprea, S.V. Bâra, A.; Ifrim, G., "Flattening the electricity consumption peak and reducing the electricity payment for residential consumers in the context of smart grid by means of shifting optimization algorithm," Vol. 122, pp. 125-139, 2018.

[5]. Lizondo, D.; Rodriguez, S.; Will, A.; Jimenez, V.; Gotay, J., "An Artificial Immune Network for Distributed Demand-Side Management in Smart Grids," Inf. Sci., Vol. 438, pp. 32–45, 2018.

[6]. Hussein Jumma Jabir, Jiashen the "Impacts of Demand-Side Management on Electrical Power Systems: A Review", Energies, Issue 11, Vol. 1050, 2018.

[7]. Q. Li and M. Zhou, "The future-oriented grid-smart grid," J. Comput., Vol. 6, Issue 1, pp. 98–105, 2011.

[8]. Rao, R.V., Savsani, V.J. & Vakharia, D.P, "Teaching-learning-based optimization: A novel optimization method for continuous non-linear large-scale problems," Information Sciences, Vol. 183, Issue 1, pp. 1-15, 2012.

[9]. Di Santo, K.G.; Di Santo, S.G.; Monaro, R.M.; Saidel, M.A., "Active demand side management for households in smart grids using optimization and artificial intelligence," Measurement, Vol. 115, pp. 152–161, 2018.

[10]. Rao RV, Patel V, "An elitist teaching-learning-based optimization algorithm for solving complex constrained optimization problems," Int J Ind Eng Comput Vol. 3, pp. 535–560, 2012.

[11]. Rao, R.V.; Savsani, V.J.; Vakharia, D.P., "Teaching-learning based optimization: A novel method for constrained mechanical design optimization problems," Comput.-Aided Des., Vol. 43, pp. 303–315, 2011.

[12]. Rao RV, Patel V, "An improved teaching-learning-based optimization algorithm for solving unconstrained optimization problems," Sci Iranica D Vol. 20, Issue 3, pp. 710–720, 2012.

[13]. Khan, B.; Singh, P. Selecting a Meta-Heuristic Technique for Smart Micro-Grid Optimization Problem: A Comprehensive Analysis. IEEE Access, Vol. 5, pp. 13951–13977, 2017.

Subject Index

Author Index

This page intentionally left blank