Advances in Intelligent Systems and Computing 1141

Aboul Ella Hassanien
Roheet Bhatnagar
Ashraf Darwish *Editors*

Advanced Machine Learning Technologies and Applications

Proceedings of AMLTA 2020

Springer

Advances in Intelligent Systems and Computing

Volume 1141

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within "Advances in Intelligent Systems and Computing" are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink ****

More information about this series at http://www.springer.com/series/11156

Aboul Ella Hassanien · Roheet Bhatnagar ·
Ashraf Darwish
Editors

Advanced Machine Learning Technologies and Applications

Proceedings of AMLTA 2020

Springer

Editors
Aboul Ella Hassanien
Faculty of Computer and Information
Cairo University
Giza, Egypt

Ashraf Darwish
Faculty of Science
Helwan University
Helwan, Egypt

Roheet Bhatnagar
Department of Computer Science
and Engineering
Manipal University Jaipur
Jaipur, Rajasthan, India

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-981-15-3382-2 ISBN 978-981-15-3383-9 (eBook)
https://doi.org/10.1007/978-981-15-3383-9

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Aboul Ella Hassanien dedicated this work to the scientific research group members in Egypt and his wife Nazaha Hassan El Saman.

Professor Roheet dedicated the proceedings to his parents, his wife, and his son Rishaan.

AMLTA 2020—Conference Committees

Chief Patron

K. Ramnarayan, Chairperson, Manipal University Jaipur, India

Patron

Gopalkrishna Prabhu, President, Manipal University Jaipur, India

Co-patrons

Niti Nipun Sharma, Pro-president, Manipal University Jaipur, India
H. Ravishankar Kamath, Registrar, Manipal University Jaipur, India
Jagannath Korody, Dean (FoE), Manipal University Jaipur, India

Honorary Chairs

Ramjee Prasad, Aarhus University, Denmark
Mohamed F. Tolba, Cairo, Egypt

General Chairs

Aboul Ella Hassanien, Cairo University and SRGE, Egypt
Rajveer Singh Shekhawat, Manipal University Jaipur, India

Organizing Chair

Roheet Bhatnagar, Manipal University Jaipur, India

Program Chairs

Sandeep Joshi, India
Vijay Pal Singh, Dhaka, India
Pankaj Vyas, India

Program Co-chairs

Devesh Kumar Srivastava, India
Uma Shankar Rawat, India

Co-chairs

Sandeep Chaurasia, India
Ashish Kumar, India
Vivek Kumar Verma, India

Publication Chairs

Vivek Kumar Verma, India
R. Jaya Krishna, India

Finance Chairs

Sandeep Chaurasia, India
Anita Shrotriya, India

Sponsorship Chair

C. S. Lamba, India

Local Organizing Chair

Venkatesh Gauri Shankar, India

Registration Chair

Akhilesh Kr. Sharma, India

Travel and Transport Chair

Priyank Singh Hada, India

Printing and Media Chair

Yogesh Gupta, India

Special Session and Workshop Chair

Dilbag Singh, India

Ph.D. Forum Chair

Pradeep Kumar Tiwari, India

Publicity Chairs

Hamidreza Boveiri, Iran
Tarun Jain, India

International Advisory Committee

Ajith Abraham, USA
P. K. Mahanti, Canada
Lipo Wang, Singapore
Nitin Kumar Tripathi, Thailand
Suphamit Chittayasothorn, Thailand
Allam Hamdan, Oman
Peter Damoah Afari, Hong Kong
Mridul Malayanil, USA
V. G. Tharinda Nishantha Vidanagama, Sri Lanka
Shalabh Bhatnagar, India
Kumkum Garg, India
Mrinal Kanti Ghose, India
Vandana Bhattacharjee, India
E. Poovammal, India
Yasmeen Pervez, India

Technical Program Committee (TPC)

Edgard Marx, Germany
Islam Amin, Australia
Ivan Ermilov, Germany
Arvind Kilaru, Canada
Ashok Rajan, USA
Moheet Bhatnagar, USA
Shoji Tominaga, Japan
Arvind Kumar Pal, Japan
Takahiko Horiuchi, Japan

Seyedali Mirjalili, Australia
Fatos Xhafa, Spain
Vaclav Snasel, Czech Republic
Janusz Kacprzyk, Poland
Tai-hoon Kim, Korea
Evgenia Theodotou, Greece
Pavel Kromer, Czech Republic
Irma Aslanishvili, Czech Republic
Jan Platos, Czech Republic
Hani M. K. Mahdi, Egypt
Essam Elfakharanyb, Egypt
Ashraf AbdelRaouf, Egypt
Mohamed Abd Elfattah, Egypt
Tahani Alsubait, UK
Evgenia Theodotou, Greece
Natalia Spyropoulou, Greece
Dimitris Sedaris, Greece
Litsa Charitaki, Greece
Elena Amaricai, Greece
Manal Abdullah, KSA
Anandavalli M, KSA
Mohammad Reza Noruzi, Iran
P. K. Upadhyay, UAE
Abdelhameed Ibrahim, Egypt
Ahmad Taher Azar, Egypt
Nitul Dutta, India
Deepak Dembla, India
Sudipta Roy, India
Sonal Jain, India
Neeraj Bhargava, India
Pijush Barthakur, India
Ritu Bhargava, India
Babita Malik, India
Nitu Bhatnagar, India
Ram Naresh Saraswat, India
Pushpendra Kumar, India
Manish Thukral, India
Abhishek Srivastava, India
Swati Srivastava, India
Swagatam Das, India
Gaurav Prasad, India
Jaya Krishna, India
Harish Sharma, India
Gulrej Ahmed, India
Sandeep Joshi, India

Devi Prasad Sharma, India
Bhavna Saini, India
Ashish Jain, India
Nanda Gopal Reddy, India
Jeril Kuriakose, India
Puspendra Singh Sisodia, India
Ajay Prasad, India
Tipendra Pal Singh, India
D. K. Lobiyal, India
Kamakshi Prasad, India
Sunita Singhal, India
Vivekanand Tiwari, India
Tarun Kumar Dubey, India
Amit Rathi, India
Ankit Srivastava, India
Rohit Kumar Gupta, India
Rahul Saxena, India
Punit Sharma, India
Priyank Singh Hada, India
Prakash Ramani, India
Pratistha Mathur, India
Pallavi Bhatt, India
Monika Jain, India
Manoj Bohra, India
Shikha Mundra, India
Anubha Parashar, India
Vivek Singh Sikarwar, India
Rishi Gupta, India
Nitish Pradhan, India
Alka Choudhary, India
Giriraj Methi, India
Rajdeep Chowdhury, India

Preface

This volume constitutes the refereed proceedings of the 5th International Conference on Advanced Machine Learning Technologies and Applications, AMLTA 2020, held in Jaipur, a UNESCO World Heritage Site and capital city of state of Rajasthan, India, during February 13–15, 2020. In response to the call for papers for AMLTA 2020, 148 papers were submitted for presentation and inclusion in the proceedings of the conference. After a careful blind refereeing process, 65 papers were selected for inclusion in the conference proceedings. The papers were evaluated and ranked on the basis of their significance, novelty, and technical quality by at least two reviewers per paper. The papers cover current research in machine learning, data sciences, deep learning, text visualization, modeling optimization complex network, renewable energy, swarm intelligence, biomedical engineering, complex control systems, cyber-security, and data mining.

We express our sincere thanks to the plenary and tutorial speakers, workshop/special session chairs, and international program committee members for helping us to formulate a rich technical program. We would like to extend our sincere appreciation for the outstanding work contributed over many months by the Organizing Committee: local organization chair and publicity chair. We also wish to express our appreciation to the SRGE members for their assistance. We would like to emphasize that the success of AMLTA 2020 would not have been possible without the support of many committed volunteers who generously contributed their time, expertise, and resources toward making the conference an unqualified success. Finally, thanks to Springer team for their support in all stages of the production of the proceedings. We hope that you will enjoy the conference program.

Giza, Egypt — Aboul Ella Hassanien
Jaipur, India — Roheet Bhatnagar
Helwan, Egypt — Ashraf Darwish

Contents

About the Editors

Aboul Ella Hassanien is the Founder and Head of the Scientific Research Group in Egypt (SRGE) and a Professor of Information Technology at the Faculty of Computer and Information, Cairo University. Professor Hassanien has more than 1000 scientific research papers published in prestigious international journals to his credit.

Roheet Bhatnagar works at Manipal University Jaipur (MUJ), where he is currently serving as a Professor at the Department of Computer Science & Engineering. He has been affiliated with the MUJ since 2012, and with the Manipal group since 2008. With vast experience spanning more than 20 years in both industry and academia, he has worked for companies like Xerox Corporation, Samsung SDS and USHA Soft in various capacities. He is a Senior Member of the IEEE and Life Member of the ISTE and ISRS.

Ashraf Darwish is a Professor of Computer Science and former Acting Head of the Mathematics and Computer Science Department at Helwan University, Egypt. Dr Darwish has a wealth of academic experience and has authored numerous publications, including papers in international journals and conference proceedings. Currently, he is the vice chair of the Scientific Research Group in Egypt (SRGE).

Modelling and Optimization

Segregating and Recognizing Human Actions from Video Footages Using LRCN Technique

Meet Pandya, Abhishek Pillai, and Himanshu Rupani

Abstract Computer vision is a vast area of research that includes extracting useful information from images or sequence of images. Human activity recognition is one such field undergoing lots of research. The practical application for this model is vast in various kinds of researches as well as actual practice. This paper proposes a two-model approach using a combination of a convolutional neural network using transfer learning and a long short-term memory model. CNN network is applied to gather the feature vectors for each video, and the LSTM network is used to classify the video activity. Standard activities contain benchpress, horse riding, basketball dunk, etc. A high accuracy level of 94.2% was achieved by the proposed algorithm.

Keywords Human activity recognition · LRCN · UCF-101 · Computer vision · Video analysis · Image processing

1 Introduction

An action is a set of movement performed by a human body in a sequential manner. From a computer vision perspective, action recognition is a classification of a video activity according to the sequence of images which collaboratively form a video. Action recognition was a significant area of research in the last few years due to its enormous applications like human interaction, security, communications and entertainment. Unlike static images, video recognition includes spatial as well as temporal features to identify the task which gives us a more coherent understanding of the actions performed. Researchers tend to incline toward action recognition due

M. Pandya · A. Pillai (✉)
U.V. Patel College of Engineering, Mehsana, India
e-mail: abhishekpillai.ap@gmail.com

M. Pandya
e-mail: meetpandya4715@gmail.com

H. Rupani
L.D. College of Engineering, Ahmedabad, India
e-mail: himanshurupani97@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_1

to its vast domain of application in surveillance, real-time gaming, video description, human interaction [1], video captioning [1, 2] and robotics [3, 4]. Sequence learning is a machine learning algorithm in which data is not independent and makes use of contextual information [5, 6]. A convolutional neural network is a conventional model for image classification but for a video each and every frame plays a critical role in analysis.

Long short-term memory networks, commonly termed LSTM, are a certain form of RNN [7], competent of grasping long-term dependency. LSTM also follows a chain-like structure where each cell contains more than one neural network and is interconnected as well as interacting with each other [8]. LSTMs help in rectifying or minimizing the error that could backpropagate through time and layers [9]. By maintaining more constant error, they allow recurrent nets to continue to learn over many time steps [10].

The below paper describes a combined model approach which includes a convolutional neural network that gives the output feature vector for video frames which is then provided to the LSTM cell, and the softmax classifier is used for classification. As a human being recognizes an activity by viewing the series of action and analyzing the attributes for some specific period of time, in this same way the model described below understands and predicts the activity. The rest of the paper is organized as follows: Sect. 2 describes the related work done on the project, Sect. 3 describes the architecture of the model used, experiment result and comparison with other model are provided in Sect. 4, and Sect. 5 concludes the paper with future research directions.

2 Related Work

Activity recognition systems are novel and highly engaging field of research and development having varied implications and applications. Nowadays, action recognition is an exclusive topic in the ongoing research works. You can refer to this [11–13] for related surveys and questions. Basic idea was to classify every frame in video and output the majority vote obtained during this process.

Two-dimensional models are the conventional 2D CNN architecture which only uses spatial features for classification. The basic approach was to take one or more than one frame of video into consideration and apply a pre-trained model to classify the image. Transfer learning algorithm was used to apply a pre-trained model on ImageNet [14] dataset to classify images. The final label could be generated by averaging out all the RGB representation and then applying to three fully connected layers for classification; if not this approach then instead the results are gained by either applying a softmax classifier or max-pooling the spatial features [15, 16].

Another method that should be taken into consideration is using CNN to classify action in-depth data directly by modeling poses in a view-invariant high-dimensional space. A synthetic 3D pose dataset [17] is created, a model is trained on hundreds of poses to create depth maps, and then a dense neural layer can be adopted for final result. Two-dimensional models consider only spatial information, and no temporal information is considered which makes it less efficient.

Three-dimensional CNN model allows capturing both spatial information and temporal information while maintaining the temporal structure of the model. A fully automated model that has no prior knowledge is trained to learn the spatiotemporal features. A recurrent neural network is trained to classify each sequence of features. In 3D convolution, the kernel used to perform convolution has an additional dimension, usually time [18, 19].

Two-stream network uses base two-stream architecture. It is a fusion of both spatial and temporal streams [20], and the spatial net can capture spatial dependency between frames while a temporal net can capture the presence of periodic motion for each spatial location in a video. These networks need to be fused at an early level such that responses for the same pixel position are put in correspondence. Different spatial fusion strategies were used like sum fusion, max fusion, concatenation fusion, conv fusion, etc., to obtain the desired results [21–24]. Two-stream 3D ConvNet architecture is created by inflating the filters and pooling kernels in order to add an extra-temporal dimension. Filters are converted from $N \times N$ to $N \times N \times N$. Also, the model takes advantage of the pre-trained model weights trained on the ImageNet dataset. The 3D model can then be explicitly trained by ImageNet dataset as pooled activation of video would be same as single static image; hence, the weights are repeated N times along the time dimension, then dividing these weights by N. UCF-101 dataset has the highest accuracy of 98% in the I3D model [25].

3 Model and Architecture

We use LRCN for action recognition challenges, which involves spatial and temporal learning. LRCN is a class of models that combines the spatial and temporal feature learning into a single model. Spatial features are extracted by an extended version of CNN, in which the Inception V3 model is applied using transfer learning, whereas temporal features are extracted by a state-of-the-art recurrent model, LSTM.

3.1 Convolutional Neural Network

Convolutional neural networks (ConvNets) are constructed of stacked feature learning stages. Each stage consists of convolution, pooling and normalization modules [8]. The benefit of CNN is its ability to learn the position and scale-invariant structures in the data, which is important when working with images. ILSVRC 2012 Alexnet introduced a CNN model for image classification which made a major breakthrough in the field of computer vision, making processes like image classification more accurate and efficient [26].

3.2 Transfer Learning

This experiment is done on a presumption that the knowledge gained from one task can be transferred to learn another task. The basic analogy that can be used here is that knowledge gained through learning what is a dot can be used to learn how to draw a line, which in turn is used to draw a rectangle, which is then further stretched to the concept of solid objects. Here, we have used the knowledge gained from learning how to recognize objects in the images in order to learn how to recognize activities in videos. Transfer learning is method in which we leverage the knowledge from a pretrained model and use it to solve a different but related problem [27].

3.3 Inception ConvNet Model

Inception is a range of models designed and trained by Google for classifying objects in the given image, using the ImageNet database. Inception V1 (GoogLeNet) architecture was introduced by Google in the ILSVRC 2014 challenge. The model has an error rate of 6.67% on the ImageNet dataset [28]. Inception model has increased depths and a new level of organization in the form of the inception module. The inception module consists of 3 different kernel sizes rather than 1. It uses 3×3, 5×5 and 1×1 filters, and max pooling function is also performed as shown in Fig. 1. The outputs from each convolution will be concatenated and sent to the next inception module. 5×5 and 3×3 convolutions are computationally very expensive. Thus to make the computational process feasible, dimensionality reduction was applied onto the image before it underwent through convolution. A 1×1 filter-sized CNN was introduced between 3×3 and 5×5 CNNs. GoogLeNet has nine inception modules

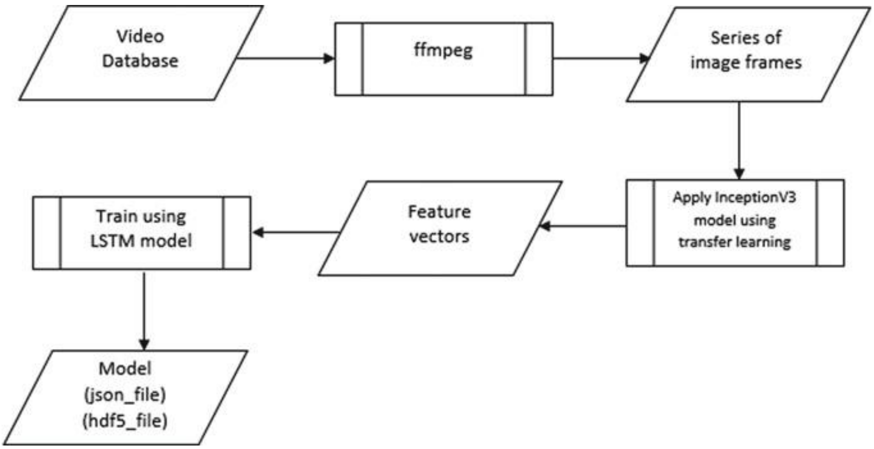


Fig. 1 Data flow diagram

stacked linearly which leads to vanishing gradient problem. To overcome that, two auxiliary classifiers were added which applies softmax on the outputs of inception modules and computes auxiliary loss.

These losses are added to total loss, so total loss is sum of real loss and two auxiliary losses.

$$\mathrm{Loss}_{\mathrm{total}} = \mathrm{Loss}_{\mathrm{real}} + 0.3 * \mathrm{loss1}_{\mathrm{aux}} + 0.3 * \mathrm{loss2}_{\mathrm{aux}} \tag{1}$$

0.3 value was described in the paper [28].

Inception V2 and V3 architectures were introduced back in 2015 with modifications to overcome representational bottleneck. The model has an error rate of 3.5% on the ImageNet dataset. Various principles were considered to improve accuracy and reduce computational burdensomeness more [29]. To eliminate bottleneck representation, an efficient grid size reduction strategy was used which consists of two parallel stride pooling and convolution which were concatenated before going into the next module. A 5×5 filter was factorized into two 3×3 filters, which proved to be less expensive, described in Fig. 2. More the factorization, cheaper the model would be; hence, asymmetric factorization was introduced. A 3×3 filter was factorized into 1×3 and 3×1 filters which proved to be 33% cheaper than usual as shown in Fig. 3. All the figures describe the three important modules of inception [30]. Inception V3 includes all the upgrades of V2 and in addition an RMS optimizer, batch normalization in the auxiliary classifier and label smoothing [29, 30].

Fig. 2 Inception block A

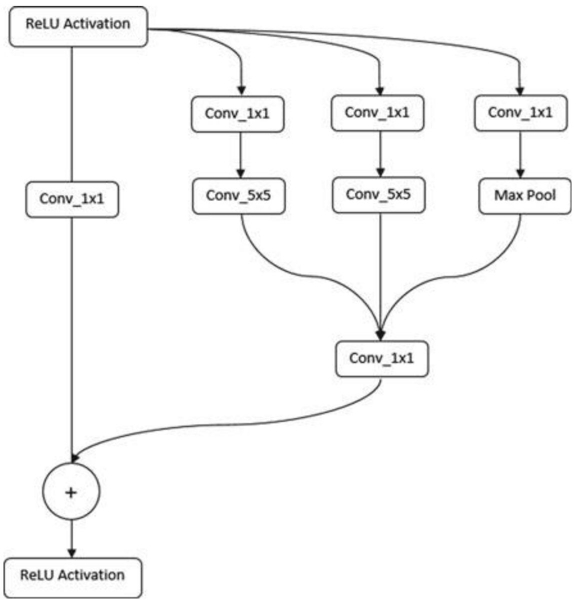
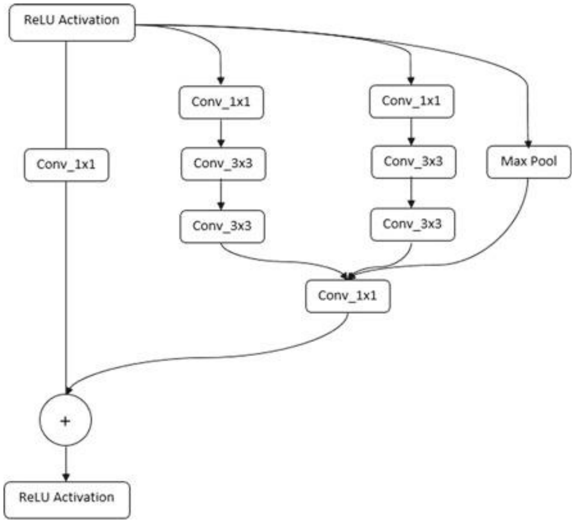


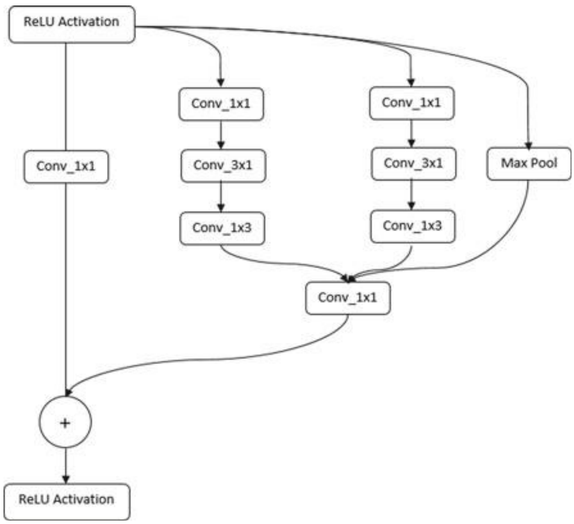
Fig. 3 Inception block B



The flow of images in Inception: Image is passed through 3 simple 3×3 CNN filters, and then it is introduced to the Inception module. There are three inception modules as shown in the above figures. Images are introduced to the module alphabetically.

Inception is known to be highly accurate for image recognition. In this model, modified version of the Inception V3 model is used in such a way that the second last layer of the model is taken as an output instead of the last layer. In other words, this model learns only up to high-level features instead of classifiers (Fig. 4).

Fig. 4 Inception block C



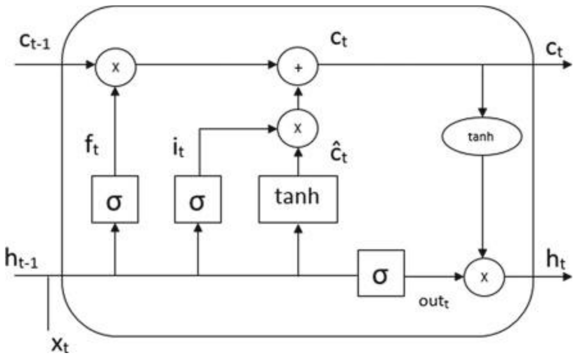


Fig. 5 LSTM neural cell

3.4 Long Short-Term Memory

Simply put, where RNN can remember only the recent past, LSTM can find which instances of the past moments are important and which are not, thus capitalizing on only that information which proves to be helpful [7]. LSTM comprises a cell state and its various gates. The cell state is the most important part of an LSTM cell. It transfers information from one time step to another and acts as a conveyor belt for important information learned from the previous inputs. The gates are the components that help the cell state to update in each time step and in the calculation of hidden states as shown in Fig. 5.

The forget gate decides which information is useful and which is not. Input from the previous hidden state and the current input is concatenated and then passed through a sigmoid function where value comes out between 0 and 1, 1 being most useful and 0 not useful or forgettable.

$$f_t = \sigma(W_f \cdot ([h_{t-1}, x_t]) + \text{bias}_f) \tag{2}$$

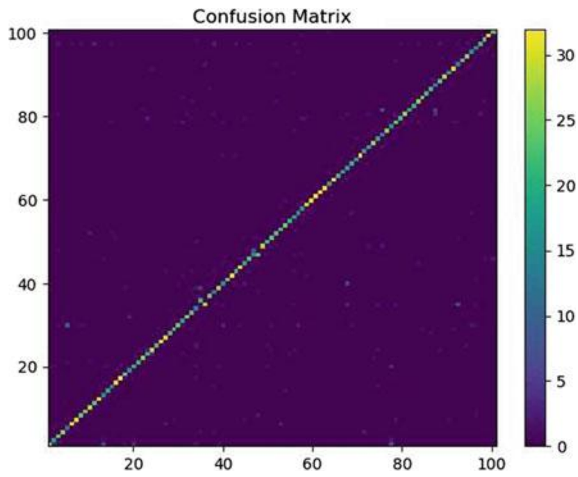
Input gate is used to update the cell state by using the current input, and tanh function is applied to the concatenated output of hidden state and input, which squishes the value between 1 and −1. A sigmoid operation is also performed on the same input to regulate the network and then performing pointwise multiplication on the output of the tanh and sigmoid functions.

$$i_t = \sigma(W_i \cdot ([h_{t-1}, x_t]) + \text{bias}_i) \tag{3}$$

Cell state updation contains two operations; first, pointwise multiplication of cell state with the forget vector for dropping values near zero, in order to discard the useless values, and second, performing pointwise addition of cell state and input vector that updates the value of cell state which the neural network finds relevant.

$$\hat{C}_t = \tanh(W_c \cdot ([h_{t-1}, x_t]) + \text{bias}_c) \tag{4}$$

Fig. 6 Confusion matrix



$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \tag{5}$$

The output gate is used to compute hidden states for the next input. A pointwise multiplication is performed between the concatenated results of the input vector and hidden state; furthermore, the result is evaluated using a sigmoid function and the output of the tanh function is applied on the newly updated cell state. The hidden state calculated at the end is a product of output gate and current cell state provided that the values of cell state have been rescaled by tanh function.

$$\text{out}_t = \sigma(W_o \cdot ([h_{t-1}, x_t]) + \text{bias}_o) \tag{6}$$

$$h_t = \text{out}_t * \tanh(C_t) \tag{7}$$

To summarize, a forget gate decides which information is useful, an input gate decides which information should be added to the current cell state and the output gate decides what the next hidden state should be (Figs. 6 and 7).

4 Experimentation

This section provides us with the results and benchmarks of the proposed model for the UCF-101 dataset [31]. The dataset is divided into two parts, training set (80%) and testing set (20%). Further, the training dataset is divided into two parts, training set (60%) and validation set (20%).

Before passing each frame of video through a deep learning model, we convert each image into a NumPy array. Here, the image has shape attributes, i.e., width,

Fig. 7 Output example

height, channel, etc. It is henceforth reshaped into an array of shapes, constituting elements like height, width, shape. After that, we expand the dimension of the array by 1°. Then, it is converted from RGB to BGR format and zero-center by mean pixel.

Instead of using the whole video for analysis, we use only 40 equally distributed frames of each video.

This model works by passing each frame of the video, respectively, through a modified Inception V3 model which gives a 2048 length feature vector. This vector contains high-level spatial features of that image. These features are then passed to the LSTM layer which generates a spatiotemporal feature vector. This processed final feature vector is then given to fully connected layers, respectively, starting from a dense layer which constitutes of 512 neurons. The dropout rate is set to 0.2, for preventing overfitting of the data. The last layer is a dense layer of 101 neurons which then results in the final classification required. Before passing the feature vector into LSTM as an input, we reshape it from batch size, seq len to batch size, time steps, seq len. The dataset used for training is UCF-101, and each activity has 100–200 videos for training. It took around 65 min to train the model on the Intel HM370 chipset with 8 GB RAM and NVIDIA Geforce GTX 1050Ti graphics processor (Table 1).

Table 1 Model comparison

Different models	Accuracy (UCF-101) (%)
PoTion: pose motion representation [32]	98.2
Spatiotemporal convolution	97
Temporal 3D ConvNets	92.3
CNN + LSTM model (our model)	94.2

5 Conclusion

In this paper, we suggest a combined approach of convolution neural network and long short-term memory network to identify action from video dataset.

First, the videos are converted into images, from that 40 equally distributed frames are selected. Those frames are used to create a 2048×40 feature vector by performing transfer learning on Inception V3, the model introduced by Google. The 40 vectors are then applied in sequence to an LSTM layer, different no. of layers were experimented, and at last a single LSTM layer was proven to be precise as well as efficient. The proposed model has been tested on the UCF-101 dataset which gained an accuracy of 94.3% on 70–30 data split.

6 Future Work

This method uses all the spatial and temporal features of a given video. In the future, we aim to perform more parameter alteration to increase accuracy and find a way to make model understand between similar types of activities like jumping jack and skipping.

References

1. Tavakkoli, A., Kelley, R., King, C., Nicolescu, M., Nicolescu, M., Bebis, G.: A visual tracking framework for intent recognition in videos
2. Wu, Z., Yao, T., Fu, Y., Jiang, Y.-G.: Deep learning for video classification and captioning (February 2018)
3. Sunny, J.T., George, S.M., Kizhakkethottam, J.J.: Applications and challenges of human activity recognition using sensors in a smart environment (September 2015)
4. Ranasinghe, S., Machot, F.A., Mayr, H.C.: A review on applications of activity recognition systems with regard to performance and evaluation (August 2016)
5. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks (September 2014)
6. A must-read introduction to sequence modelling (with use cases), Analytics Vidhya. https://www.analyticsvidhya.com/blog/2018/04/sequence-modelling-an-introduction-with-practical-use-cases/
7. Sherstinsky, A.: Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network (August 2018)
8. Hochreiter, S., Schmidhuber, J.: Long short-term memory (1997)
9. Lipton, Z.C., Berkowitz, J.: A critical review of recurrent neural networks for sequence learning (June 2015)
10. Varol, G., Laptev, I., Schmid, C.: Long-term temporal convolutions for action recognition (2016)
11. Zhang, H.-B., Zhang, Y.-X., Zhong, B., Lei, Q., Yang, L., Du, J.-X., Chen, D.-S.: A comprehensive survey of vision-based human action recognition methods (February 2019)
12. Thankaraj, S., Gopalan, C.: A survey on human activity recognition from videos (February 2016)

13. Asadi-Aghbolaghi, M., Clapes, A., Bellantonio, M., Es-calante, H., Ponce-Lpez, V., Bar, X., Guyon, I., Kasaei, S., Escalera, S.: Deep learning for action and gesture recognition in image sequences: a survey (January 2018)
14. Krizhevsky, A., Sutskever, I.: ImageNet classification with deep convolutional neural network
15. Zha, S., Luisier, F., Andrews, W., Srivastava, N., Salakhutdinov, R.: Exploiting image-trained CNN architectures for unconstrained video classification (May 2015)
16. Wang, L., Xiong, Y., Wang, Z., Qiao, Y., Lin, D., Tang, X., Van Gool, L.: Temporal segment networks: towards good practices for deep action recognition (August 2016)
17. Zolfaghari, M., Oliveira, G.L., Sedaghat, N., Brox, T.: Chained multi-stream networks exploiting pose, motion, and appearance for action classification and detection (2017)
18. Ji, S., Xu, W., Yang, M.W., Yu, K.: 3D convolutional neural networks for human action recognition (2010)
19. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3d convolutional networks. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 44894497. IEEE (2015)
20. Ma, C.-Y., Chen, M.-H., Kira, Z., AlRegib, G.: TS-LSTM and temporal-inception: exploiting spatiotemporal dynamics for activity recognition (March 2017)
21. Simonyan, K., Zisserman, A.: Two-stream convolutional networks for action recognition in videos
22. Feichtenhofer, C., Pinz, A., Zisserman, A.: Convolutional two-stream network fusion for video action recognition (April 2016)
23. Gammulle, H., Denman, S., Sridharan, S., Fookes, C.: Two stream LSTM: a deep fusion framework for human action recognition (2017)
24. Feichtenhofer, C., Pinz, A., Zisserman, A.: Convolutional two-stream network fusion for video action recognition (September 2016)
25. Diba, A., Fayyaz, M., Sharma, V., Karami, A.H., Arzani, M.M., Yousefzadeh, R., Van Gool, L.: Temporal 3D convnets: new architecture and transfer learning for video classification (November 2017)
26. O'Shea, K.T., Nash, R.: An introduction to convolutional neural networks
27. Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C.: A survey on deep transfer learning (August 2018)
28. Szegedy, C., Liu, W., Jia, Y.: Going deeper with convolutions (September 2017)
29. Szegedy, C., Vanhouck, V., Ioffe, S., Shlens, J.: Rethinking the inception architecture for computer vision
30. A simple guide to the versions of the inception network, towards data science. https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202
31. UCF101—Action recognition data set, University of Central Florida. https://www.crcv.ucf.edu/data/UCF101.php
32. Choutas, V., Weinzaepfel, P., Revaud, J., Schmid, C.: PoTion: pose moTion representation for action recognition. In: CVPR 2018 - IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, United States, pp. 7024–7033 (Jun 2018). https://doi.org/10.1109/CVPR.2018.00734. hal-01764222

Fall Alert: A Novel Approach to Detect Fall Using Base as a YOLO Object Detection

Kiran P. Kamble, Swapnil S. Sontakke, Himanshu Donadkar, Ritik Poshattiwar, and Abhinavram Ananth

Abstract In this work, a novel approach to detect irrecoverable fall of a person especially during night by leveraging convolutional neural network (CNN)-based object recognition and image processing techniques are proposed. YOLO (You Only Look Once) object detection model is trained on the COCO dataset to detect the desired object from input frames extracted from a video stream and is tested whether it has succumbed to an irrecoverable fall or not for fall over multiple frames. The paper begins by introducing the core idea behind proposed approach and its importance followed by a review of previously done work in a nutshell. Then, the proposed approach which gave appreciable results is presented. Accuracy of the proposed approach is found to be 93.16%. The work also shows an experimental ratio of the height of camera to the distance of the person from the camera. It is found to be 5:6 for its successful fall detection that gives a competitive performance as compared to other state-of-the-art approaches.

Keywords Fall detection · YOLO · COCO · Convolutional neural network (CNN) · Image processing

1 Introduction

Image processing and deep learning techniques have proven to obtain significant results to various problems in domains like retail, security, health care, banking, and industrial automation. Recognition of activities of daily living (ADL) has come to prominence that can be solved using image processing and deep learning. By tracking various ADL actions such as standing, hand movements, sitting, walking, running, and falling, we can create solutions for these use-cases. We focused on

K. P. Kamble · S. S. Sontakke (✉) · H. Donadkar · R. Poshattiwar · A. Ananth
Department of Computer Science and Engineering, Walchand College of Engineering, Sangli, India
e-mail: sontakke.swapnil31@gmail.com

K. P. Kamble
e-mail: kirankamble5065@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_2

the fall detection of a person. Such a use-case is very beneficial for various allies, and one such is being personnel in army and security forces. Armed forces are at continuous vigil at important and sensitive locations such as military bases, secret safe-houses, and hideouts. At such locations, during vigilance especially during night time, when illumination is relatively low, the enemy can attack the personnel in silence and stealth. This triggers a need to surveil whether the security personnel have come across an irrecoverable fall from such attacks or medical emergency and take subsequent measures. Thus, our objective is to detect and alert irrecoverable fall of the security personnel at night. This will help the injured personnel to get the medical aid as early as possible. In case of adversary attack, this will alert armed forces to call for expeditious actions against that attack.

Additionally, keeping the above discussion in mind, it is, hence, essential to continue advancing in the field of image processing and deep learning so as to find novel and as accurate as possible solutions to problems which were earlier not solvable using traditional methodologies.

2 Literature Review

As of now, there has been ample amount of research work done to detect the fall of desired object. Fall detection activities proposed in recent years are based on three classes, namely vision based, sensor based, and sound based. Two kinds of devices are used by these systems for fall detection and alert. Those are wearable and non-wearable devices. Sensors such as accelerometer and gyroscopes are wearable sensors carried by human body to detect the fall automatically and differentiate between normal ADLs and accidental fall by setting some threshold [1]. Su et al. [2] label systems that integrate several sensors composed of nodes carried on the body parts such as waist and thighs to detect the fall. One of the simplest wearable devices is a "push-button" which can be activated physically by falling person when fall is detected. These kinds of devices work well in indoor as well as outdoor environments [3]. However, the downside of those systems is that aged person may forget to carry those wearable devices all the day or even feel awkwardness by wearing it. Non-wearable sensors include video cameras, infrared sensors, floor vibration sensors, and sound sensors. Out of which floor vibration and sound sensor systems are based on signal processing and pattern-recognition algorithms. They use features like shock-response spectrum and mel-frequency cepstral coefficients to classify fall and non-fall activities [4]. To detect the fall and non-fall using acoustic sound signals as an input, Li et al. [5] use microphone array sensors. Liu et al. [6] show that radar systems are also helpful for reliable detection of fall. Doppler signatures and features extracted from the input frames help in vigilance the real-time fall. Tzeng et al. [7] stated, environmental conditions may affect the floor pressure and infrared and sometimes the sensors may break down. To deal with these limitations, different computer vision strategies are extensively used. One of the important benefits of following computer vision approaches is no necessity of wearing the devices on body. Different types of cameras help to capture scene in the visible range. Numerous classification techniques like artificial neural network are used to develop customized networks as per need to detect the fall of the person [8]. Foroughi et al. [9] demonstrate Eigen space-based human fall detection system that follows the approach of multi-class support vector machine to precisely classify motions and detect the fall. Rougier [10] focuses on Gaussian mixture model (GMM) that uses shape deformation feature to classify the fall and non-fall activity. Thome et al. [11] propose a layered hidden markov model (LHMM) based multi-view fall detection system. The system implemented motion analysis and solved the problem of intrusion.

Earlier machine learning practices were being used which mostly used input data from Microsoft's Kinect sensor which generate depth images. Other type of methodology involved in the detection of fall is by wearable sensor devices on human body. Nizam et al. [12] took advantage of Kinect infrared device for acquiring depth images at a frame rate of 30 FPS with Microsoft SDK v1.7. The contribution of this paper is that it uses floor plane equation and the joint coordinates. These coordinates are extracted from the skeleton data outputted by the Kinect runtime. Generated data is used to figure out the velocity and acceleration of the human body, the distance from the head to floor plane, and the position of the other joints. These measurements contribute in identifying the accidental fall. The results obtained from experiment depict that the accuracy of the approach is of 94.43%. With this accuracy, it can accurately distinguish the fall movement from other daily activities. Nizam et al. [13] use depth information from Microsoft Kinect Sensor in the fall detection algorithm to compute velocity and position of the subject. Each consecutive frame is contributed in computing the velocity of the body to identify any irregular activity uninterruptedly within the sight of the sensor. Nizam et al. [14] detect the fall with the help of three components viz., velocity of the head, the distance from head to floor plane, and position of other joints. The proposed approach can detect the fall of the subject if it stays on the floor crossing the threshold of 5 s resulting into triggering of alarm. The average accuracy of the approach is found to be 96.55%. The proposed approach gains a sensitivity of 100% with a specificity of 95%. The constraint of following the above work is the special need of Kinect equipment to acquire depth images.

3 Proposed Fall Detection Approach

Our preliminary approach involved training a custom model to detect a person falling or not. For this, the proposed work tried to find datasets of ADL's such as person falling and walking in order to train the model on custom object. But we didn't get any proper dataset to work on. Hence, we created our own dataset which includes images of persons' laying and walking in different postures and positions. Then, we created our own custom model using Faster R-CNN which was proposed by Ren et al. [15] and YOLO first introduced by Redmon et al. [16]. But the results we obtained were unsatisfactory. This happened because we could not bring much diversity in environment of the dataset. Also, we found that the custom object training models are fairly less accurate than the pre-trained models. Therefore, we decided to use a pre-trained model followed by some image processing approach which fairly justifies fall detection.

3.1 Object Detection for Fall Alert

Before moving to our actual fall detection approach, we built a strong platform for reorganization and detection of desired object to declare it as fall or non-fall. We experimented with traditional image processing techniques to detect objects in the image using Haar Cascade by Viola and Jones [17] well known for face detection. Pre-trained Haar cascade is disseminated with the OpenCV library. With this integrated approach we achieved fast results but the limitation of this approach is that it has excessive false-positive detection rate and it can also miss the object entirely, if it is centered on the parameters provided to the cascade. Dalal and Triggs proposed Histogram of Oriented Gradients for Human Detection [18] more familiar with HOG + Linear SVM. This approach focuses on human detection based on histogram of image and its gradients. The HOG + Linear SVM method is easy to train as compared to Haar Cascade. Including higher detection rate, model needs to train with less parameter tuning during inference and lower false-positive rate. The weakness of HOG + Linear SVM is that they incline to be slower than Haar Cascade. The approach of Dalal and Viola shares two crucial components, first sliding windows and image pyramids. A sliding window moves from left corner to right corner of the provided image and classifies each region of interest. Whereas image pyramid sequentially reduces the size of our input image, if input image holds desired object at fluctuating scales. The above approach is a most promising approach of the image processing field. However, deep learning algorithm outperforms these traditional object detection algorithms at every step of the way. When it approaches deep learning-based object detection, there are three primary object detectors we came across, those are—R-CNN family, Single Shot Detector (SSDs) [19], and YOLO. In the first R-CNN which is two-stage detector uses selective search algorithm to propose candidate bounding boxes that could contain objects. These regions were then passed into a CNN for classification, ultimately leading to one of the first deep learning-based object detectors. The next version of R-CNN becames a true end-to-end deep learning object detector by removing the Selective Search requirement and instead relying on a Region Proposal Network (RPN) that is (1) fully convolutional and (2) can predict the object bounding boxes and objectness scores (i.e., a score quantifying how likely it is a region of an image may contain an image). The outputs of the RPNs are then passed into the R-CNN component for final classification and labeling. While R-CNNs tend to very accurate, the biggest problem with the R-CNN family of networks is their speed, they were incredibly slow, obtaining only 5 FPS on a GPU. To increase the speed of deep learning-based object detectors, Single Shot Detectors (SSDs) and YOLO use one stage detector strategy. The YOLO object detection model treats the problem as a regression problem, taking a given input

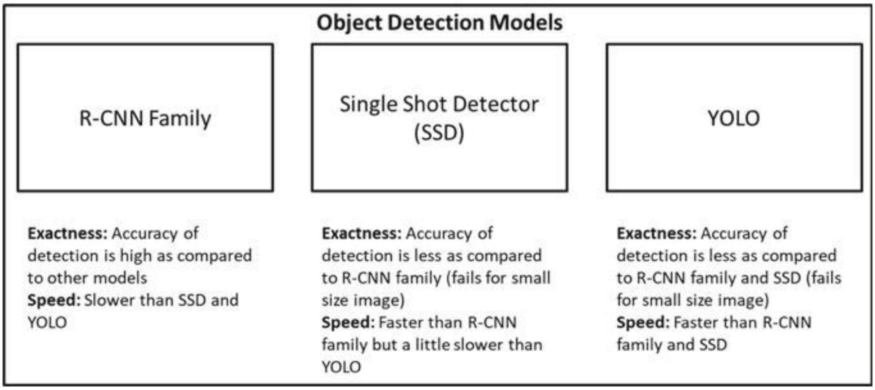


Fig. 1 Comparison of object detection models based on exactness and speed

image and simultaneously learning bounding box coordinates and corresponding class label probabilities. In general, single-stage detectors tend to be less accurate than two-stage detectors but are significantly faster that for Live stream Applications. It is very useful as compared to the R-CNN family and Single Shot Detectors. YOLO takes a given input image and simultaneously learns bounding box coordinates and corresponding class label probabilities obtaining 45 FPS on a GPU. Hence, in the proposed approach, bothers are about the speed and not the exactness (Fig. 1).

3.2 Algorithm YOLO-Width–Height

Input for the YOLO object detection model is a continuous video stream and output is the detection: whether fall or not. The weights and configured files of the model are first initialized. A video stream is input and for each frame, we followed the following procedure.

Step1: Convert the frame into a Blob and pass it through the classifier to update the weights of each output node.
Step 2: For each output in the output layer:

(1) Check detection in the current output node (for various classes).
(2) Update the boxes, confidences, and classIDs lists for the corresponding detections.
(3) Apply Non-Maxima Suppression to keep single detection per object detected in the image.

Step 3: Only consider objects that are humans and use them for further processing.
Step 4: Extract the coordinates of their bounding boxes and calculate the width and height.

Fig. 2 Person detected

Step 5: If width is greater than height for a significant number of frames, detect as fall and alert the concerned authority.

3.3 Results

Using the above mentioned approach, we were able to achieve the following results for the two classes—fall and non-fall (Figs. 2, 3). Accuracy: 93.16%, Recall: 1.0, and Precision: 0.98 over 6792 sample images. Following that confusion matrix depicts the evaluation parameters (Table 1).

3.4 Experiments

We tested this methodology in concern with orientation of camera as follows. With the camera on our testing, we found out that if a person falls directly in the line of sight of the camera, our system fails to detect the fall. Assuming: (1) The width of a person is 3 ft. (2) The height of the person on an average is 6 ft. (3) Height at which the camera is present is 10 ft. (4) The distance of the body from the camera must be about 12 ft for a successful detection. Hence, we need to place the camera appropriately. The following are our calculation for the above assumptions (Fig. 4).

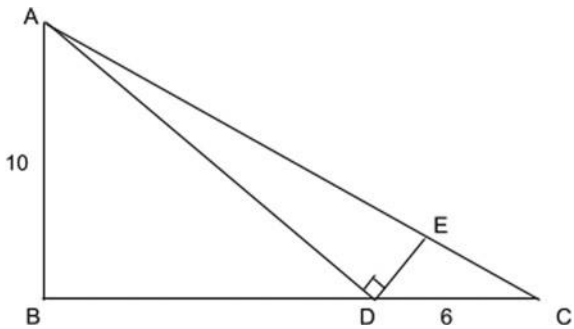
Scenario
Person is fallen directly in the line of sight of the camera. Assuming width of person,

Fig. 3 Fall alert

Table 1 Confusion matrix

Number of test samples: 6792	Fall	No-fall
Fall	4027	0
No-fall	63	2702

Fig. 4 Sample geometric representation of position of object and camera



DE = 3 ft, and height of person, CD = 6 ft. Let AB be elevation of camera = 10 ft. Let D be the point where the person is situated. Let ED be the height of the bounding box detected. For a successful detection, the height detected must be less than 3 ft for a fall; hence, we need to find a position for the person (BD) such that ED is about 3 ft. Consider, BD (Distance of person from camera) = 12 ft (Say).

$$AD = \sqrt{(100 + 144)} = 15.62\,\text{ft}$$

$$\angle ACB = \tan^{-1}(AB/BC) = \tan^{-1}(10/18) = 29^\circ$$

$$\angle ADB = \tan^{-1}(AB/BD) = \tan^{-1}(10/12) = 40^\circ$$

$$\angle CDE = 180^\circ - \angle ADB - 90^\circ = 90^\circ - 40^\circ = 50^\circ$$

$$\angle CAD = 180^\circ - \angle ADC - \angle ACD = 180^\circ - 90^\circ - 50^\circ - 29^\circ = 11^\circ$$

$$\tan \angle EAD = ED/AD = ED/15.62$$

$$ED = 15.62 \times \tan(11^\circ) = 3.03\,\text{ft.}$$

Hence, BD has to be about 12 ft if camera is at a height of 10 ft. In other words, the ratio of the height of camera to that of the distance of the person to the camera is 5:6. Another way to solve this bottleneck is to use multiple cameras perpendicular to each other (Fig. 5).

4 Conclusion

With our novel approach, we were able to get some very impressive results using minimal hardware requirement. Though we had to make some assumptions (minimum distance of the person requirement), we could detect fall of security personnel accurately if the said assumptions were met. We also discussed some possible bottlenecks and some solutions to the said problems. Some improvements as future scope could be considered as follows: (1) Adapt the solution for old-age homes (2) Multithreading (3) Improve fall detection accuracy using better image processing and deep learning techniques. With the implementation discussed thus far, it is evident that image processing combined with object detection could prove fruitful for many applications, one of which has been discussed above.

Acknowledgements We would like to acknowledge CSE department of Walchand College of Engineering, Sangli for supporting us for successful execution of this work.

(a) (b)

Fig. 5 **a, b** False result for vertical fall (toward camera)

References

1. Bourke, K., O'Brien, J.V., Lyons, G.M.: Evaluation of a threshold based tri-axial accelerometer fall detection algorithm. Gait Posture **26**(2), 194–199 (2007)
2. Su, Y., Liu, D., Wu, Y.: A multi-sensor based pre-impact fall detection system with a hierarchical classifier. In: 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pp. 1727–1731 (2016)
3. Belshaw, M., Taati, B., Snoek, J., Mihailidis, A.: Towards a single sensor passive solution for automated fall detection. In: 33rd Annual International Conference of the IEEE EMBS Boston, MA, USA, pp. 1773–1776, August 30–September 3 (2011)
4. Zigel, Y., Litvak, D., Gannot, I.: A method for automatic fall detection of elderly people using floor vibrations and sound—proof of concept on human mimicking doll falls. IEEE Trans. Biomed. Eng. **56**(12), 2858–2867 (2009)
5. Li, Y., Ho, C., Popescu, M.: A microphone array system for automatic fall detection. IEEE Trans. Biomed. Eng. **59**(2), 1291–1301 (2012)
6. Liu, L., Popescu, M., Skubic, M., Rantz, M., Yardibi, T., Cuddihy, P.: Automatic fall detection based on Doppler radar motion signature. In: Proceedings of 5th International Conference on Pervasive Computing Technologies for Healthcare, Dublin, Ireland (2011)
7. Tzeng, H.-W., Chen, M.-Y., Chen, J.-Y.: Design of fall detection system with floor pressure and infrared image. In: 2010 International Conference on System Science and Engineering, pp. 131–135. IEEE (2010)
8. Yodpijit, N., Sittiwanchai, T., Jongprasithporn, M.: The development of artificial neural networks (ANN) for falls detection. In: 3rd International Conference on Control, Automation and Robotics, pp. 271–275 (2017)
9. Foroughi, H., Yazdi, H.S., Pourreza, H., Javidi, M.: An eigenspace-based approach for human fall detection using integrated time motion image and multi-class support vector machine. In:

IEEE 4th International Conference on Intelligent Computer Communication and Processing, pp. 83–90 (2008)
10. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Robust video surveillance for fall detection based on human shape deformation. IEEE Trans. Circuits Syst. Video Technol. **21**(5), 611–622 (2011)
11. Thome, N., Miguet, S., Ambellouis, S.: A realtime, multiview fall detection system: a LHMM-based approach. IEEE Trans. Circuits Syst. Video Technol. **18**(11), 1522–1532 (2008)
12. Nizam, Y., Mohd, M.N.H., Jamil, M.M.A.: Classification of human fall from activities of daily life using joint measurements. J. Telecommun. Electron. Comput. Eng. **8**(4), 145–149 (2016)
13. Nizam, Y., Mohd, M.N.H., Jamil, M.M.A.: Human fall detection from depth images using position and velocity of subject. Procedia Comput. Sci. **105**, 131–137 (2017)
14. Nizam, Y., Mohd, M.N.H., Tomari, R., Jamil, M.M.A.: Development of human fall detection system using joint height, joint velocity and joint position from depth maps. J. Telecommun. Electron. Comput. Eng. **8**(6), 125–131 (2016)
15. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017)
16. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp. 779–788 (2016)
17. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, Kauai, HI, USA, p. I (2001)
18. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05), vol. 1, San Diego, CA, USA, pp. 886–893 (2005)
19. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: SSD: single shot multibox detector. In: European Conference on Computer Vision, pp. 21–37. Springer (2016)

Evaluation of Automatic Text Visualization Systems: A Case Study

Priyanka Jain, R. P. Bhavsar, Karimullah Shaik, Ajai Kumar, B. V. Pawar, Hemant Darbari, and Virendrakumar C. Bhavsar

Abstract We have developed an automatic text visualization (ATV) system, named Preksha, that takes natural language text in Hindi as input and produces a run-time interactive 3D scene based on it. Preksha is the only ATV system which deals with complex processing of morphologically rich input in Hindi, a language of free-word-order nature. Its design and approach make Preksha extendible to other Indian languages. In this paper, we present challenges for evaluation of an ATV system and propose a subjective evaluation methodology. This evaluation design includes intelligibility, fidelity and complexity aspects of the scenes generated. Subsequently, Preksha is evaluated by using a total of 10,220 user responses through an online evaluation survey. The results show that Preksha is able to generate scenes with very high levels of intelligibility and fidelity.

Keywords Text-to-scene conversion · Automatic text visualization · Natural language processing · Cognitive methods

1 Introduction

Automatic text visualization (ATV) is the process of text-to-scene conversion similar to human mental imagery cognition. It understands the input text by natural language processing (NLP) and generates the corresponding visual form. The visual results may be in the form of a static image, 3D scene or multimedia animation. We have developed Preksha—a Hindi language text visualizer [5, 16, 20, 21]. Although other ATV systems have been developed for English and many other languages, Preksha

P. Jain (✉) · K. Shaik · A. Kumar · H. Darbari
Centre for Development of Advanced Computing, Pune, India
e-mail: priyankaj@cdac.in

R. P. Bhavsar · B. V. Pawar
KBC-North Maharashtra University, Jalgaon, India

V. C. Bhavsar
University of New Brunswick, Fredericton, New Brunswick, Canada

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_3

is the first complete ATV system for the Hindi language. Considering the complexity and richness of the Hindi language, Preksha is enriched with a strong linguistic rule-based approach. The architecture of Preksha is modular with a good balance of cohesion and coupling of three major components, namely language, knowledge and scene engines. Considering evaluation of an ATV as a cognitive process [14, 15], this paper describes cognitive methods for evaluating Preksha. We explore the questions of usability of the Preksha system by evaluating its fidelity, intelligibility and complexity.

Section 2 gives a brief overview of related work carried out in the field. Section 3 covers challenges in the evaluation of an ATV system. The methods involved in conducting cognition-based usability analysis are detailed in Sect. 4 and are included in our Preksha evaluation plan. Section 5 discusses user responses and gives their analysis. Finally, Sect. 6 concludes the paper.

2 Related Work

A survey on some of the available ATV systems for non-Indian languages is given in [12]. The latest research on ATV work using machine learning is presented in [17–19]. Subjective evaluation is done in [6, 11] by asking subjects to rate the degree of matching generated scenes with the input text. An educational tool was used by the WordsEye text-to-scene system [13] to help students develop language skills. Further, this system was evaluated comparing its output with simple search methods to find a picture to illustrate a sentence [7].

The earlier research on ATV systems [1–4, 6] is mostly for non-Indian languages. Pertaining to Indian language (IL) family, Hindi has different issues like argument scrambling and morphological richness [8, 9]. Therefore, the results of Preksha cannot be tested by entering the translated Hindi text into other reported research works. Since other available ATV systems have limited working examples and these have no comparability of linguistic behaviour, our evaluation does not include relative evaluation, i.e., comparison with other systems. As there is no other ATV system for Indian languages, we do not provide comparative study results for Preksha.

3 Challenges

Evaluation of an ATV system is an equally difficult task as building it. This fact has been underlined in the operational challenges with various factors, which govern the evaluation of a multi-dimensional ATV system. Unlike text-to-text machine translation system, we cannot count matched and unmatched words in a text-to-scene translation system. The output of the ATV system is a 3D scene. Building references by manual scene creation and comparing it with the machine output for evaluation is also not feasible. The reason behind this is that the visualization has almost unlimited possibilities of reference scene generation. Both in case of mental and machine output, comparing semantic of the visualized scene is almost impossible in today's technical world. Interpretations of input text by different humans and corresponding visualization of the virtual world are another challenge. In this case, the justification for evaluating the machine-generated output with one of the many possibilities of references is challenging.

Another important challenge is to compare the visualization result with human cognition. This is a problem in real life also, where different readers process, comprehend and visualize the information according to their perspectives, past experience, visualizing power and cognition abilities. The generated scene by an ATV system may or may not be exactly mapped to the scene imagined by the user. There are many-to-many relations in the linguistic semantic and its visualization because a scene generated from a text has the possibility of multiple representations. It is also true vice versa, i.e., a scene can be interpreted by two different observers in a different manner.

4 Preksha Evaluation Plan

4.1 Test Suite

A questionnaire was circulated to different class of social groups as an online survey, where 130 users (evaluators) responded voluntarily for evaluation. In this qualitative evaluation, we follow the methods using visual examples as evidence regarding the correctness. For this purpose, we have presented 10 input texts and corresponding scenes generated through Preksha to the evaluators. The evaluator has an option to opt out after evaluating five scenes.

4.2 Input Texts

To illustrate the capability of Preksha, a test set of input texts is prepared with a variety of possible scenes. Preksha constructs a scene using object selection based on object (default/mentioned) qualities, colour, shape and size. Preksha positions objects as per their spatial relations and quantities, and appropriately selects a background image. The input texts correspond to indoor and outdoor scenes. The sentence constructions in a total 10 scenes are a combination of simple, compound and complex sentences having reference of few or multiple objects. These sentences have different degrees of visualization, complexity and length. The average number of sentences in an input text is 3.5, and the average number of words per sentence is 9.8.

4.3 Questionnaire

The test suite is prepared to assess the capability of Preksha in inferring the background and objects to be visualized as per the input text. Each scene was associated with 10 questions, and each question is concerned with a feature of the system. Features of evaluation in the questionnaire are classified with three evaluation criteria as intelligibility in the first three questions, accuracy in the next six questions and complexity in one question for each scene as given in Table 1. Intelligibility is a measure to know that without any reference of input text, how 'understandable' the generated scene is. Fidelity is information retained in the generated scene with reference of the original input text.

Table 2 presents the questionnaire with the description and classification of questions into the evaluation criteria used for the Preksha evaluation. Users were asked to rate how well the generated scene matched the input description on a five-point Likert scale [10].

Our measures have five distinct values. Table 3 represents the measurement of accuracy, intelligibility and complexity as per the ranking scale.

5 User Responses and Analysis

After receiving the user responses, we have observed the user rankings and analysed the evaluation data in two forms. In the first analysis, the question-wise each attribute of the Preksha is observed. Preksha is evaluated by a total 10,220 responses of 130 evaluators through an evaluation survey. We obtained 3,033 responses corresponding to the first three questions of intelligibility (see Table 2), 6,132 responses corresponding to the next six questions of fidelity and 1,022 responses corresponding to last one question of complexity. The details of the various parameters, number of responses and notations used in our result analyses are as follows:

- Total unique numbers of evaluators in the survey $e = 130$
- Total number of scenes (C's) in the evaluation survey $|C| = 10$
- Total questions (Q) per scene in the evaluation survey $|Q| = 10$
- Response received for c-th scene and q-th question $= r_{c,q}$,
  - Since the evaluators have the option for quitting the survey after attempting evaluation of minimum five scenes, each scene has a varied number of responses received. However, each scene has a fixed number (10) of questions to be compulsorily attempted. Therefore, there are a fixed number of responses received for all the questions.
  - Total number of responses received for a qth question from all scenes is calculated by Eq. (1)

Table 1 Preksha evaluation five-point tests score sheet

Scale	Meaning	Interpretation
Five-point intelligibility test scale score sheet		
4	Very intelligent	Completely understandable and quite clear without any ambiguity
3	Fairly intelligent	Information conveyed is understandable, but the generated scene is not natural in visualization
2	Intelligent	Information conveyed is understandable with some efforts. It is generally clear with few inaccuracies in placement and orientation of the objects
1	Barely intelligent	Central idea is clear only after considerable study. It has inappropriate object selection and inaccuracies in size, shape and colour of the objects
0	Not intelligent	Nothing can be understood after any amount of efforts. Completely weird
Five-point accuracy test scale score sheet		
4	Very accurate	Completely accurate. All information from input text is transformed, and a natural scene is generated
3	Fairly accurate	Fairly accurate, most of the original information is passed. A scene is generated with selection of appropriate objects according to input text. The scene may not be very natural in this case
2	Accurate	Accurate, approximately 50% of the original information is passed on. Object placement and orientation of the objects are not correct according to input text
1	Barely accurate	Barely accurate, less than 50% of the original information is passed on. Object selection, size, shape and colour of the objects are not appropriate as per the input text
0	Inaccurate	Completely inaccurate, does not make sense at all. Scene has no resemblance with input text
Five-point complexity test scale score sheet		
4	Very complex	This is very difficult to understand and visualize the input text. The language is complex which has reference of situation with many complicated objects
3	Fairly complex	This is little difficult to understand and visualize the input text with considerable efforts
2	Moderate	This is 'OK' to understand and visualize the input text with some efforts
1	Easy	The input text is simple in terms of language used and visualization of some objects referred in it
0	Very easy	The language used in input text is very simple

Table 2 Questions in English from Preksha questionnaire

Four-point intelligibility test scale
1. Whether the scene is able to convey any message without taking reference of input text? 2. Whether the scene composition is realistic? 3. Whether the object orientation with respect to camera view is appropriate in the scene?
Four-point fidelity test scale
4. Whether the scene is able to convey the same message of input text? 5. Whether the scene has sufficient number of objects as per the requirement to show the meaning of input text? 6. Depending on the input text, whether the objects shown in the scene are properly placed? 7. Whether the colour mentioned of objects in the scene is appropriate according to input text? 8. Whether the size, shape and texture of mentioned objects in the scene are appropriate according to input text? 9. Whether the background of scene is chosen properly according to input text?
Four-point complexity test scale
10. How much difficult is the input text to be visualized by human?

Table 3 Preksha ranking scale distribution

Scale	Intelligibility	Accuracy	Complexity
$3 > x \leq 4$	Very intelligible	Very accurate	Very complex
$2 > x \leq 3$	Fairly intelligible	Fairly accurate	Fairly complex
$1 > x \leq 2$	Intelligible	Accurate	Complex
$0 > x \leq 1$	Barely intelligible	Barely accurate	Barely complex
0	Unintelligible	Inaccurate	Not complex

$$\sum_{i=1}^{10} r_{i,q} = 1022 \tag{1}$$

Here, $r_{c1,q} = 130$, $r_{c2,q} = 130$, $r_{c3,q} = 130$, $r_{c4,q} = 130$, $r_{c5,q} = 130$, $r_{c6,q} = 84$, $r_{c7,q} = 77$, $r_{c8,q} = 71$, $r_{c9,q} = 70$ and $r_{c10,q} = 70$.

- Total number of responses received for all questions from all scenes is calculated by Eq. (2)

$$\sum_{i=1}^{10} \mathrm{r}_{i,*} = 10220 \tag{2}$$

Here, $r_{c1,*} = 1300$, $r_{c2,*} = 1300$, $r_{c3,*} = 1300$, $r_{c4,*} = 1300$, $r_{c5,*} = 1300$, $r_{c6,*} = 840$, $r_{c7,*} = 770$, $r_{c8,*} = 710$, $r_{c9,*} = 700$ and $r_{c10,*} = 700$.

- Rank value of an evaluator's response $R = \{0, 1, 2, 3, 4\}$.
- Ranking value of the ith response for the q-th question belonging to the c-th scene is $R_{c,q,i.}$
- Average of $R_{c,q,i}$ for c-th scene and the q-th question is represented by Eq. (3) where i is ranging from 1 to r_c (total responses received for c-th scene)

$$\bar{R}_{c,q,*} = \frac{1}{r_c}\sum_{i=1}^{r_c} R_{c,q,i} \tag{3}$$

- Standard deviation of $R_{c,q,i}$ is represented by Eq. (4)

$$\sigma_{c,q,*} = \sqrt{(}\frac{\sum_{i=1}^{r_c}\left(R_{c,q,i} - \bar{R}_{c,q,*}\right)^2}{r_c}) \tag{4}$$

- Average of the averages of all the responses for the qth question belonging to all scenes is represented by Eq. (5), where i is ranging from 1 to r_c (total responses received for cth scene) and j is ranging from 1 to c (total number of scenes)

$$\bar{R}_{*,q,*} = \frac{1}{c}\sum_{j=1}^{c}(\frac{1}{r_c}\sum_{i=1}^{r_c} R_{j,q,i}) \tag{5}$$

- Standard deviation of $R_{j,q,i}$ is represented by Eq. (6)

$$\sigma_{*,q,*} = \sqrt{(}\frac{\sum_{j=1}^{c}\sum_{i=1}^{r_c}\left(\bar{R}_{j,q,i} - R_{*,q,*}\right)^2}{c * r_c}) \tag{6}$$

The evaluation process is designed to carry out the measurement of the fidelity and intelligibility of the Preksha system. Table 4 shows the feature-wise user's responses for the intelligibility questions, and a corresponding graph is presented in Fig. 1. The captured user responses are analysed in the subsequent section. The first three questions Q1, Q2 and Q3 are designed to observe the intelligibility of the system. These questions received total 3,066 responses from an evaluation survey. It is observed that the average of the three intelligibility parameters score is 3.02 with a standard deviation of ~1. It means Preksha is very intelligible. Table 5 and Fig. 2 represent the analysis of 6,132 user responses on the fidelity evaluation from question Q4 to Q9. The average of the six fidelity parameters is 2.91 with a standard deviation of ~1. We conclude that Preksha is fairly accurate. The complexity evaluation on 1,022 user responses is shown in Table 6 and Fig. 3. These demonstrate that the evaluators found the scenes 'complex' to be visualized by a human.

Note that 4.4% responses are falling into (0–1) rank and 54.9% responses are falling into (2.1–3) rank scale. Thus, our results show that Preksha 3.02/4 'Very

Table 4 Feature analysis of intelligibility (questions 1–3)

Q	C_1	C_2	C_2	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	$\bar{R}_{*,q,*}$ / $\sigma_{*,q,*}$
Q_1 $\bar{R}_{c,1,*}$	3.07	3.14	3.10	2.95	2.91	2.60	2.68	2.96	3.38	3.07	2.99
Q_1 $\sigma_{c,1,*}$	0.89	1.03	0.98	1.12	1.05	1.14	1.09	1.00	0.82	0.99	1.03
Q_2 $\bar{R}_{c,2,*}$	3.36	3.33	3.02	3.14	3.11	2.64	3.03	2.90	3.28	3.12	3.12
Q_2 $\sigma_{c,2,*}$	0.80	0.92	0.93	0.95	0.95	1.10	0.97	0.99	0.87	1.06	0.96
Q_3 $\bar{R}_{c,3,*}$	2.91	3.02	3.00	3.02	2.98	2.66	2.80	2.91	3.23	2.91	2.95
Q_3 $\sigma_{c,3,*}$	1.00	1.07	1.02	0.99	0.98	1.09	1.08	1.09	1.00	1.07	1.04

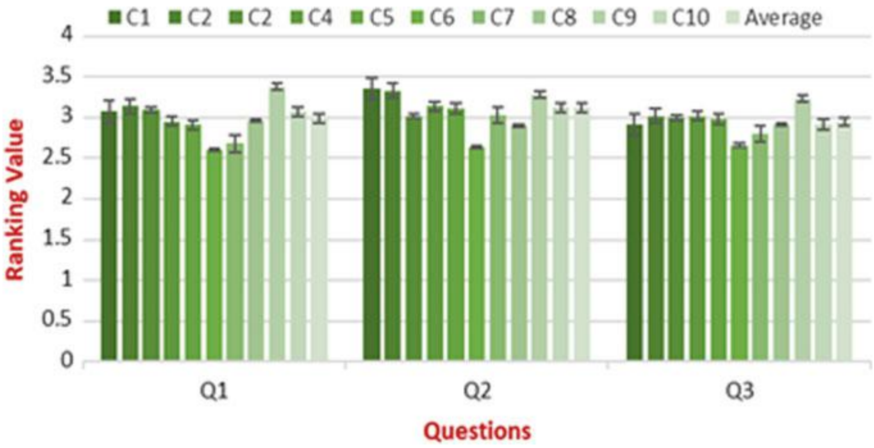


Fig. 1 Feature-wise analysis of intelligibility from questions 1 to 3

Intelligible' and 2.91/4 'Fairly accurate' for generating scenes. It is also observed that the standard deviation of all the ranks has a minimum of 0.68 and the maximum is 1.36. This means that all the users have provided similar ranks without much conflicts in their views. The consistency in human evaluation reflects the reliability of the system to some extent.

Figure 4 is a sample of the Preksha test set. This scene has multiple objects with different attributes, cardinality and spatial positioning. The input Hindi text is ' काफ्रेस टेबल पर सफेद लैपटाप, सिल्वर लैपटाप और सीडी है. कमरे में ब्लैकबोर्ड हैं. एक कोने में पोडियम के पीछे महिला टीचर हैं.पोडियम पर माइक के पास पेंसिल स्टैंड और दो पेन ड्राइव हैं. सफेद लैपटॉप के आगे किताबों पर चाय की प्याली हैं .' The corresponding English text is '*There are white laptop, silver laptop and CDs on the conference table. The room has a blackboard. There is a female teacher in the corner behind the podium. Near the mic on the podium, there is a pencil stand and two pen drives. There are cups of tea on the books ahead of the white laptop.*'

6 Conclusion

Preksha is the only known ATV system for an Indian language, viz. Hindi, and it is extendible to many other Indian languages. The area of automatic text visualization is a comparatively new discipline. It is much difficult in the case of Indian scenario because of linguistic diversity and resource sparsity. Further, this area does not have much-proven work towards finalizing standards and strategies for the evaluation process.

After a short discussion on the evaluation challenges and design methods, especially for Indian languages, this paper has described the evaluation process, results

Table 5 Feature analysis of fidelity (questions 4 to 9)

Q	C_1	C_2	C_2	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	$\bar{R}_{*,q,*}$ / $\sigma_{*,q,*}$
Q_4 $\bar{R}_{c,4,*}$	2.91	3.17	2.99	3.02	3.09	2.86	2.86	2.91	3.04	3.06	3.00
Q_4 $\sigma_{c,4,*}$	1.09	0.88	1.00	0.95	0.96	1.01	1.04	1.05	0.95	0.87	0.98
Q_5 $\bar{R}_{c,5,*}$	2.71	2.87	3.00	3.02	2.87	2.71	2.89	2.96	2.87	2.88	2.88
Q_5 $\sigma_{c,5,*}$	1.03	1.03	0.99	0.98	1.03	1.03	0.99	0.98	1.04	0.90	1.01
Q_6 $\bar{R}_{c,6,*}$	3.10	3.15	3.15	3.19	3.02	2.83	3.03	3.06	3.35	3.23	3.11
Q_6 $\sigma_{c,6,*}$	0.95	1.02	1.01	0.91	0.99	1.06	1.05	1.03	0.90	0.96	0.99
Q_7 $\bar{R}_{c,7,*}$	3.00	2.96	3.15	2.81	2.80	2.55	2.79	2.77	2.96	2.94	2.89
Q_7 $\sigma_{c,7,*}$	1.01	1.09	0.85	1.10	1.07	1.16	0.98	0.98	1.01	0.97	1.04
Q_8 $\bar{R}_{c,8,*}$	2.49	2.86	2.93	2.78	2.71	2.61	2.80	2.90	3.00	3.01	2.79
Q_8 $\sigma_{c,8,*}$	1.13	1.10	1.04	1.05	1.08	1.09	1.01	0.95	1.03	0.85	1.06
Q_9 $\bar{R}_{c,9,*}$	2.66	2.80	2.95	2.95	2.79	2.67	2.80	2.86	2.97	2.86	2.83
Q_9 $\sigma_{c,9,*}$	1.04	1.10	0.95	0.90	1.01	1.00	1.03	0.98	1.00	1.05	1.01

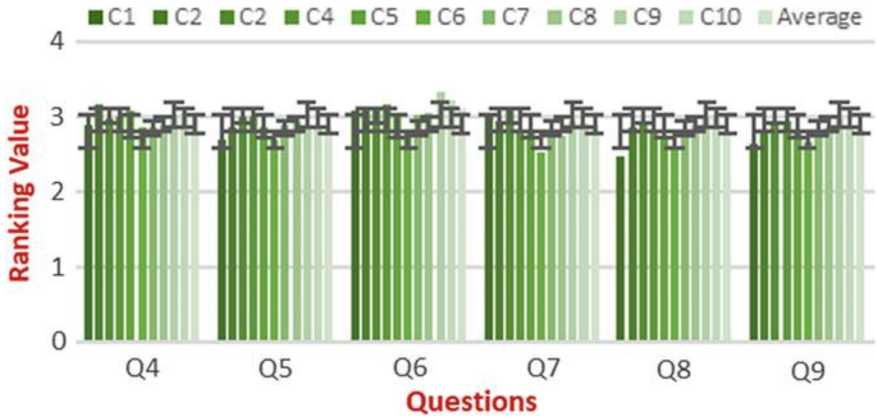


Fig. 2 Feature-wise analysis of fidelity from questions 4 to 9

Table 6 Feature analysis of complexity (question 10)

Q	C_1	C_2	C_2	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	$\bar{R}_{*,q,*}$ / $\sigma_{*,q,*}$
Q_{10} $\bar{R}_{c,10,*}$	1.68	1.74	1.78	1.79	1.79	1.87	1.82	1.79	1.68	1.74	1.77
Q_{10} $\sigma_{c,10,*}$	1.22	1.23	1.36	1.39	1.28	1.22	1.19	1.20	1.37	1.24	1.27

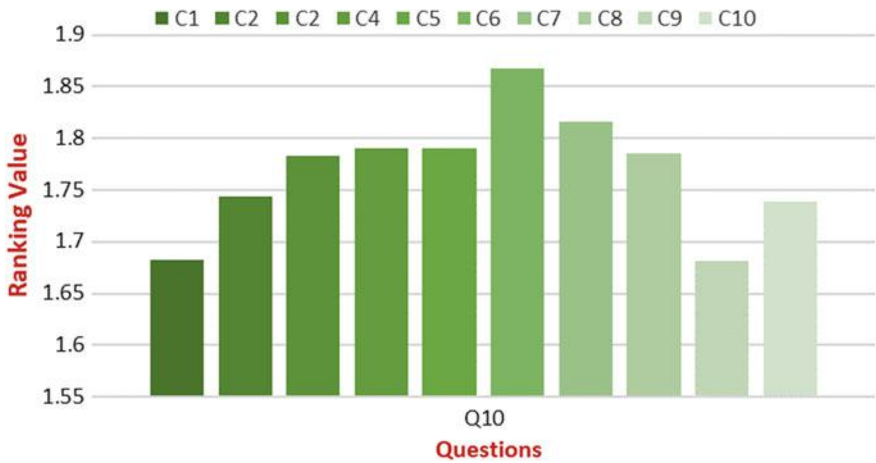


Fig. 3 Feature-wise analysis of complexity for question 10

Fig. 4 Conference room scene generated by Preksha

and error analysis of Preksha. The evaluation plan of Preksha is carried out through an evaluation survey with a total of 10,220 user responses. It is observed with many results that Preksha is ‘Very Intelligible’ (3.02/4 score) and ‘Fairly accurate’ (2.91/4 score) for generating scenes. Our work gives a strong foundation to carry out further work on ATV systems with language learning and cognitive support.

References

1. Sevens, L.: ‘Words divide, pictographs unite’: pictograph communication technologies for people with an intellectual disability. Ph.D. thesis, University of Leuven, jury member (2018)
2. Coyne, B., Sproat, R.: WordsEye an automatic text-to-scene conversion system. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH ’01, Computer Graphics Proceedings, New York, USA, pp. 487–496 (2001)
3. Karkar, G.A.: An ontology based text-to-picture multimedia m-learning system. Ph.D. Thesis, Khalifa Al-Khalifa, Dean, College of Engineering, Qatar University (2018)
4. Seversky, L.M., Yin, L.: Real-time automatic 3D scene generation from natural language voice and text descriptions. In: Multimedia ’06 Proceedings of the 14th Annual ACM International Conference on Multimedia, New York, USA, pp. 61–64 (2006)
5. Jain, P., Bhavsar, R.P., Kumar, A., Pawar, B.V., Darbari, H., Bhavsar, V.C.: Tree adjoining grammar based parser for a Hindi text-to-scene conversion system. In: 4th International Conference for Convergence in Technology (I2CT), and IEEE Xplore (2018)
6. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F.: Shapenet, an information-rich 3d model repository. arXiv:151203012 (2015)

7. Coyne, B., Eric, R., Cecilia, S., Michael, B., Hirschberg, J.: Evaluating a text-to-scene generation system as an aid to literacy. Workshop on Speech and Language Technology in Education (SlaTE) at Interspeech (2011)
8. Bharati, A., Sangal, R.: Parsing free word order languages in the paninian framework. In: Proceedings of ACL (1993)
9. Husain, S.: A generalized parsing framework based on computational Paninian grammar. Ph.D. thesis, IIIT-Hyderabad, India (2011)
10. Likert, R.: A technique for the measurement of attitudes. Arch. Psychol. **140**, 155 (1932)
11. Zitnick, C.L., Parikh, D., Vanderwende, L.: Learning the visual interpretation of sentences. In: IEEE International Conference on Computer Vision, IEEE Xplore and Digital Library (2013)
12. Hassani, K., Lee, W.S.: Visualizing natural language descriptions, a survey. In: ACM Computing Surveys (CSUR) Surveys Homepage archive, vol. 49, issue 1, article no. 17, NY, USA (2016)
13. Ulinski, M.: Leveraging text-to-scene generation for language elicitation and documentation. Ph.D. thesis, Columbia University (2019)
14. Scriven, M.: Evaluation as a cognitive process. School of Behavioral and Organizational Sciences. J. MultiDiscipl. Eval. **4**(8) (2007). ISSN 1556–8180
15. Runco, M.A.: Creativity and cognition. In: International Encyclopedia of the Social & Behavioral Sciences. Elsevier, pp. 2892–2895 (2001)
16. Jain, P., Darbari, H., Bhavsar, V.C.: Vishit: a visualizer for Hindi text. In: 4th International Conference on Communication Systems and Network Technologies (CSNT), IEEE Xplore, pp. 886–890 (2014)
17. Li, W., Zhang, P., Zhang, L.: Object-driven text-to-image synthesis via adversarial training. arXiv:1902.10740v1 [cs.CV] 27 Feb 2019
18. Dong, H., Zhang, J., McIlwraith, D., Guo, Y.: I2T2I: learning text oo image synthesis with textual data augmentation. arXiv:1703.06676v3 [cs.CV] 3 Jun 2017
19. DAR, P.: Thanks to AI, you can now create cartoons from text based descriptions. https://www.analyticsvidhya.com/blog/2018/04/this-ai-create-cartoons-text-description/. Accessed 13 Oct 2019
20. Jain, P., Darbari, H., Bhavsar, V.C.: Cognitive support by language visualization a case study with Hindi Language. In: 2nd International Conference for Convergence in Technology (I2CT), IEEE Xplore (2017)
21. Jain, P., Shaik, K., Kumar, A., Darbari, H., Bhavsar, V.C.: Cascaded finite-state chunk parsing for Hindi language. In: International Conference on Communication and Information Processing (ICCIP) 17th–18th, May 2019 (to be published in Elsevier-SSRN)

Face Recognition-Based Attendance System Using Real-Time Computer Vision Algorithms

Darshankumar Dalwadi, Yagnik Mehta, and Neel Macwan

Abstract The face identification system is one of the most emerging methods for authentication of user; it is drawing wide attraction to the surveillance system which reflects innovation in a video surveillance system. This system here represents the automated attendance system using real-time computer vision algorithms and adaptive techniques to track the faces during a specific period of time. Our system works on eigenface recognizers and Intel's Haar cascades which make the attendance-taking process easier and less time-consuming rather than the traditional process. Our system provides the cheapest solution rather than a previous biometric system like fingerprint authentications. The recorded data is being compared with the training dataset, and the attendance is recorded if the match is found with the help of Python libraries. The camera is being installed at the entry location, so that attendance is recorded as soon as the match of the person entering the particular area is found. So, our main aim is to provide an alternative which is very much convenient to process the attendance and also is very much safe and authentic to have faced as a security option.

Keywords Computer vision · Haar cascades · Eigenface recognizer · Convolutional neural network (CNN) · Internet of things

1 Introduction

The primary aim of this paper is to explain the issues experienced in the old participation framework while replicating a fresh, out-of-the-box, new, creative, and shrewd framework that can give comfort to the establishment. In this task, a brilliant gadget will be created which is fit for perceiving the personality of every person and in the end record down the information into a database framework. Aside from that, a site will be created to give visual access to the information. The task is created depending

D. Dalwadi (✉) · Y. Mehta · N. Macwan
Birla Vishvakarma Mahavidyalaya Engineering College, V.V. Nagar, Anand, Gujarat 388120, India
e-mail: darshan.dalwadi@bvmengineering.ac.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_4

on the IoT (Internet of Things). Open Source Computer Vision is an idea where a savvy gadget is utilized to oversee frameworks. IoT is commonly about various gadgets being interconnected extraordinarily in the current Web framework where data is shared among them. It tends to be seen as a sensory system that connects anything or everything together. The following are the objectives:

- To develop a simple attendance system that would easy to use for people of all domains.
- To provide visual support while taking attendance, so that people know the attendance is recorded.
- To provide the system software by which one can just record attendance by initiating the Python script.
- The future plan is to provide this along with integration with a Web site where people can maintain on the cloud platform to overcome storage issues.

There are two major things of concern in our model—one is Open Source Computer Vision and another is Python distribution environment which supports different packages.

Open Source Computer Vision is the way through which the detection of person or object is possible; along with it, we have implemented Haar cascade recognizer which was developed in order to train many positive and negative images. Features like eyes, nose, and lips can be trained with this recognizer; for this, we have to apply each and every feature for training images—more the data for training, more the accuracy we can attain. Generally, the cascade is very similar to convolutional kernels where each feature is obtained by subtracting the sum of pixels under a bright region from the sum of pixels under the dark region. For setting up Python distribution environment, we have to install many libraries like Numerical Python, OpenCV, Pickle through Python installed packages.

2 Literature Review

Customer participation was taken physically which is very tedious and frequently prompts human mistake. Moreover, there are numerous vulnerabilities toward the wellsprings of the participation records which truth be told, and the majority of the participation records are not recovered from the genuine circumstance. In our institution, we were previously using radio-frequency identification (RFID) tags for marking the attendance, and with the help of Attendance Scanner, each class will then be taken by touching or moving these tags on the lecturer mobile phone. So, in "Face Recognition-Based Attendance Marking System," here the previous database is used so when the camera captures the image, the previous data is used to compare the photograph with data and mark the attendance in the record [1].

This framework is the place where participation is checked whether it is profoundly secure where nobody can stamp the participation of others. Additionally, in this proposed framework, the face identification calculation is improved by utilizing the skin shading arrangement system to build the exactness of the recognition procedure [2]. Albeit more endeavors are put resources into the exactness of the face location calculation, the framework is yet not convenient. This framework requires an independent PC which will require a steady power supply that makes it not versatile [3]. This kind of framework is reasonable for denoting staff's participation as they just need to report their quality once per day, dissimilar to understudies who require reporting their participation at each class on a specific day, it will be badly arranged if the participation stamping framework is not versatile. Subsequently, to illuminate this issue, the entire participation of the board framework can be created on an implanted plan so it tends to be work comparatively with just batteries that makes it compact and less tedious than the previously used system like fingerprint-based attendance system or RFID tag-based attendance system [4].

3 System Blueprint

Our system is divided into two segments—one is software and another is hardware—before to detect face from camera, we have to set up our system that it could run the script and we have used our laptop as processing unit and webcam as camera unit to make our product just cost-effective. Specialty of our product is we have not invested even single penny to make our system execute so it can be considered as best option for educational purpose. We have to use our system processor which is i3 generation Intel Core processor containing AMD graphics processor which was sufficient to run our python algorithms for Haar cascade. For running the algorithms, we have used the Anaconda IDE for running the whole project. Anaconda bundles a whole bunch of Python packages that are commonly used by people using Python for scientific computing and/or data science. It provides a single download and an install program/script that install all the packages in one go. The alternative is to install Python and individually install all the required packages using pip (Fig. 1).

The face database is an important step to be done before any further process is done in our paper. We have first created and defined the path for images which will serve as dataset then training of images. It is to be done with the help of Numerical Python library and array in anaconda IDE; then, we have to select the area of interest, so that we have use pickle class for saving the data or the image data after all of this, and we have implemented Intel's Haar cascade recognizer. It actually detects the face of the human after the detection is done, and we have saved the trained model in the .yml file format; below is the flowchart for fetching the data [5].

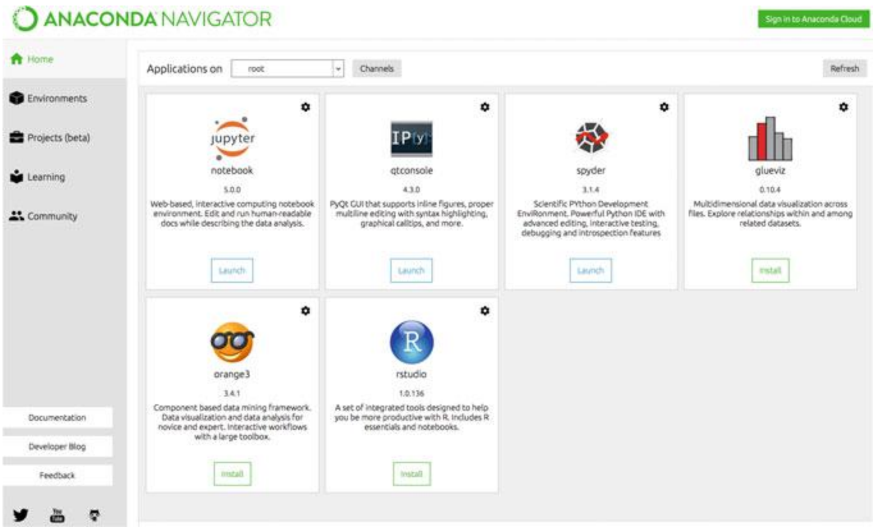


Fig. 1 Anaconda software

4 Software Development

Specific Requirements
The followings are the basic software requirements:

- OpenCV 3.4, Python 3
- Tkinter package—It is used to provide the user interface
- Python Image Library package—It is used for capturing a portrait
- NumPy package—It is a numerical python package which allows the mathematical calculation for Haar cascades
- Cv2—Package is for accessing the OpenCV 3 in our anaconda ide
- Pandas—It is used to generate the CSV or XLS file containing the recorded attendance.

Our Methodology
There is a lot of information that should have been inputted into the framework which basically comprises the person's essential data which is their ID and their countenances. The main method of picture obtaining should be possible by utilizing the workstation Camera to catch the essence of the person. That captured images are converted into a grayscale [1]. We have also made a special directory in local Processing unit which contains the stored data which is used for training and testing below is a glimpse of our test data sets Fig. 2a, b.

The face database is an important step to be done before any further process is done; in our paper, we have first created and defined path for images which will

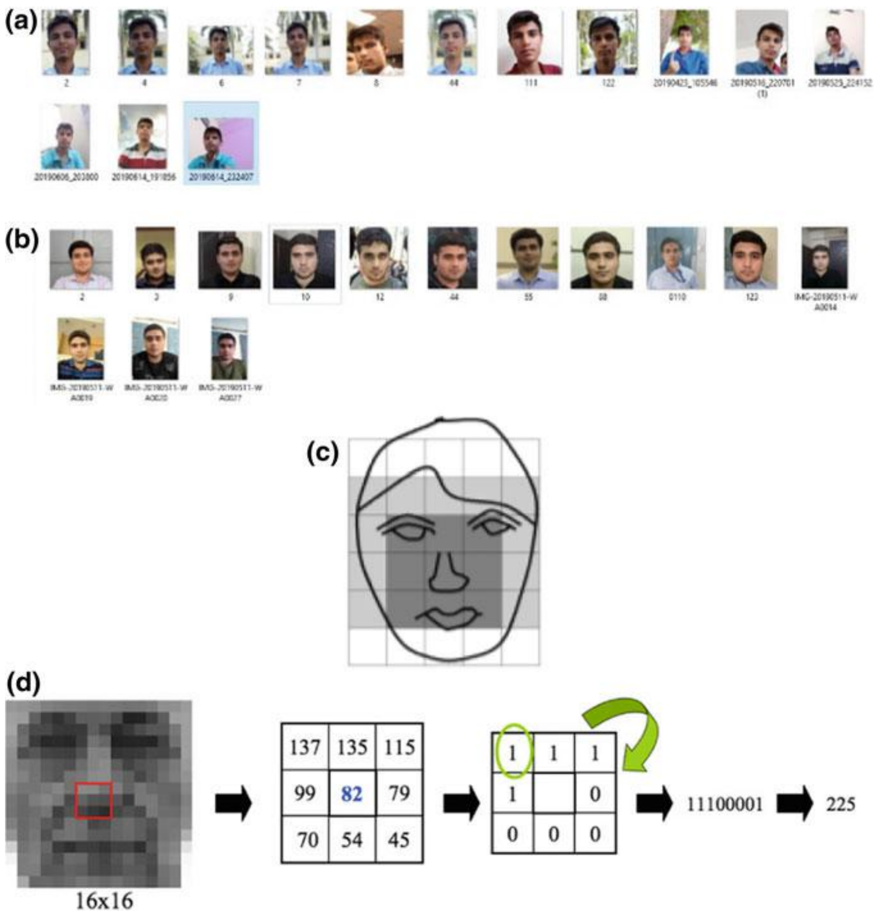


Fig. 2 **a** Image template, 1. **b** Image template, 2. **c** Typical face result obtain based on the intensity distribution, and **d** process of generating a unique key for the particular feature using LBPH

serve as dataset, then training of images is to be done with help of NumPy library and array in anaconda IDE, then we have to select the area of interest, and then we have use pickle class for saving the data or the image data; after all of this, we have implemented Intel's Haar cascade recognizer which actually detects the face of the human after the detection is done.

The Haar cascade algorithm is based on subtracting the pixels under brightness by the sum of pixels under the darker region. This means that whenever the feature is generated, it consists of some value which is a subtraction of these two values. Equation of Haar cascade is as below Fig. 2c:

$$X_i = \sum_{a}^{b} X_{lightpixels} - \sum_{a}^{b} X_{darkpixels}$$

where P is an individual pixel part which can be dark or light and xi is the feature value of that group of pixels in a selected area of interest that is human face [6]. We have used LBPH process of face recognition for training and testing of the face data model where the 3×3 window is taken in which the face of the person is being acquired; it is converted to grayscale where the whole face feature is converted to values; from the given values, we will set a value; let say 82 then feature value above 82 will be considered as binary 1 below it will be considered as 0 and then we get key for each face dataset which you can see below Fig. 2d [7].

Process of Taking Attendance

The rest of the procedure should all be possible through a Jupyter scratchpad. Consequently, the participation-taking methodology will likewise be finished with the assistance of Pandas and NumPy cluster. This is to give a well-disposed UI to the administrator in while having the option to direct an execution on the PC and Processor to do participation taking without the need to control the workstation from a terminal which will be vague for most user [8]. Along these lines, just with a tick of catch on the screen get report, a python content will be executed which it will dispatch a progression of, for example, stacking the prepared information to the recognizer and so on. The attendance-taking process will then proceed in a loop to acquire, identify and mark the attendance for each of the students that are obtained from the integrated camera of laptop python script will be initiated after that Haar cascade algorithm is implemented on area of interest which will initiate recognizer it will compare the face with the existing database stored in locally inside laptop after that it will load the labels if match is found it will export the label as name of present student in CSV format and a report is generated if the matched is not found it acquires the face detected in the frame and generate label and thus attendance is stored in CSV format.

5 Training and Testing of Model

Here there are flowcharts for various process which are going to execute in this paper like installing OpenCV in system portrait acquisition and exporting labels to the CSV. This is flowchart describing how OpenCV is installed where we have created python distribution environment also, we have installed the anaconda environment the virtual environment is created through pip and OpenCV package is installed (Fig. 3).

Next Flowchart is about image acquisition process which allows our script to capture the images from the script and more where we have defined the path to the database also, we have trained the dataset using the Numerical Python. Then we have picked up an area of interest for recognition and implemented the pickle library for saving the data after the whole process the Haar Cascade Eigen Recognizer is being implemented (Fig. 4a).

Below is how the whole process of our model works: First the Python script is being initiated, which initialized the camera to start; now the Intel cascade is

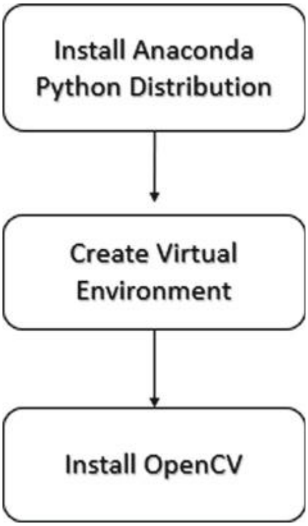


Fig. 3 OpenCV installation flowchart

applied, the labels are loaded, and then, it compared the captured video data with exiting database of trained model person. If the match is found, then it adds a label to CSV report through the Pandas library and data is stored in a local folder with the help of pickle [6]. Here hence, the data is being acquired it is being trained which shows in below image (Figs. 4b and 5).

6 Experimental Results

Here when a script is initiated as we have trained the model of Neel, here it compares his face from preexisting loaded data of him and generates the label; finally, the label is exported in the report, and it is our final output [7] (Fig. 6).

And the below image is output report generated by Pandas after the execution of the whole script where the names repeated suggest that the following person is present in front of the camera for the following seconds; here, we have set the delay time at 1.00 s means at one second it recaptures and redos the above process (Fig. 7).

Here is our whole system layout, and it shows how our model looks after the whole process (Fig. 8).

This model's output efficiency can be improved with better graphics processor, and better lighting conditions are the parameters on which the accuracy of the system depends. In this model, we are able to achieve 80% of accuracy under the normal lighting conditions and with minimum available resources.

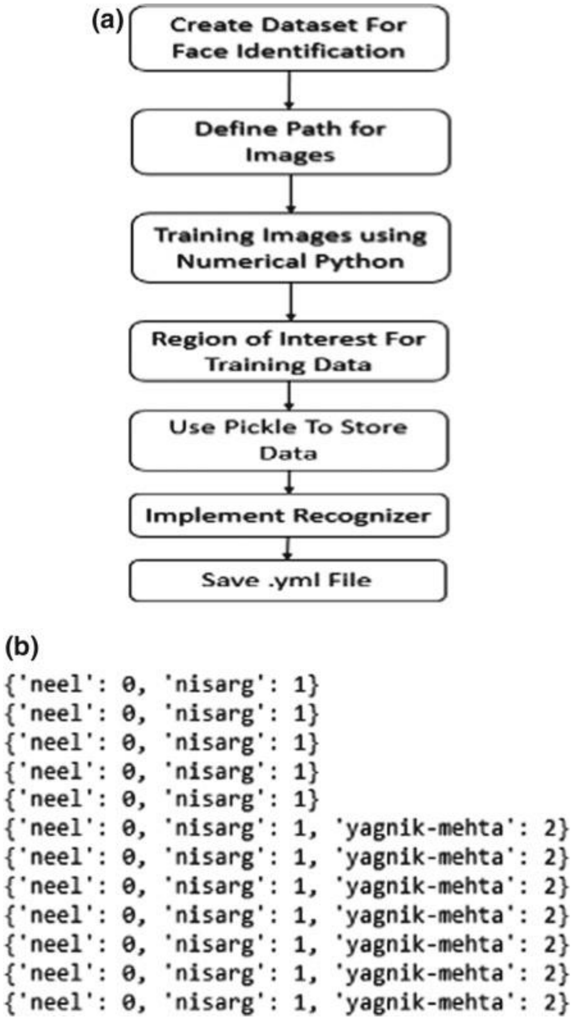


Fig. 4 **a** Flowchart for the image acquisition process. **b** Training flow of data

7 Conclusion

This innovation can diminish the exertion of upholding understudies to at end classes a several thing is computerized. This had given accommodation not only to the speaker, yet in addition to the understudies in light of the fact that the participation-taking procedure for a class of roughly 100 understudies should be possible in 5 min which is far more quickly than the old strategy for passing participation sheet around in the study hall which made a ton of issues to the establishments and burden to the

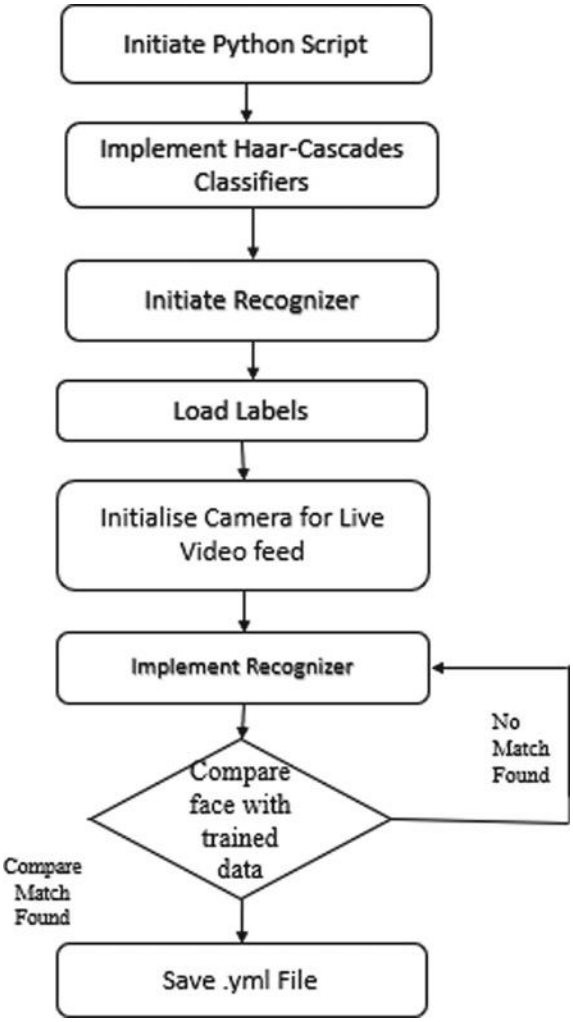


Fig. 5 Flowchart of process execution

understudies. There are numerous escape clauses during the time spent gauging participation utilizing the old technique which made numerous issues the vast majority of the establishments like legislative or private. Therefore, the facial acknowledgment highlight implanted in the participation checking framework cannot just guarantee participation to be taken precisely and furthermore disposed of the defects in the past framework. By utilizing innovation to vanquish the deformities cannot only spare assets yet, in addition, lessens human intercession in the entire procedure by taking care of all the muddled undertaking to the machine. The main expense of this

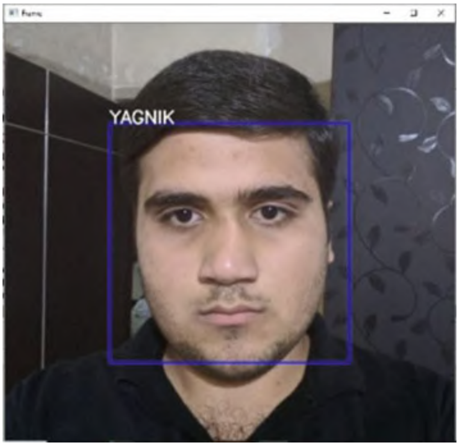


Fig. 6 Label generated along with comparison

	A
2	neel
3	neel
4	neel
5	neel
6	neel
7	neel
8	neel
9	neel
10	yagnik
11	yagnik
12	yagnik
13	yagnik
14	yagnik
15	yagnik

Fig. 7 Output file

arrangement is to have adequate space into the store every one of the appearances into the database stockpiling.

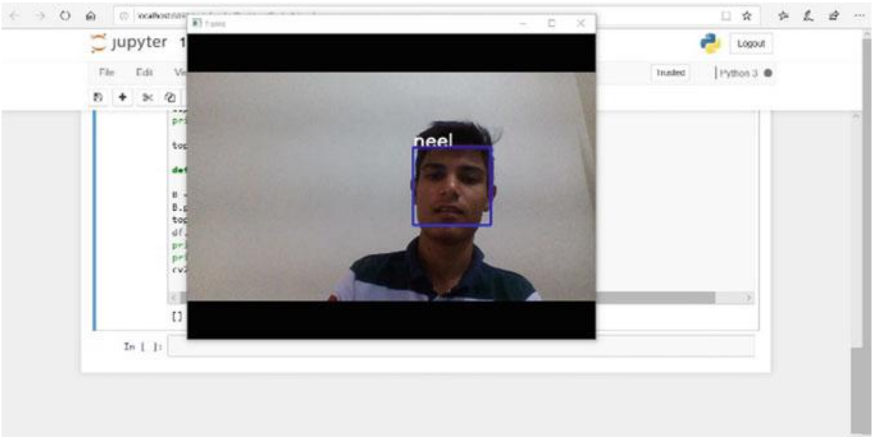


Fig. 8 Output frame

References

1. Raja, R.: Face detection using OpenCV and Python. SuperDataScience—Big Data—Analytics Careers—Mentors—Success (2017). https://www.superdatascience.com/opencv-face-recognition/. Accessed 19 June 2019
2. Docs.opencv.org.: Face recognition with OpenCV—OpenCV 2.4.13.4 documentation (2017). http://docs.opencv.org/2.4/modules/contrib/doc/facerec/facerec_tutorial.html#facehdatabase. Accessed 13 June 2019
3. Libelium.com.: Top 50 internet of things applications—Ranking—Libelium (2016). http://www.libelium.com/resources/top_50_iot_sensor_applications_ranking/. Accessed 6 June 2019
4. Rosebrock, A.: Practical Python and OpenCV: learn computer vision in a single weekend. Pyimagesearch.com (2017). https://www.pyimagesearch.com/practical-python-opencv/?src=sidebar-face-detection. Accessed 18 June 2019. 60 BCS (HONS) Computer Engineering Faculty of Information and Communication Technology (Perak Campus), UTAR
5. Shehu, V., Dika, A.: Using real-time computer vision algorithms in automatic attendance management systems. https://ieeexplore.ieee.org/document/5546452
6. GitHub.: iashris/Face-Detection-OpenCV (2017). https://github.com/iashris/Face-Detection-OpenCV/blob/master/FaceDet.py. Accessed 13 June 2019
7. Python tutorials 1 documentation. http://opencv-python.tutroals.readthedocs.io/en/latest/py_tutorials/py_gui/py_image_display/py_image_display.html. Accessed 13 June 2019
8. Ieeexplore.ieee.org.: Internet of things for smart cities—IEEE Xplore document (2014). http://ieeexplore.ieee.org/document/6740844/. Accessed 6 June 2019

The Adopting Knowledge Management Proves Vital for Bahrain's Governmental Sector

Abdulaziz Showaiter, Anjum Razzaque, and Allam Hamdan

Abstract This study sheds light on how the knowledge management (KM) revolution contributed in evolving the governmental performance in Bahrain since few decades. The study was essential to contribute to the limited literature and establish possible contributions that KM has made in the government sector in Bahrain. Hence, a quantitative research followed an adopted survey; piloted through a non-structured in-depth interview, for testing the credibility of this study's adopted framework. The collected data was analyzed and revealed that KM in Bahrain was at 50%, though such initiatives significantly affect innovation. In terms of responsiveness, this study found out that those who recorded less information retrieval time had also recorded a high maturity level, and that KM was found to be significant in supporting Bahrain vision 2030. In this vein, the study revealed a strong correlation between competitiveness and KM, since KM enhances management of human capital, vital for Bahrain's economy. Precisely, the study confirmed that a knowledge creation and management lead to economic transformation from the traditional to a knowledge-based sector. Study recommendations aim for the government to plan for the upcoming organizations amidst stiff competition.

Keywords Knowledge management · Government sector's performance · Bahrain

1 Introduction

Leading organizations report that employees' talent and interests help achieve goals with higher productivity, with lower turnover and more employee satisfaction. Strategic Human Resources Management, SHRM, is not only about introducing human capital plans. It is more concerned with strategies' implementation [1]. It is involving in monitoring employees' behavior toward daily strategies to meet organizational

A. Showaiter
The George Washington University, Washington, USA

A. Razzaque · A. Hamdan (✉)
Ahlia University, Manama, Bahrain
e-mail: allamh3@hotmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_5

goals for ensuring best practices and documentation of the lessons learnt. SHRM is a toll to design business processes' approaches to managing employees' behavior and performance to focus the long-term requirements and considering the continuous change in the organization operating context [2]. And to be more specific, SHRM is a mindset which only comes alive in the organization's managerial levels when managers concentrate and pay more attention to the behavior of the employees toward the organization's vision, mission, goals, values, and aims. Today, globalization has made the nature of the businesses on edge to make continuous changes in improving their environments, technologies, business processes, ways of operations and employees' skills, all of which to maintain their competitive advantage in the market [3–5]. On top of that, organizations today must focus on escalating their level of performance while trying to cut on costs, improve quality level, and create novel products and\or services. Bearing in mind to implement all these changes around the capabilities of the workforce to ensure the harmony between the human capital and the business nature to run smoothly [6, 7], changes usually face resistance, but an organization can minimize that by being smart in effectively managing its human resources in terms of their commitment, engagement, motivation and by providing the required synergy if needed. Being an organization run by smart management tends to apply the adequate HRM strategies to maintain the competitive advantage and to shape employees' attitude and positive behavior [8]. In addition, knowledge management (KM) is being the selective application of knowledge from past experiences of making current and future decisions for the main purpose of improving the effectiveness of the organization. The goal of KM is to identify knowledge that is critical and apply it in the appropriate situations. Acquire the critical knowledge in an organizational memory or knowledge base, determine how effective the knowledge being applied is, and adjust the use of knowledge to improve on effectiveness [9]. Knowledge revolution is a phase of capitalism that is dependent on brainpower, education, skills, human capital information, knowledge, intellectual capital, intangibles ideas, and brand names. It can only be compared to the industrial or agricultural revolution. It is based on the current postindustrial era where there is a lot of automation and routine. Value is being derived from education. Organizations need to develop their KM capabilities which will allow them to support a few of the important operations and activities of the organization. KM can help the organizations in storing, utilizing, and acquiring of processes like strategic planning, dynamic learning, planning, problem solving and decision making. It is publicly known, in Bahrain, that variations on quality levels of delivery have been spotted on official and non-official performance and service reports in the government sector in the Kingdom of Bahrain. A rationale is applied in other contexts as well [10]. According to that, this study aims to determine the reflection of the KM revolution on the Bahrain governmental sector in terms of improving performance and manage human capitals.

2 Results of Analysis

For the purpose of this research, a sample of 55 participants was randomly chosen based on availability. This sample was divided into categories, i.e., civil servants and members of the government sector. This sample was divided this way in order to capture both the beneficiaries and the implementers of the program. The survey determined organizational KM awareness and initiatives. Table 1 depicts the descriptive statistics (N = 18, 32.7% males and N = 14, 25.5% females), between 20 and 30 years of age.

2.1 Correlation

Correlation is a statistical ranking tool that measures the degree by which two or more variables are related. The questions were categorized under the following groups, and their average was taken to make the analysis simpler. The result of analysis shows that Pearson's r correlation between age and the four categories are negatively correlated. On the other hand, there is a statistical significance seen between gender and management (p-value = 0.014) and gender and application (p-value = 0.027) using a level of significance = 0.05. This means that male participants agree that there is management support given and ample application of KM. Furthermore, there is high statistical significance ($\alpha = 0.01$), and the three categories are understanding and recognition, training and promotion, and management support (Table 2), indicating that respondents believe that to apply KM to daily activities, top management involvement is necessary in training and cascading information.

Table 1 Sample characteristics

Characteristics		Frequency	Percent
Gender	Female	14	25.5
	Male	18	32.7
	Missing	23	41.8
	Total	**55**	**100**
Age	Less than 20 years old	3	5.5
	20–30 years old	29	52.7
	31–40 years old	12	21.8
	41–50 years old	8	14.5
	More than 50 years old	3	5.5
	Total	**55**	**100**

Table 2 Correlation matrix

	1	2	3	4	5	6
1. Gender	1					
2. Age	−0.034	1				
3. Understanding and recognition	0.044	−0.012	1			
4. Training and promotion	0.24	−0.001	0.757[b]	1		
5. Management support	0.331[a]	−0.211	0.625[b]	0.834[b]	1	
6. Application	0.299[a]	−0.031	0.713[b]	0.952[b]	0.882[b]	1

[a]Correlation is significant at the 0.05 level (two-tailed)
[b]Correlation is significant at the 0.01 level (two-tailed)

2.2 Regression

Regression analysis aims to determine which factors have a significant effect on the dependent variables. The result in Table 3 shows that 59.3% of the variability seen in the understanding and recognition of KM is predicted by the model. Since the p-value is < 0.05, the model is significant but only the training and promotion variables can predict the dependent variable since p-value = 0.16 < alpha = 0.05.

Table 4 presents that 92% of the variability seen in the training and support needed for KM is predicted by the model. Since the p-value is < 0.05, the model is significant but only the *Application* variable can predict the dependent variable since p-value = 0.00 < alpha = 0.05.

Table 5 presents that 81.7% of the variability seen in management support given to KM is predicted by the model. Since the p-value is < 0.05, the model is significant but only the *Application* (p-value < 0.00) and *Age* (p-value = 0.005) variables can predict the dependent variable since their p-values < alpha = 0.05.

Table 3 Understanding and recognition

Model 1: Understanding and recognition			
Variables	Beta	t-test	Sig.
(Constant)		1.259	0.214
Gender	−0.152	−1.556	0.126
Age	−0.009	−0.090	0.928
Application	−0.001	−0.002	0.999
Management support	0.039	0.182	0.856
Training and promotion	0.761	2.508	0.016
F-test	14.289		
Sig. (F)	0.000		
R	0.770		
R square	0.593		
Adjusted R square	0.522		

Table 4 Training and promotion

Model 2: Training and promotion			
Variables	Beta	t-test	Sig.
(Constant)		0.313	0.756
Gender	−0.021	−0.466	0.643
Age	0.027	0.615	0.541
Understanding and recognition	0.149	2.508	0.016
Application	0.851	8.422	0.000
Management support	0.002	0.022	0.983
F-test	12.934		
Sig. (F)	0.000		
R	0.959		
R square	0.920		
Adjusted R square	0.912		

Table 5 Management support

Model 3: Management support			
Variables	Beta	t-test	Sig.
(Constant)	0.794	3.089	0.003
Gender	0.087	1.081	0.285
Age	−0.112	−2.966	0.005
Understanding and recognition	0.015	0.182	0.856
Application	0.005	0.022	0.983
Management support	0.821	4.047	0.000
F-test	43.735		
Sig. (F)	0.000		
R	0.904		
R square	0.817		
Adjusted R square	0.798		

Lastly, Table 6 depicts 93.5% of the variability seen to apply KM: predicted by the model. Since the p-value is < 0.05, the model is significant but only the *Management Support* (p-value < 0.00) and *Training and Promotion* (p-value < 0.00) variables can predict the dependent variable since their p-values $<$ alpha $= 0.05$.

Table 6 Application

Model 4: Application			
Variables	Beta	t-test	Sig.
(Constant)	−0.138	−0.813	0.420
Gender	0.043	0.869	0.389
Age	0.021	0.858	0.395
Understanding and recognition	0.000	−0.002	0.999
Training and promotion	0.695	8.422	0.000
Management support	0.305	4.047	0.000
F-test	14.047		
Sig. (F)	0.000		
R	0.967		
R square	0.935		
Adjusted R square	0.928		

3 Discussion of Results

Data analysis indicates the status of government organizations with regard to insight of KM and how they spread its importance for the others in the organization. According to the analysis, 55 randomly selected individuals provided views of KM. The data was coded into SPSS to analyze the responses from different individuals. The responses were grouped according to respondents' age and gender. According to the analysis, the participants feel that there is a management support and an adequate application of KM. The high statistical significance between application and the three categories is from the impact of KM revolution in the Bahrain's government sectors. The analysis confirms that due to KM revolution; individuals expect that the top management and employees must always be involved in the organization's daily activities. The KM revolution has also increased awareness of employees on the training they need to receive in order to improve their performance of work. From the survey analysis, the employees believe that for proper KM, there is a need for training and support. This is depicted in Table 6 that presents the percentage of the variability seen in the training and support needed for KM. The analysis also indicates that the revolution of KM has affected the way government sectors in Bahrain consider management support given to KM. Table 5 depicts the percentage of variability seen in management support that is given to KM in government sectors. The KM practices may be used to influence initiatives that help in ensuring better government development. It is, therefore, important for government sector to continue with research on KM and how its maturity can lead to improved performance in the government sector. According to the survey analysis, the KM revolution also has had a significant impact on the application of KM. Table 6 indicates the percentage variability seen in the application of KM as predicted by the model. Different government sectors need to address continuously the KM initiatives that would lead to the knowledge-based government sector. The analysis demonstrates that positive KM is an important tool in the government sector in the Kingdom of Bahrain. The current research provides information about KM that may be used by government sectors, leaders, and other decision-makers in the government to understand and focus on the specific indicators that are in line with the government programs. The research supports leaders and other officials in the government to understand the consequences of the implementation of KM. For government sectors to improve performance and effectiveness, there is a need to select and implement the necessary KM framework; this will assist the government sector to tackle challenges from changing environment: Making the development of a KM frame is not only desirable, but also necessary for the government sector because such a framework will establish a methodology that would link different disciplines in the sector to KM, thus improving the performance of the entire government. KM can be referred to as the activities through which the government sector get value from the knowledge-based assets and intellectual capital available. To implement KM strategies, the government sector can carry out practical programs that have clear milestones and KM strategy that will be practically effective in improving the government sector's performance. The government sector of Bahrain is expected by the citizens to deliver high-quality services about education, infrastructure, and health care. Proper KM initiatives in the government sector can lead to highly efficient and effective government because the government sector is considered to have vast knowledge. Such KM can increase the country's use of internal resources, thus improving the overall economy. KM is especially important in cases where tacit knowledge is not available. KM utilizes the internal resources of the government sector to ensure relevant initiatives are supported, and the required development practices take place. The current study may be used to analyze how government sector makes use of KM and its influences to improve the performance of the government. Therefore, this study may be used as a tool by different government sectors in the Kingdom of Bahrain for providing a better understanding of the discipline of KM by ensuring adoption and facilitation of KM initiatives. KM will also ensure that the appropriate initiatives are prioritized. Initiatives developed through the KM processes can be regarded as important internal assets in the different government sectors. The results of this study encourage the government sector to support KM initiatives as a tool that they can use in creating success and improved government development practices. Appropriate management of knowledge will help ensure that there is reduced opportunity loss through processes that ensure that initiative selection is linked to the overall objectives of the government sector. If the government sector is able to optimize its initiatives, it is highly likely that the outcomes and value proposition will be more sustainable and tangible. The adoption of appropriate KM can be used to ensure that positive influence on the performance of government sector. Such KM can increase the innovation and learning in the government sector, thus improving the performance of the entire government. This is in line with the issues of KM infrastructure where the capabilities of knowledge process lead to improved organizational performance and organizational learning. This research provides a model that may be used by government sectors in ensuring that there is efficient and appropriate utilization of resources and provision of services to the citizens. The study analysis detail on what leaders and government officials manage knowledge to improve organizational performance. If the officials in the government sector embrace the spirit of KM practices, then they can understand how KM practices add value the initiatives to improve efficiency and performance of the government sector. The proposed KM model that may be used in the government sector will generate better results and will improve the relationship between KM and the internal resources with different development practices in the government sector. The current research provides a direction for future studies focused on the strengths associated with integration of different disciplines that form a comprehensive framework that can be used to tackle different problems associated with government sectors. There are different human resource management strategies that may be used to improve the performance of government about KM. However, a single approach may not be appropriate to ensure improved performance due to the difference in employees' capabilities and motivation. Therefore, there is a need to mix different approaches. The mixed approaches should be used to provide a model that will facilitate the transformation of the government sector toward improving its performance and being more effective and efficient. The study indicates that KM is a crucial source of an organization capacity in the government sectors. KM can be considered as an important influence that can enhance the performance and overall competitiveness in the government sectors. The current study confirms this by indicating how KM practices have revolutionized the government sectors. Currently, government sector in the Kingdom of Bahrain is facing radical changes; thus, there is a need to adapt to such change to ensure continued effectiveness in this sector. However, there is a challenge in linking the different practices with the KM. The existing lack of positive approaches that may be used to influence KM implementation and thus getting its value is an important issue that needs to be considered in future research.

4 Conclusion

The current research can also help the different government sectors identify the actions to take regarding how KM and how government sector performance improvement can be targeted. If this research is conducted and enhanced in different sectors, it aids management in government and private sectors in improving the performance and effectiveness of their organizations. The current research helps practitioners and leaders understand the central role played by KM through the generation of initiatives that have different developing values and classifying them accordance with their priority. Future research may be focused on developing KM best practices that consider the relationships between cause and effect which is still considered a challenge in different government sectors. The study finding may be developed further, so that they can be used practically for the self-assessment of different government sectors to ensure optimum KM initiatives; consequently, improving value creation in the different parts of the government. KM can improve the competitiveness of a private or government sector whether direct or indirectly. For instance, KM can influence resources and internal values that can make the government sector to be highly efficient and effective. Knowledge can be assessed by the use of different methods; future research is recommended to focus on how knowledge may be managed based on values and how such knowledge can be used to increase competitiveness in the private sector as well as in the government sector. Government sector should understand the capability of KM and how to effectively manage knowledge assets. The government sector creates knowledge to ensure performance. Future research should inquire on how government sectors initiate KM processes to enhance performance to facilitate the KM capability to boost performance.

References

1. Araimi, F.A.: Power of Human Resources. AuthorHouse (2011)
2. Razzaque, A.: Organizational learning through Knowledge sharing in virtual communities thru learner readiness and their social capital. In: Albastak, Y., Al-Alawi, A., Al-Bassam, S. (eds.) Implementing Knowledge Management Strategy in the Public Sector. IGI Global Publications (2019)
3. Razzaque, A.: Virtual learning enriched by social capital and shared knowledge, when moderated by positive emotions. Int. J. Electron. Bank. **1**(4), 358–373 (2019)
4. Razzaque, A., Jalal-Karim, A.: The influence of knowledge management on EHR to improve the quality of health care services. In: European, Mediterranean & Middle Eastern Conference on Information Systems (EMCIS 2010). Abu-Dhabi, UAE (2010)
5. Lucas, L.: The evolution of organizations and the development of appropriate knowledge structures. J. Knowl. Manag. **14**(2), 190–201 (2012)
6. Armstrong, M.: Armstrong’s Handbook of Strategic Human Resource Management. Kogan Page (2011)
7. Razzaque, A., Eldabi, T.: Assessing the impact of physicians’ virtual communities on their medical decision making quality. In: 51st Hawaii International Conference on System Sciences—HICSS-51. University of Hawaii at Manoa, Hawaii (2018)
8. Harzing, A.P.: International Human Resource Management. SAGE Publishers (2010)
9. Zheng, W., Yang, B.: Linking organizational culture, structure, strategy and organizational effectiveness: mediating role of knowledge management. J. Bus. Res. **63**(1), 763–771 (2011)
10. Razzaque, A., Eldabi, T.: Assessing the impact of physicians’ virtual communities on their medical decision making quality. HICSS-51: Jan 3–6, 2018. Honolulu, Hawaii: HICSS-51. http://engagement.hicss.org:81/papers/0398/paper0398.pdf (2018)

Video Surveillance for the Crime Detection Using Features

Aastha Chowdhary and Bhawana Rudra

Abstract This paper aims at extending the comparison between two images and locating the query image in the source image by matching the features in the videos by presenting a method for the recognition of a particular person or an object. The frames matching the feature (not feature its query) object in a given video will be the output. We describe a method to find unique feature points in an image or a frame using SIFT, i.e., scale-invariant feature transform method. SIFT is used for extracting distinctive feature points which are invariant to image scaling or rotation, presence of noise, changes in image lighting, etc. After the feature points are recognized in an image, the image is tracked for comparison with the feature points found in the frames. The feature points are compared using homography estimation search to find the required query image in the frame. In case the object is not present in the frame, then it will not present any output.

Keywords Video similarity search · SIFT · Object recognition · Feature extraction · Video surveillance

1 Introduction

Video tracking solves many challenging problems and is a key technology with wide-ranging applications in the field of image processing such as crime scene tracking, VIP tracking and security. It is the process of locating an object which is moving over a period of time. For locating objects in a frame, there is a need for finding distinct local feature points which are invariant to image scaling, rotation, changes in illumination, noise, etc. These feature points are used for image detection. Image detection is the method of locating a given query image in a source image by making

A. Chowdhary (✉) · B. Rudra
Department of Information Technology, National Institute of Technology Karnataka, Mangalore, India
e-mail: aasthac67@gmail.com

B. Rudra
e-mail: bhawanarudra@nitk.edu.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_6

use of feature matching. In this, we try to find some of the distinct features with the help of SIFT that finds the local feature points which are invariant to the points mentioned above. A large number of such features can be extracted from an image by using SIFT. These features are very distinct and provide a high probability of being correctly matched against a large set of features. Along with these features, an image similarity search algorithm is used for image detection. Among the many algorithms for similarity search, we prefer to use the homography estimation search.

This algorithm is used to find the best matches between the two images and find one of the images in another image. These good matches which provide the correct result are caller inliers and the others are called the outliers. A detailed explanation of the method used is explained in the methodology.

The paper is organized into four sections. The first section gives the introduction which explains about the importance of video tracing followed by a literature survey in Sect. 2. Section 3 explains the methodology adopted to solve the problem and the results and finally the conclusion.

2 Literature Survey

A method was described [1] for finding distinctive invariant features from images which can be used for reliable matching. These features are found using SIFT. They have come up approach by using these features for object recognition. The keypoints found by using SIFT algorithm are different and helps to find the correct match from the large dataset [2]. This is achieved by using the high-dimensional vector presentation within the local image [3]. Large number of keypoints can be extracted to identify the small objects in the clutter. This leads to robustness and computation is efficient. Hough transform for identifying clusters were used along with least squares with final verification [4, 5]. Scale-invariant feature transform (SIFT) was used to solve visual tracking problem and consists of five stages: Scale-space extrema detection; keypoint localization; orientation assignment; keypoint descriptor; keypoint matching. The object is selected and SIFT features are extracted.

The Euclidean distance is calculated among the SIFT objects and the frames can be computed. It explains the various steps involved in finding local features using SIFT algorithm and then uses K-d tree algorithm to search each keypoints nearest two feature points and based on these checks if the match is accepted or rejected. This paper provides mathematical approach and steps of video object matching but does not come up with the code for the same. The results are satisfactory with high robustness and achieve real-time performance. This approach described was implemented using VC++ and OpenCV. The result cost was only 0.21 s when SIFT algorithm is used. The results were compared with other methods, and the cost was 0.21 s from 4.3 s. The algorithm proves to be robust and accurate across various scales, orientation and illumination conditions. This plays an important role in tracking and retrieval of the objects from the videos [6–10]. They focused on detecting the duplicate media using an auto dual threshold for the elimination of redundant video frames. After extraction of the features, the local features of F-SIFT were used for content description. It is a content-based copy detection method for the similarity search. The local features extracted using the F-SIFT are large enough causing computational cost. SVD method is used to compute the similarity between the query and the target video. This helps in finding the longest path in the graph that is obtained [11, 12].

Another analysis depends on previous approaches toward object tracking, detection, recognition techniques, feature descriptors and segmentation method based on video frames and various tracking techniques. They have discussed the limitations and future scopes of each of the techniques. The authors used new ideas for the detection of the objects and lead to new research. The paper has analyzed and reviewed the previous approaches using various phases. They have noticed that some of the methods are accurate with more computational complexity [13, 14]. A new algorithm called target image search based on local features (TISLF) compares the target images from the video source using local features. The algorithm consists of three stages, i.e., video segmentation, recognition and estimation. The concept is to cut the video into segments based on the continuity of the video. This algorithm can be used for various applications like tracing the appearance, searching and labeling painting in documents and searching the landmarks and so on. This is compared with a deep learning method and resulted in competitive performance [15, 16].

David G. Lowe presented a new method for image feature generation called SIFT which previously lacked invariance to scale and were more sensitive to projective distortion as well as illumination change [17]. A recognition system is formed using the local image features and these features are invariant to image scaling, translation and rotation. The features will share the similar features with the neurons in the inferior temporal cortex which will be used for the recognition of the primate vision. A staged filter is used for the detection of features at stable points in scale space. The image keys will help to create the local geometric deformations using the blurred images in multiple orientation planes at multiple scales. The keys were used as the input to the nearest neighbor for the identification of the candidate objects. The verification is achieved by finding a low residual least squares from each match for the unknown model. The results show that the recognition can be achieved from the partially occluded images under 2 s of the computation time [17]. The recognition can be improved using the new SIFT features for incorporating the color, texture and edge groupings, sizes and offsets. The indexing and verification allow for the types of scale and rotation invariant features into a single model. The robustness will be maximum by detecting many feature types that rely on indexing and clustering to select the most useful from a particular image [18]. A new video similarity search was proposed for the content-based video retrieval. This method was used to measure the similarity search method that can be used for the computation of the image. This uses the spatial-temporal distribution for the sequences from the video [19]. The similarity will be calculated on the number of similar video components. The results for the large dataset are efficient and effective. An efficient search method based on the index table was used to achieve more efficiency [20–22].

The authors used a correlation between the current image and the background image for the categorization of the pixels as foreground and background using a modified method. The background is considered based on the highly correlation blocks. The remaining block will be considered and pixel-based comparison is performed for the categorization of the foreground or the background. The experiments proved to find the accuracy with speed. More research is in process to improve the accuracy in shape and edge accuracy [23]. The authors discussed a morphological operation for the detection of the objects. The tests were conducted based on the images collected by the static camera. This method focused on obtaining the moving target images from the video frames [24]. A model was proposed for the detection of the objects and the classification by using a combination of TensorFlow and SIFT approach along with the use of deep neural network (DNN). DNN will handle high-dimensional data along with various parameters like a human brain. The obtained results using simulation produced more accurate results than the use of existing methods [25].

Referring to these various papers and open sources related to various image similarity search algorithms it is concluded that most of the papers have come up with a mathematical computation for object detection and proposed various efficient methodologies for image similarity search but none of them have come up with a concrete solution.

3 Proposed Methodology

Feature matching and image detection is an essential and fundamental concept of computer vision which has a wide range of applications extending from motion tracking, object recognition, matching two images to create a panorama, image alignment to 3D reconstruction, robot navigation, etc. For addressing all these wide-ranging applications, we need to find distinct and reliable local feature points which are invariant to object alignment, changes in illumination, i.e., brightness and exposure and 3D camera viewpoint, image scaling and rotation, etc. In order to perform the operation, the following are the steps are considered, i.e., video collection, fey feature detection, object recognition and object tracking. All the steps as shown in Fig. 1 are explained in detail along with the results.

Fig. 1 Methodology

3.1 Video Collection

The video sequence for our analysis can be found from various CCTV cameras. For example, if the analysis of a crime scene is taking place then the video footages from the nearby CCTV cameras within the range of 1 km are taken into consideration and this will become a series of source images for our work, i.e., each frame as a separate source.

3.2 Keypoint Detection

After the collection of videos, our next step involves keypoint detection. This step involves finding local features.

The reasons for using local features are as follows:

1. Local features are robust to occlusion and clutter.
2. They are distinctive, i.e., unique which leads to unambiguous matches.
3. They are huge in number, i.e., a large number of feature points can be found for a single image making it easier for comparison and better detection of an object.
4. They are efficient.

Such feature points can be found using SIFT, i.e., scale-invariant feature transform, an application of OpenCV and computer vision. SIFT is a feature detection algorithm which is used to detect and describe local features in an image. The feature points found by SIFT are much more reliable for matching any other feature detection algorithm.

The following are the four major steps for finding feature points through SIFT:

1. Scale-space extrema detection: This is the first step which involves searching over all scales and image locations. SIFT uses difference of Gaussians for this computation. Difference of Gaussian (DoG) is obtained as the difference of Gaussian blurring of an image with two different alpha, let it be alpha and k * alpha, where alpha acts as a scaling parameter. This process is done for different octaves of the image in Gaussian pyramid. Once the DoG is found, the image is searched for local extrema over scale and space. If it is a local extremum, it is a potential keypoint. It basically means that keypoint is best represented in that scale.
2. Keypoint localization: Once the keypoints are found, they must be processed to get better results. This stage attempts to eliminate some points from the candidate list of keypoints by finding those that have low contrast or are poorly localized on an edge. If the function value at z is below a threshold value this point is excluded. This threshold is called contrast threshold. Edge keypoints are also removed. After this step, only the best keypoints remain.
3. Orientation assignment: This contributes to the stability of matching. An orientation is assigned to each keypoint depending on the scale, and the gradient

Fig. 2 Figure on the left is the input image and the right contains the output image with SIFT feature points marked with colored circles

magnitude and direction are calculated in that region to achieve invariance to image rotation.

4. Keypoint descriptor: A total of 128 bin values are available which are represented as vectors to represent keypoint descriptors.

This is the inner working of SIFT which has been incorporated in the OpenCV's SIFT library and the user of this library does not specifically have to set the value of alpha, i.e., the scaling parameter or any other parameter. The library can be directly used and is known to have accurate results.

The features will be extracted using the following algorithm.

Algorithm:

1. Read the image whose SIFT features are to be computed.
2. Grayscale the image by using OpenCV's inbuilt function.
3. Create a SIFT object.
4. Use this object to detect the keypoints.

The sample output for the input with some brief explanation is as follows:

OpenCV functionalities and the NumPy array are imported. The imported image is converted to a gray image. SIFT object is created so that its functionality can be used. sift.detect() a function which finds the keypoints in the image. cv.drawKeypoints() is an OpenCV function which draws small circles around the locations of the keypoints. Finally, the output image is being saved as shown in Fig. 2.

3.3 Object Recognition

The next step after keypoint detection is object recognition. Features are extracted from the frames and a similarity search is done with the query image and matching results are analyzed, and the results are returned. The similarity search algorithm used is the homography estimation search. The homography estimation search has a wide range of applications like panorama stitching, camera pose estimation for augmented reality, perspective correction, etc. Firstly, if enough number of matches are found between the features of source and the query image, then perspective transform is found between the two planes, i.e., the two images to reduce the error. The k-NN algorithm is used to find the best set of frames that match. This step along with the keypoint detection is repeated for every frame in the video sequence.

Our next step is to compare two images, i.e., the source and the query image and check if the query image is present in the source image. For this step, the feature points for both the query and the source image will be found separately and then compared using homography estimation search. If matches are above the minimum value which is set, then it is concluded that the two images match.

Algorithm for matching two images:

1. Read the source and the query images.
2. Grayscale both the source and the query images.
3. Create a SIFT object.
4. Use this object to compute the keypoints and the descriptors for each of the images separately and store them.
5. Use homography estimation search to compare the two descriptors. Use a k-NN matcher to find the matches between the descriptors.
6. Find the good matches among the matches that are found is Step 5. These should satisfy the condition m.distance < 0.7 * n.distance.
7. Check if the number of good matches is greater than a threshold value which we have taken as 10. If yes then the two images match else they do not match.

Initially, the SIFT features for both the source and the query images are found as explained earlier in keypoint detection steps. The sift.detectAndCompute() directly finds keypoints and descriptors in a single step. des1 and des2 are the descriptors of source and query image, respectively. These descriptors are compared and stored in an array which stores the matches. The array good[] is then used to store all the good matches which satisfy the condition as stated in point 6 of the algorithm. If the size of the array good[] is greater than the minimum match count which is 10 in our case, then it is considered that the images match else the images do not match. Then, showing the matches between the source and query image becomes out output image.

Results:
Case 1: The query and source image are the same (Fig. 3).

Case 2: The query image is contained in the source image (Fig. 4).

In this case as shown in Fig. 4, the dogs face is contained in the dog's image so it shows a match. This shows that the query image is contained in the source image.

Case 3: The source and the query images do not match (Fig. 5).

In this case as shown in Fig. 5, the feature points of the source and query image do not have good enough matches to be claimed as matching images. This is the reason why no matches are shown in the output image.

Fig. 3 Left figure is the input image (source and query image are the same) and the right is the output image (output on terminal: match). These green lines shown in this figure show the features that match between the two images. In this case, there are a lot of matching features as the images are same

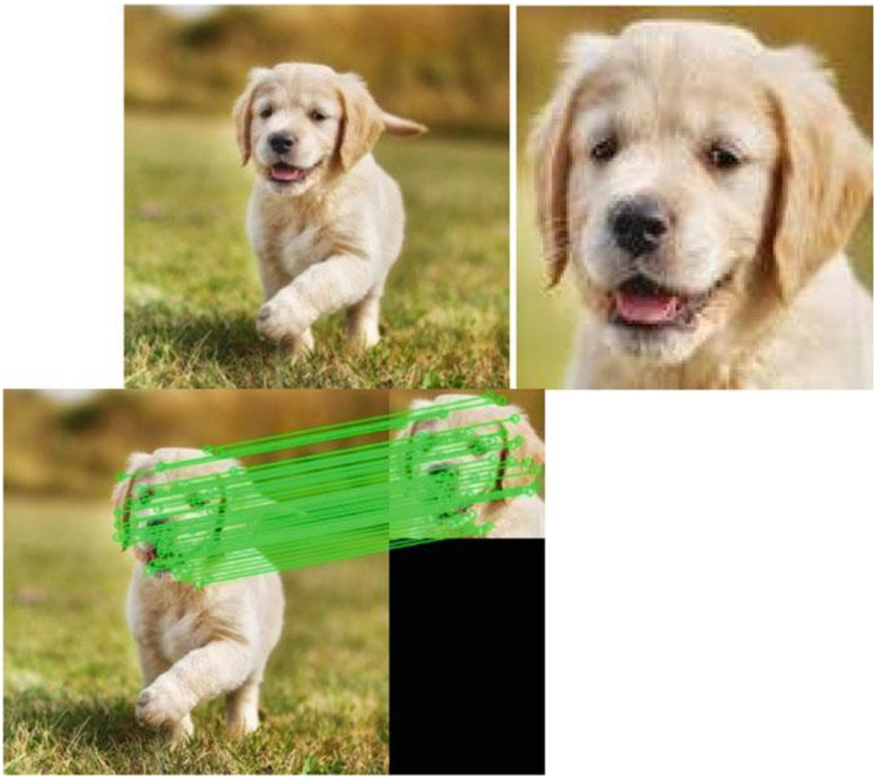


Fig. 4 First image is the source image, the center image is the query image and the right most image is the output image (output on terminal: match)

Fig. 5 First image is the source image, the center image is the query image and the right most image is the output image (output on terminal: no match)

3.4 Object Tracking

In the above case, we were comparing two images to check if they match. Now instead of the two images, we will have a query image and a source video. The last step involves object tracking. In this step, only the video frames from the various video sequences which contain the query image are saved along with the frame number which helps in finding the frames in which our query image is present. In this step, the matches are drawn between the matched source and query image showing the location of the query image in the source frame. Only the frames with a match are a part of the output. In case the query image does not match a frame of the video, that frame is not a part of the output.

Algorithm:

1. Load the query image first and grayscale that image.
2. Create a SIFT object and use that object to detect the keypoints in that image.
3. Load the video and run an infinite loop to work on every frame of the video separately as each frame is an image.
4. First important condition is to check if there are no more frames left in the video then break from the for loop.
5. If step 4 is not true then grayscale the frame and compute the keypoints and store them.
6. The next step is matching the features. In this step, we calculate the descriptors and the keypoints.
7. Another condition to be checked here is if the number of keypoints for both the images is greater than 2 only then continue with the algorithm else move on to the next frame.
8. Use homography estimation to check if the two images match. Use the k-NN matcher to find the matches and then use the same condition as for images to find the good relevant matches.
9. If the number of good matches is greater than the min threshold value, then the two images match else they do not match.

10. If the images match then output this frame number and also map the matches between the source and the query images and save these.
11. Repeat step 3–10 for all the frames of the video sequence.

These steps lead to the final outputs which are the frames which contain the query image. This has a wide range of applications like tracking a VIP person or criminal tracking and much more as the output gives us the frame numbers along with the frames in which the object we wish to find is present. The final step input is a video sequence and the output will be traced path. All the video frames are tested and results are verified and accurate.

4 Conclusion

This paper proposes a method for image tracking which plays a key role for security purposes. The method used to do so is using SIFT keypoints which are specifically useful due to their distinct features. This provides a high probability of accurate matching followed by using a similarly search algorithm, namely homography estimation search that finds the inliers, i.e., the good matches and results in accurate output.

References

1. Lowe, D.G.: Distinctive image features from scale—invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
2. Funt, B.V., Finlayson, G.D.: Color constant color indexing. IEEE Trans. Pattern Anal. Mach. Intell. **17**(5), 522–529 (1995)
3. Brown, M., Lowe, D.G.: Invariant features from interest point groups. In: British Machine Vision Conference, Cardiff, Wales, pp. 656–665 (2002)
4. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scaleinvariant learning. In: IEEE Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, pp. 264–27 (2003)
5. Weber, M., Welling, M., Perona, P.: Unsupervised learning of models for recognition. In: European Conference on Computer Vision, Dublin, Ireland, pp. 1832 (2000)
6. Hu, X., Tang, Y., Zhang, Z.: Video object matching based on SIFT algorithm. In: IEEE International Conference Neural Networks & Signal Processing, Zhenjiang, China, June 8–10 (2008)
7. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the International Conference on Computer Vision, Corfu, Greece, pp. 1150–1157 (1999)
8. Rucklidge, W.J.: Efficiently locating objects using the Hausdorff distance. Int. J. Comput. Vis. **24**(3), 251–270(20) (1997)
9. Grabner, M., Grabner, H., Bischof, H.: Fast approximated SIFT. In: Proceedings of Asian Conference on Computer Vision (ACCV 2006), Hyderabad, India, Springer, LNCS 3851, pp. 918–927 (2006)
10. Shao, H., Svoboda, T., Tuytelaars, T.: HPAT indexing for fast object/scene recognition based on local appearance. In: Proceedings of International Conference on Image Video Retrieval, Urbana, IL, pp. 71–80 (2003)

11. Anju, P.S., Varma, S., Paul, V., Sankaranarayanan, P.N.: Video copy detection using F-sift and graph based video sequence matching. Int. J. Comput. Sci. Inf. Technol. (IJCSIT) **6**(1), 152–158
12. Zhao, W.-L., Ngo, C.-W.: Flip-invariant SIFT for copy and object detection. IEEE Trans. Image Process. **22**(3) (2013)
13. Tiwar, M., Singhai, R.: A review of detection and tracking of object from image and video sequences. Int. J. Comput. Intell. Res. **13**(5), 745–765 (2017). ISSN 0973–1873
14. Susar, R., Dongare, M.: Moving object detection, a succinct review. Int. J. Adv. Res. Comput. Commun. Eng. **4**, 334–336 (2015)
15. Guan, B., Ye, H., Liu, H., Sethares, W.: Target image video search based on local features
16. Holland vs argentina 2014 semi-final full match 1 youtube. YouTube video, July 10 2014. Accessed 1 Aug 2018
17. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the International Conference on Computer Vision, Corfu, Sep (1999)
18. Pope, A.R., Lowe, D.G.: Learning probabilistic appearance models for object recognition in early visual learning. In: Nayar, S., Poggio, T. (eds.) Oxford University Press, pp. 67–97 (1996)
19. Cao, Z., Zhu, M.: An efficient video similarity search algorithm. IEEE Trans. Consum. Electron. **56**(2) (2010)
20. Seo, K.-D., Park, S., Jung, S.-H.: Wipe scene-change detector based on visual rhythm spectrum. IEEE Trans. Consum. Electron. **55**, 831–838 (2009)
21. Ke, Y., Sukthankar, R., Huston, L.: An efficient parts-based near duplicate and sub-image retrieval system. In: Proceedings of the 12th Annual ACM International Conference on Multimedia, pp. 869–876 (2004)
22. Yang, X., Tian, Q., Chang, E.-C.: A color fingerprint of video shot for content identification. In: Proceedings of the 12th Annual ACM International Conference on Multimedia, pp. 276–279 (2004)
23. Ramya, P., Rajeswari, R.: A modified frame difference method using correlation coefficient for background subtraction. Procedia Comput. Sci. **93**, 478–485 (2016). https://doi.org/10.1016/j.procs.2016.07.236
24. Risha, K.P., Kumar, A.C.: Novel method of detecting moving object in video. Procedia Technol. **24**, 1055–1060 (2016). https://doi.org/10.1016/j.protcy.2016.05.235
25. Najva, N., Bijoy, K.E.: SIFT and tensor-based object detection and classification in videos using deep neural networks. Procedia Comput. Sci. **93**, 351–358 (2016). https://doi.org/10.1016/j.procs.2016.07.220
26. OpenCV-Python Tutorials. https://opencv-pythontutroals.readthedocs.io/en/latest/py_tutorials/py_feature2d/py_feature_homography/py_feature_homography.html
27. OpenCV: Introduction to SIFT. https://docs.opencv.org/3.4/da/df5/tutorial_py_sift_intro.html

Real-Time Neural-Net Driven Optimized Inverse Kinematics for a Robotic Manipulator Arm

Jui Shah and Arjun Bhasin

Abstract This paper proposes a method that optimizes the inverse kinematics needed for the trajectory generation of a 4-DOF (degrees of freedom) robotic manipulator arm to give results in real time. Due to the many-to-one mapping of the angle vector which describes the position of the manipulator joints and to the coordinates of the end-effector, traditional methods fail to address the redundancy that exists in an efficient way. The proposed method is singular, and in that it (1) Generates the most optimal angle vector in terms of maximum manipulability, a factor which determines the ease with which the arm moves, for a given end-vector. (2) Proposes a novel approach to inculcate the complexity of dealing with real coordinate system by proposing a machine learning technique that uses neural networks to predict angle vector for practically any end-effector position although it learns on only a few sampled space. (3) Works in real time since the entire optimization of the manipulability measure are done offline before training the neural network using a relevant technique which makes the proposed method suitable for practical uses. (4) It also determines the shortest, smooth path along which the manipulator can move along avoiding any obstacles. To the best of the authors' knowledge, this is the first neural-net-based optimized inverse kinematics method applied for a robotic manipulator arm, and its optimal and simple structure also makes it possible to run it on NVIDIA Jetson Nano Module.

Keywords Inverse kinematics · Neural networks · Robotic manipulator arm · Manipulability · Redundancy · Degree of freedom · Optimization

---

J. Shah (✉)
Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India
e-mail: shahjui2000@gmail.com

A. Bhasin
Fero DWC-LLC, Dubai, United Arab Emirates
e-mail: Arjun@fero.ai

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_7

1 Introduction

A robotic manipulator arm performs several tasks that essentially deal with picking and placing of objects which has applications in a wide variety of fields such as automation, surgery, space and applications with respect to bio-hazardous, radioactive materials or for use in inaccessible places. The movements are adapted from observing how similar manual tasks are handled by a fully-functioning human arm [1, 2]. Just as the upper and the lower human arm are joined by a fixed elbow-joint, a robotic manipulator arm is a set of joints connected by links of certain thickness which can rotate around the joints to reach different points within the operational space bounded by constraints, if any (Fig. 1).

The order of the vector space it operates on is called the operational space, and the number of joints it consists of is its degree of freedom. This paper deals with a 4-DOF manipulator arm in a two-dimensional vector space. Since, the number of degrees of freedom is more than the operational space, it will result into the redundancy of the system in a way that for one end-effector position, and there would be many angle vectors. For general physical systems, redundancy might help in more reliability of the systems, but it increases the cost and complexity of the system design causing failure in performance [3]. Hence, it becomes important to remove the redundancy which causes failure of the neural network as its error significantly increases if it trains on a redundant training set.

Redundancy optimization is discussed in great detail in [4–6]. We solve the redundancy by using manipulability factor of the robotic manipulator arm which measures the extent to which it moves freely [7, 8]. The idea of manipulability is described in Sect. 3.

In the field of robotic kinematics, forward kinematics and inverse kinematics are the two main methods of defining configuration. Forward kinematics maps the operational space with the angular space (DOFs) and the inverse kinematics vice-versa. Since, forward kinematics cannot be simplified to a linear mapping, and since the destination and start positions are generally known in practical applications, the inverse kinematics approach is more appropriate to use; however, it is even more complex. It is described in Sect. 2.

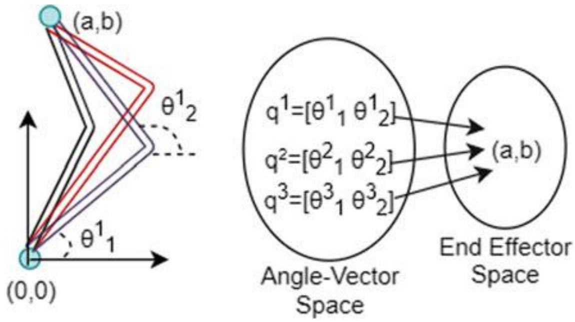


Fig. 1 Many-to-one mapping

Classically, inverse kinematics equations were solved using numeric approach by Newton's method as discussed in [9], but the iterative technique with high computational time and cost is not feasible for real-world applications. Later, Jacobian transpose methods and pseudo-inverse methods were introduced, but the calculation of the Jacobian and the pseudo-inverse matrices proves to be analytically difficult [10]. With the great comeback of neural networks, numerous approaches to solve inverse kinematics problem have been proposed as in [11–14]. But since the mapping from the operational space to the angular space is many–one, the neural networks fail to predict the mapping accurately resulting in huge errors. Reference [15] discusses various methods of using modified back-propagation for solving inverse kinematics using neural networks.

The main reason of failure of the previous techniques like in [16] is that the solution may not exist or a slow convergence for some initial values. It also has limitations for some of the target position and is yet to be implemented for higher DOF manipulator arm.

This paper proposes the use of manipulability measure to optimize the training set which we then feed into the deep neural network (DNN) described in Sect. 4. The optimization is done offline, and hence, the many–one mapping from the operational space to the angular space is turned into one–one mapping. This reduces the error caused by the neural network significantly. The use of the neural network results in fast real-time calculations and results. It also proposes calculation of obstacle cost and minimum curve length cost from each trajectory randomly generated using intuition derived from STOMP algorithm [17] in Sects. 5 and 6, respectively, instead of determining by calculus of variation which has high analytical cost, and then chooses the minimal cost trajectory which is the most optimal.

Results showing high efficiency are given in Sect. 7. The code was developed in Python.

2 Kinematic Modelling

2.1 Direct Kinematic Modelling

Let our frame of reference be the x-axis and the y-axis in the first quadrant and the robotic manipulator arm having 4-DOF be fixed at the point (0, 0) as shown in Fig. 2. Suppose the arms be of length l_1, l_2, l_3, l_4, respectively. Now, the configuration of the robot at a particular instance is given by the angle vector $q \in \Re^4$ or the position of end-vector of each joint $\xi \in \Re^2$:

$$q = [\theta_1\ \theta_2\ \theta_3\ \theta_4], \qquad \text{and} \qquad \xi = [\xi_1\ \xi_2\ \xi_3\ \xi_4] \tag{1}$$

where $\xi_1 = [x1 \quad y1]^T, \xi_2 = [x2 \quad y2]^T, \xi_3 = [x3 \quad y3]^T, \xi_4 = [xEE \quad yEE]^T$.

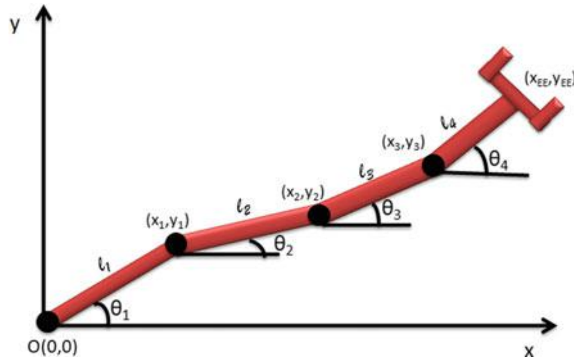


Fig. 2 A robotic manipulator arm in 2D plane

Forward Kinematics Given the angle vector, the end-position vector is be calculated using forward kinematics

$$x_1 = l_1 c_1, \quad \text{and} \quad y_1 = l_1 s_1$$

where c_1 and s_{34} denote $\sin\theta_1$ and $\cos\theta_3 + \theta_4$, respectively. Similarly, all x_i and y_i can be calculated by taking projections of the arm-length along the respective coordinate axis. Since, l_1, l_2, l_3, l_4 are known as $\xi_4 = f(q)$:

$$\begin{bmatrix} x_{EE} \\ \\ y_{EE} \end{bmatrix} = \begin{bmatrix} l_1 c_1 + l_2 c_{12} + l_3 c_{123} + l_4 c_{1234} \\ \\ l_1 s_1 + l_2 s_{12} + l_3 s_{123} + l_4 s_{1234} \end{bmatrix} \tag{2}$$

The matrix contains sine and cosine terms which makes the mapping nonlinear, and hence, writing it in the form of $\xi_a = A(q)q_a$ is impossible where $A \in \mathcal{R}^{m\times n}$. Consequently, forward kinematics is not preferred over instantaneous model.

2.2 *Instantaneous Direct Kinematic Model*

Since it is easier to deal with linear systems, we take the time derivative of the forward kinematic model $\xi = f(q)$ to get the instantaneous direct kinematic model which gives us a linear mapping between the joint-angle velocity vector $\dot{q}$ and the end-effector velocity vector $\dot{\xi}$.

$$\dot{\xi} = \frac{\partial f(q)}{\partial q}\dot{q} = J(q)\dot{q} \tag{3}$$

where $J = \frac{\partial f}{\partial q}$ is the $m \times n$ Jacobian matrix.
This implies that

$$\dot{q} = J^{-1}(q)\dot{\xi} \tag{4}$$

Equation 4 defines the inverse mapping from joint-angle velocity vector to the end-effector velocity vector that describes inverse kinematics which has real-world applications since we know the end-coordinates of a vector rather than its angular positions.

Traditionally, to calculate inverse of a Jacobian matrix, we use the right pseudo-inverse (Moore–Penrose inverse) of a matrix:

$$A^+ = A^T(AA^T)^{-1} \tag{5}$$

However, directly computing the inverse of the matrix may result in numerical instabilities while calculating the inverse of $A^T A$ if the matrix A is ill-conditioned. Instead, it is the best to use the singular value decomposition of A to find out the pseudo-inverse of A. The singular value decomposition (SVD) of the matrix A is defined as follows

$$A^+ = VS^{-1}U^T \qquad \text{when} \qquad A = USV^T. \tag{6}$$

Jacobian The Jacobian matrix is defined as:

$$J = \begin{pmatrix} \frac{\partial x_{EE}}{\partial \theta_1} & \frac{\partial x_{EE}}{\partial \theta_2} & \frac{\partial x_{EE}}{\partial \theta_3} & \frac{\partial x_{EE}}{\partial \theta_4} \\ \frac{\partial y_{EE}}{\partial \theta_1} & \frac{\partial y_{EE}}{\partial \theta_2} & \frac{\partial y_{EE}}{\partial \theta_3} & \frac{\partial y_{EE}}{\partial \theta_4} \end{pmatrix} \tag{7}$$

We know that,

$$\frac{\partial x_{EE}}{\partial \theta_1} = -l_1 s_1 - l_2 s_{12} - l_3 s_{123} - l_4 s_{1234} = -y_{EE}, \text{ and} \tag{8}$$

$$\frac{\partial y_{EE}}{\partial \theta_1} = l_1 c_1 + l_2 c_{12} + l_3 c_{123} + l_4 c_{1234} = x_{EE} \tag{9}$$

By similar calculations of the other partial derivatives, we get the Jacobian matrix as

$$J_4 = \begin{pmatrix} -y_{EE} & y_1 - y_{EE} & y_2 - y_{EE} & y_3 - y_{EE} \\ x_{EE} & x_{EE} - x_1 & x_{EE} - x_2 & x_{EE} - x_3 \end{pmatrix} \tag{10}$$

3 Manipulability

Manipulability is the ease with which the end-effector of an arm can move arbitrarily in any direction. Let us take into consideration a general case of a manipulator with n degrees of freedom with a m-dimensional operational space. If $m < n$, the manipulator will be redundant, where $n - m$ are the redundant degrees of freedom.

It is discussed more precisely in [18, 19] for a mobile manipulator. Consider the m-dimensional Euclidean space of the operational space parameters. Now, at a given configuration q_a,

$$\sqrt{\dot{q_1}^2 + \dot{q_2}^2 + \dot{q_3}^2 + \dot{q_4}^2} \le 1 \quad \text{, i.e.,} \quad \|\dot{q}\| \le 1 \quad \text{forms an ellipsoid.} \tag{11}$$

From Eq. 5, we get $\dot{q_a} = J_a(q_a)^+ \dot{\xi_a}$, where $J_a(q_a)^+$ is the pseudo-inverse of $J_a(q_a)$. Hence, the set of velocities satisfying $\|\dot{q}\| \le 1$ is given by

$$\xi_a^T (J_a(q) J_a(q)^T)^{-1} \xi_a \le 1. \tag{12}$$

This is called the manipulability ellipsoid because it describes the arm's ability to move in any arbitrary direction. The end-effector moves at high speed along the major axis of this ellipsoid and at low speeds along the minor axis of the ellipsoid. Hence, higher the manipulability, better the functionality.

Also, given singular value decomposition (SVD) of the analytical Jacobian matrix $J_a(q_a)$, where $U_a(q_a) = [u_1\ u_2\ \cdots\ u_m]$ and $V_a(q_a)$ are orthogonal and $\Sigma_a(q_a)$ is the ordered singular values' diagonal matrix with $\sigma_{a1} \ge \sigma_{a2} \ge \cdots \ge \sigma_{am}$ is

$$J_a(q_a) = U_a(q_a) \Sigma_a(q_a) {V_a}^T(q_a) \tag{13}$$

The axes of the manipulability ellipsoid are given by $\sigma_{a1} u_1, \sigma_{a2} u_2, \ldots, \sigma_{am} u_m$.

We define a measure of manipulability as w_1, proportional to the volume of the ellipsoid, and as follows:

$$w = \sqrt{det(J(q) J(q)^T)}. \tag{14}$$

$J(q)J(q)^T$ is a positive semi-definite matrix, and hence, this measure is suitable. We add constraints to the space to eliminate singularities where $w = 0$.

4 Neural Network

The aim is to get an optimal angle vector q^* for a particular ξ_4. The neural network is trained upon the data of ξ_{train} and corresponding q^*_{train}. Data is generated in three main steps:

1. By discretizing q^*_{train} in the first quadrant, i.e., q_{θ_0} ranges from 5° to 90° and $q_{\theta_1}, q_{\theta_2}, q_{\theta_3}$ range from 5° to 360°; all with a step size of 5°.
2. Calculating the end-effector position by forward kinematics and storing the angle vector corresponding to that end-effector.
3. Finding the optimal angle vector by finding maximum manipulability measure provided by the given set of angle vectors corresponding to each end-effector

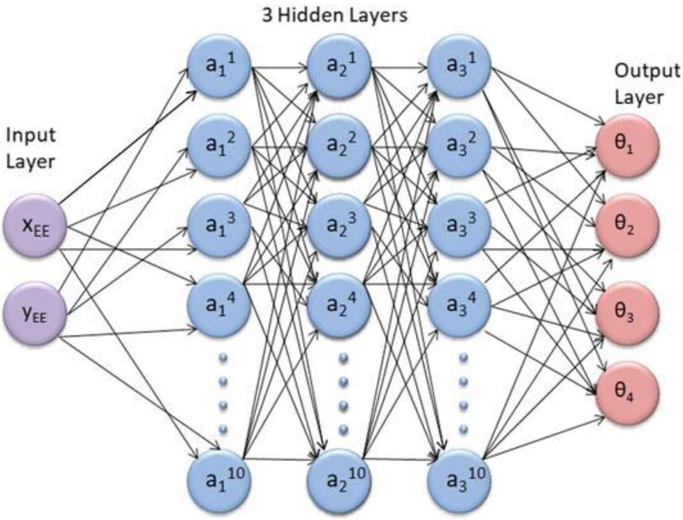


Fig. 3 Neural network

position by using Cobyla optimizer. This network has 2 inputs: xEE, yEE, 3 hidden layers each with 10 neurons and the output layers that contain 4 nodes: $\theta_1, \theta_2, \theta_3, \theta_4$ which are the optimal angles for the given end-effector coordinates. ADAM optimizer is used with a learning rate of 0.01 while minimizing the loss function (mean squared error) (Fig. 3).

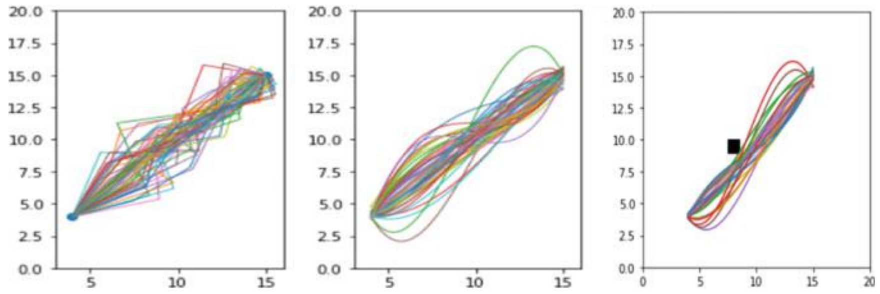
Using neural networks makes real-time prediction very fast, and it also inculcates the entirety of the real coordinate axis which cannot be done by a normal look-up table that stores angle-effector pairs of discrete data.

5 Generating Random Trajectories

We generate random trajectories by deriving intuition from STOMP algorithm [17]. Random values of epsilon ϵ are generated from a Gaussian distribution having co-variance R^{-1} such that $R = AA^T$ where A is a finite differencing matrix chosen in such a way that the resultant curve is a smooth curve. Then, we add the values of our random variable $epsilon$ to the mean values to get randomly generated points.

$$x = \theta_x + \epsilon_x, \quad \text{and,} \quad y = \theta_y + \epsilon_y \tag{15}$$

The generated points are then joined using curve-fitting algorithm such as polynomial fit or spline fitting (Fig. 4).



(a) Linearly connected random points (b) Curve fitting (c) Obstacle Avoidance

Fig. 4 Generated estimated paths

6 Cost Function

6.1 Obstacle Detection

Since we want to avoid obstacles under any circumstances, if a path collides with an obstacle, we make its cost as infinity else 0.

It is very important to note that we do not check the collision of the path with the obstacle, rather the collision of the robotic manipulator arm following the path with the obstacle externally.

We check obstacle collision externally with the robotic manipulator arm. First step is approximating the obstacle to the nearest polygon and dividing it into triangles. Each link of robotic manipulator arm is also divided into triangles. Next step is to check whether triangle-triangle collision occurs which reduces the problem to simply checking if an edge of triangle is inside another, i.e., if a point is inside a triangle or not using dot and cross products.

6.2 Minimal Curve Length

The length of the path followed by the manipulator should be minimum, and since, we cannot use Euclidean distance, we calculate the curve length from $x = a$ to $x = b$:

$$L = \int_a^b \sqrt{1 + \left(\frac{\partial y}{\partial x}\right)^2} \partial x \tag{16}$$

6.3 Total Cost

Let, C is the cost function,

$$C = O + \lambda L \tag{17}$$

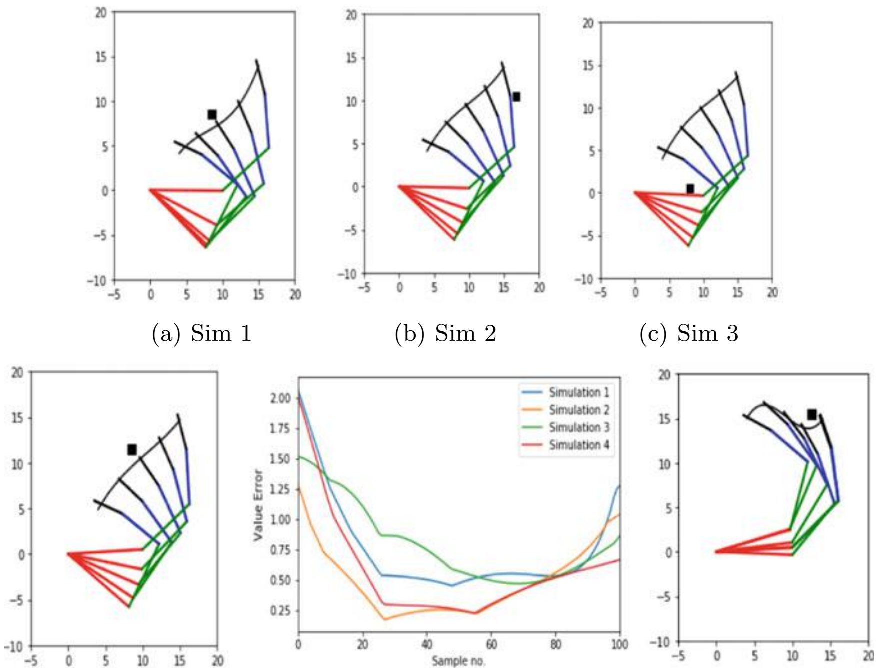
where O is the obstacle cost and L is the curve length cost.

7 Results and Simulation

The specifications of the simulated robotic manipulator arm used are: $l_1 = 10$ units, $l_2 = 8$ units, $l_3 = 6$ units and $l_4 = 4$ units. The manipulator has been fixed as $O(0, 0)$, the initial position of its end-effector is $A(4, 4)$,and the destination is set to be $B(15, 15)$. First, random points are generated using the formula

$$(x_i, y_i)_j = (\theta_x + \epsilon_x, \theta_y + \epsilon_y), \tag{18}$$

where i spans the entire path and j spans the number of paths.

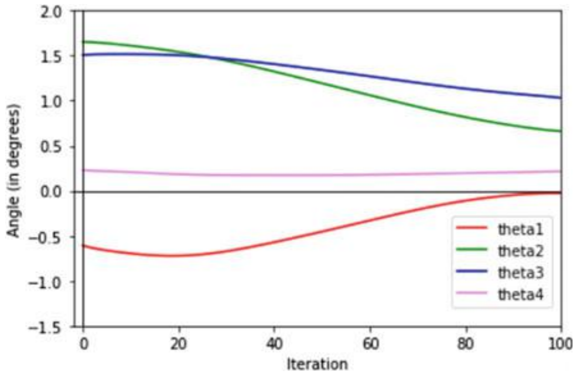


(a) Sim 1 (b) Sim 2 (c) Sim 3

(d) Sim 4 (e) Error in sim (f) Sinusoidal Sim

Fig. 5 Various simulations and their error at every instant

Fig. 6 Value of theta in Sim2



The neural network described in Sect. 4 is used to generate the configuration of the manipulator arm which is fitted with a batch-size of input as 128 and number of epochs as 300.

The generated path clearly avoids the obstacle and is of minimal length. Also, the error generated in reaching the final position can be practically ignored. The results obtained are highly efficient, accurate and fast.

All the simulations of the proposed algorithm as shown in Fig. 5 were developed using Jupyter Notebook Python3.

From Fig. 5e, it can be inferred that the error is high at the end-positions because of singularities in the constrained end-effector space. Also, Fig. 6 depicts the change of value of the angle vector in a particular simulation, which varies smoothly and hence validates our method that using neural network provides a solution for the entire real axis because of continuity/smoothness.

The entire model can be executed on NVIDIA Jetson Nano Module.

References

1. Manipulator (device). https://en.wikipedia.org/wiki/Manipulator_(device)
2. Kaur, K.: Basic robotics: what is a robotic manipulator? (2013). http://www.azorobotics.com/Article.aspx?ArticleID=138
3. White papers-redundant system basic concepts, national instruments (2008). http://www.ni.com/white-paper/6874/en/
4. Nakamura, Y.: Advanced Robotics: Redundancy and Optimization. Addison-Wesley (1991)
5. Nakamura, Y., Hanafusa, H.: Optimal redundancy control of robot manipulators. Int. J. Robot. Res. (1987)
6. Yoshikawa, T.: Foundation of Robotics: Analysis and Control. MIT Press (1990)
7. Yoehikawa, T.: Translational and rotational manipulability of robotic manipulators. In: American Control Conference (1990)
8. Yoehikawa, T.: Dynamic manipulability of robotic manipulators. In: IEEE International Conference on Robotics and Automation (1985)

9. Kurfess, T.R.: Robotics and Automation Handbook. CRC Press LLC (2000)
10. Buss, S.R.: Introduction to inverse kinematics with Jacobian transpose pseudoinverse and damped least squares methods (2004)
11. Duka, A.-V.: Neural network based inverse kinematics solution for trajectory tracking of a robotic arm (2013)
12. Dinh, B.H.: Approximation of the inverse kinematics of a robotic manipulator using a neural network (2009)
13. Murray, R.M.: A Mathematical Introduction to Robotic Manipulation. CRC Press (2017)
14. Serrezuela, R.R., Chavarro, A.F., Cardozo, M.A.T., Toquica, A.L., Martinez, L.F.O.: Kinematic modelling of a robotic arm manipulator using Matlab. ARPN J. Eng. Appl. Sci. (2017)
15. Tejomurtula, S., Kak, S.: Inverse kinematics in robotics using neural networks. Information Sciences (1998)
16. Srisuk, P., Sento, A., Kitjaidure, Y.: Inverse kinematics solution using neural networks from forward kinematics equations. In: 9th International Conference on Knowledge and Smart Technology (KST) (2017)
17. Kalakrishnan, M., Chitta, S., Theodorou, E., Pastor, P., Schaal, S.: STOMP: stochastic trajectory optimization for motion planning (2011)
18. Martinez, J.: Manipulability ellipsoids in robotics
19. Rozo, L., Jaquier, N., Calinon, S., Caldwell, D.G.: Learning manipulability ellipsoids for task compatibility in robot manipulation. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2017)

A Deep Learning Technique to Countermeasure Video-Based Presentation Attacks

Thomas Danner, Prosenjit Chatterjee, and Kaushik Roy

Abstract Presentation attack detection (PAD) on faces is crucial to countermeasure face recognition system from a security breach. The increase in convenience to our lives is not without its own additional avenue for exposure to a security breach. The presentation attack (PA), or spoofing attack is the act of using artificial/synthetic materials to get unauthorized access into a secured system. For example, an individual could use a 3D printed mask, or even a digital or printed photograph to circumvent a facial recognition authentication system, for the case of iris-detection an intruder could use custom contact lenses, and for fingerprints, digital prosthetics to penetrate a secured system. Previously, many deep learning approaches utilize a very deep complex model to handle face PAD. To countermeasure the above issue, in this paper, we apply a deep learning approach to mitigate the presentation attack. In our proposed approach, we implement a lightweight 'modified-AlexNet' and obtained the highest test accuracy of 99.89% on the Spoof in the Wild (SiW) dataset.

1 Introduction

The two categories of presentation attacks (PA) can roughly categorized as direct attacks and indirect attacks [1]. The direct face PA includes printing a face on paper (print attack), re-iterating a face video on a digital device (replay attack), using a mask to pretend as the real user (mask attack), etc. An indirect face PA method involves

T. Danner (✉) · P. Chatterjee · K. Roy
Department of Computer Science, North Carolina A & T State University, Greensboro, NC 27411, USA
e-mail: tdanner@aggies.ncat.edu

P. Chatterjee
e-mail: pchatterjee@aggies.ncat.edu

K. Roy
e-mail: kroy@ncat.edu

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_8

targeting system architectures and the in between the handshaking point where subsystems are loosely coupled. In this study, we are focused on countermeasure the direct attacks on the face PA to authenticate the secure system.

Recently, deep learning approaches have been applied to mitigate the PAs and demonstrate the promising performances. This research attempts to create a deep-learning driven solution to identify direct presentation attacks during the authentication process, particularly, with respect to facial recognition/authentication. By first using a simple approach of supervised classification with a substantial dataset to train a shallow model, such as AlexNet, it is expected that it can yield high accuracies on a multiclass framework with the expectation to expand the algorithm that intruders cannot intercept.

2 Related Work

Prior to the deep learning approach to facial anti-spoofing, most anti-spoofing methods relied on handcrafted filters such as LBP, HOG, and SIFT in order to verify the images presented to a sensor and were typically classified with traditional machine learning (ML) techniques such as SVM. Schwartz et al. [2] researched texture, shape, and color of the face region and applied the partial least square (PLS) classifier for deciding whether a biometric sample was invalid or original; however, this application was exclusive to the print attack PA type. Menotti et al. [3] explored the relationship between the CNN architecture and the filter weights to optimize a CNN for handling PA. Their experiments found a 98.93% accuracy utilizing a simple CNN architecture and good understanding of filter optimization. Freitas Pereira et al. [4] proposed an anti-spoofing solution based on the dynamic texture, a spatio-temporal version of the original LBP. Results showed that LBP-based dynamic texture description has a higher effectiveness than the original LBP.

Pinto et al. [5] first explored video-based face spoofing detection. They introduced visual rhythm analysis to capture temporal information on face spoofing attacks. Yang et al. [6] provided further work extending into the facial anti-spoofing in regards to video-based classification. In their methodology, there were three primary focuses that accounted for the entire anti-spoofing system: face localization, spatial augmentation, and temporal augmentation. Particularly, the latter two here are of great interest to the methods of anti-spoofing, with spatial augmentation evaluating carefully at the areas beyond the face that is being authenticated. This involves looking into the background details and carefully surveying if the image, based on its environment, is genuine. The group research into temporal augmentation had a large part in future anti-spoofing with its focus on studying the use of feeding the convolutional neural network (CNN) multiple frames vs single frames.

Nguyen et al. [7] proposed the integration of popular handcrafted filters with the deep learning approach. In their studies, they used a modified VGG Net-19 architecture in tandem with a multi-level LBP (MLBP) to solve the PAD problem. The MLBP filter used obtains feature vectors containing 3732 components for each facial image allowing more texture to be extracted. Using the trained VGG Net-19 model, they extracted a 4096 component image feature vector and classified their results with SVM. With extensive experiments on dataset distribution, they were able to get their percent error as low as 1.696%.

Liu et al. [8] were not the first to explore the use of remote photoplethysmography (rPPG); however, they used the prior research to apply to the applications of facial anti-spoofing. The rPPG plays a vital role in detecting the liveliness of a subject exposed to sensor. They were able to train their dataset using the videos by combining CNN and RNN applications, the CNN to train with and the RNN was given feature maps with rPPG supervision to check for temporal variability among the video frames provided. One of the primary innovations of this experiment was in the RNN's ability to predict the rPPG signals provided in the future frames from the data it is already been provided. Their group was also the originator of the Spoof In The Wild (SiW) dataset [9] (Figs. 1 and 2).

Liu et al. [8] also utilized depth supervision techniques in order to further aid the classification of video given to the CNN. The depth maps were used with a depth loss function with CNN in order to establish three-dimensional elements of the data provided to the CNN and compared against a projected two-dimensional depth map generated by dense face alignment (DeFA) methods. These depth maps were used as a ground truth for the purpose of their research.

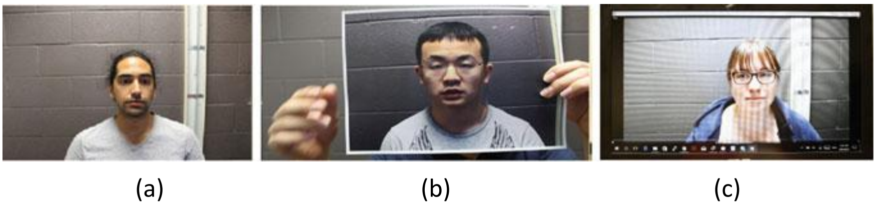


Fig. 1 Visual demonstration of the three classes from SiW dataset in order: **a** live, **b** print, and **c** replay

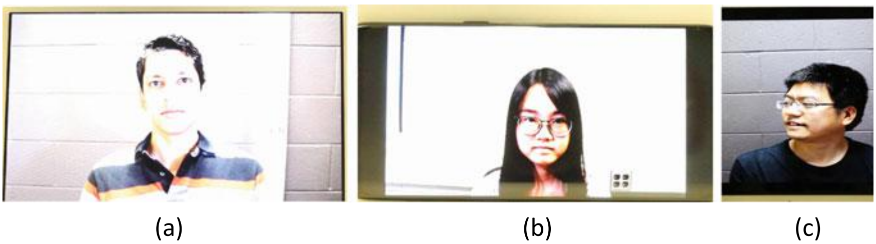


Fig. 2 Further examples of the replay attack replayed on devices: **a** Asus MB168b Monitor, **b** Samsung Galaxy S8 Phone, **c** iPad Pro Table

3 Methodology Used

Very few studies have been conducted on extensive datasets such as SiW. In many cases, we see CASIA-MFSD [10] and replay attack [11] used as the primary benchmark. It was our goal to use a larger benchmark that can provide a higher number of samples to test the effectiveness in speed and accuracy of a lightweight model. It is expected that by reducing the complexity of the model to a tolerable level, we can achieve faster speed for both training and validation without sacrificing the accuracy of the output. As previously discussed, a simple binary classification approach to anti-spoofing will not provide a CNN with enough data to reliably classify a frame in video as a particular PA type. Our proposed approach was to use a lightweight 'modified-AlexNet' multiclass approach that can handle two PA types and the live class provided with the SiW dataset [9], and also uses the origin videos for training the CNN as a baseline. We predicted that this would not only increase the overall prediction accuracy, but also increase the accuracy for particular replay and print attacks which do not disclose the beveled edges of a photograph or a screen used for replay attacks.

The SiW dataset [9] is comprised of 165 subjects. Each subject has 8 live videos each video consisting of the subject conducting several ranges of movement with their face and head, and between 12 and 20 videos of spoofing between the print and replay PA types. The replay attacks are replayed on four different devices and recaptured with two different sensors adding diversity to how they appear to the CNN. Each video is approximately 20–30 s long recorded at 30 frames per second (FPS) at 1080p resolution. This gives us approximately 600–900 frames per video that the CNN can sample from. Due to memory constraints with our machines, we created an algorithm to sample every 15th frame giving us approximately 40–60 frames per video, or one frame per half second of video.

For our methodology, we extracted our frames from the video and re-sized the images appropriately as well as converting to grayscale for our CNN. During this time, we can also instantiate an adjacent array to perform a classic supervised classification method for our model to train and validate with. After all pre-processing was conducted, we trained a 'modified-AlexNet' (Fig. 3) to extract the dataset features, using a rectified linear units (ReLU) activation function to compute the loss, followed

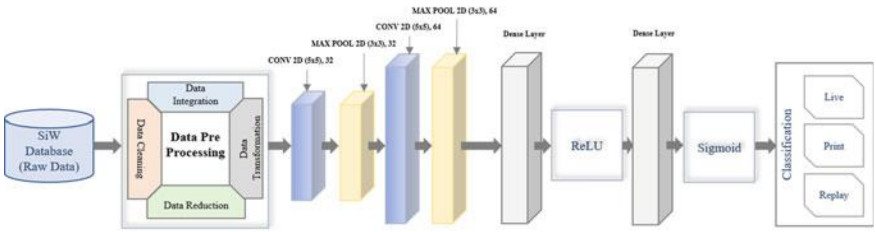


Fig. 3 Data flow diagram including a detailed look at the composition of the 'modified-AlexNet' architecture

by an additional pairing of a dense layer and a sigmoid activation function layer. For our model, we reduced the depth of the AlexNet to resolve some complexity issues and to increase the speed of extraction across the dataset.

4 Experimental Results

With the large size of the SiW dataset, we used subsets of the dataset in order to classify the entire dataset and avoid issues of memory overflow. The subsets comprised of between 40 and 50 subjects with each subject's videos being organized between live sample, and print, and replay PA types. Each subset generated very high accuracies (Fig. 4) and acceptable losses (Fig. 5).

Subset 1 consisted of 11,134 frames resulting in no false positives, and only 3 false negatives. It produced 99.97% accuracy, 100% precision, and 99.97% sensitivity. Subset 2 consisted of 9,344 frames yielding 25 false positives, and 8 false negatives. Subset 2 produced 99.64% accuracy, 99.73% precision, and 99.91% sensitivity. The third subset consisted of a total of 9,486 frames yielding 2 false positives and 2 false negatives. The accuracy produced was 99.95%, precision was measured at 99.97%, and sensitivity was also measured at 99.97%. The final subset consisted of

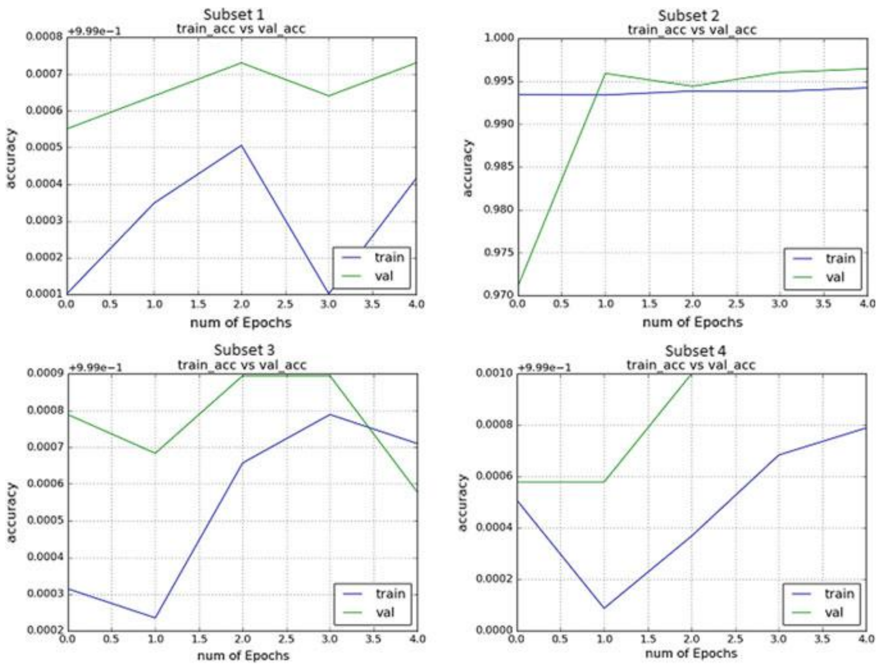


Fig. 4 Training and validation accuracies per subset (accuracy vs. no. of epochs) for subset 1, 2, 3, and 4

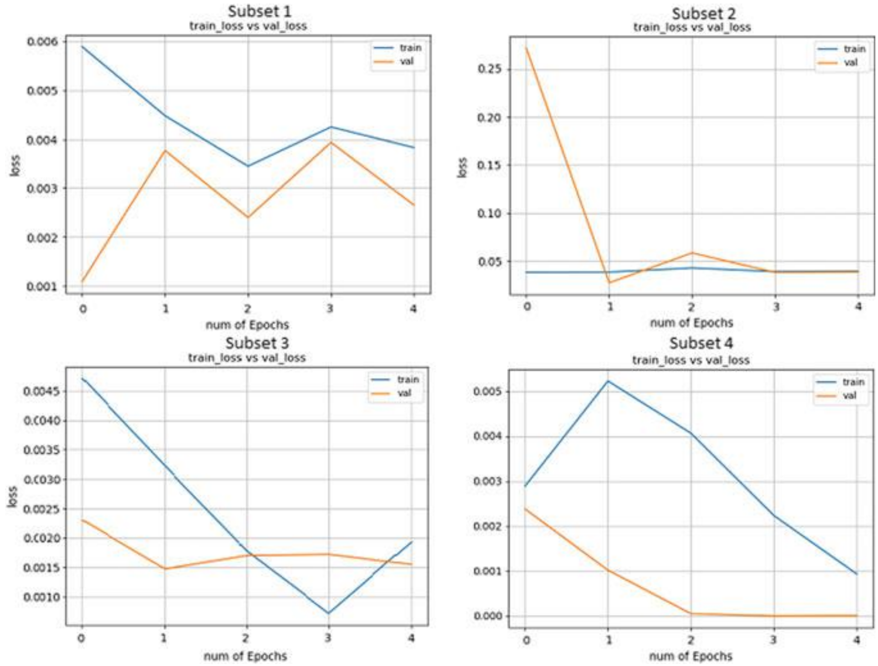


Fig. 5 Computed training and validation losses per subset (loss vs. no. of epochs) for subset 1, 2, 3, and 4

7,131 frames producing no false positives or false negatives. This gave the subset a measured accuracy, precision, and sensitivity of 100%. It should be noted that Subset 4 was a more skewed subset than the others. There are no print attack cases contained in Subset 4 due to an imbalanced distribution of material in the SiW dataset toward the end of the subjects recorded. Although this does simplify the final subset to a binary classification, we believe it has little impact on the overall findings for our testing.

The tests provided an overall average of 99.89% accuracy. With only 40 out of 30,795 samples misidentified, 27 of which were false positives and only 13 were identified as false negatives (Fig. 6 and Table 1). The precision was measured at 99.92%. Sensitivity was measured at an average of 99.96%.

This strongly supports the notion of a simple supervised classification implementation providing high effectiveness over a large sample of data. However, the similarity between many samples for one subject could influence these results.

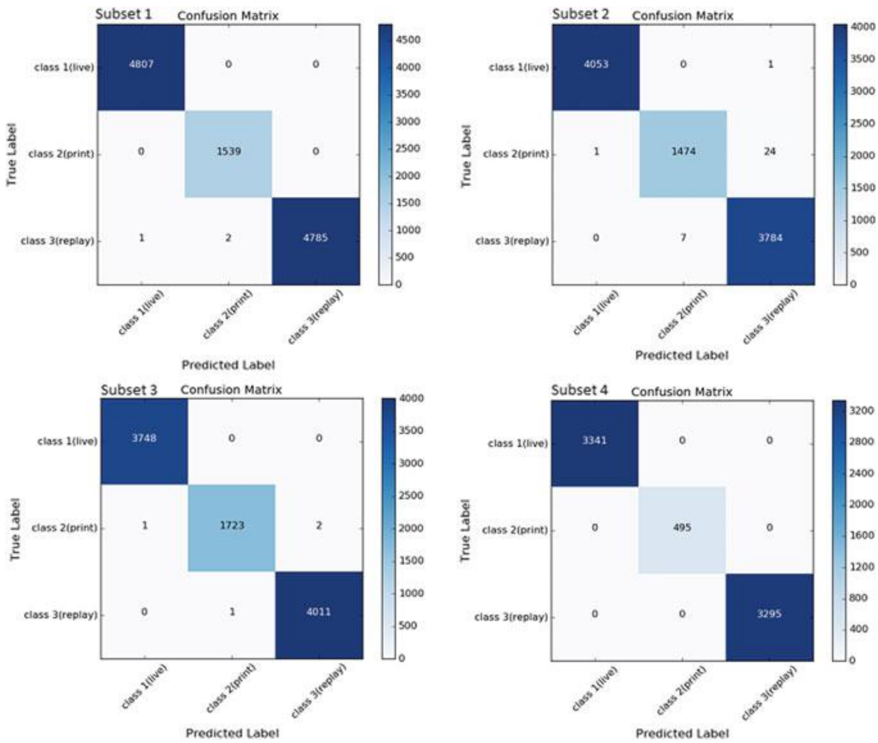


Fig. 6 Post-classification confusion matrices per subset (subset 1, 2, 3, and 4)

Table 1 Statistics

Subsets	Number of samples	True positives	False positives	False negatives	Accuracy	Precision	Sensitivity
Subset 1	11,134	11,131	0	3	0.99973	1	0.99973
Subset 2	9,344	9,311	25	8	0.99646	0.99732	0.99914
Subset 3	9,486	9,482	2	2	0.99957	0.99978	0.99978
Subset 4	7,131	7,131	0	0	1	1	1
Overall	37,095	37,055	27	13	**0.99982**	**0.99927**	**0.99964**

5 Conclusion and Future Work

Initially, we predicted that a low complexity lightweight CNN could still yield high accuracies in a, relative to the benchmark, short period of time. As an outcome, our modified AlexNet yielded an overall 99.9% accuracy. The training and validation times took approximately 30 min per epoch. Relative to the size of the benchmark, the time lapsed is reasonable. In conclusion, we believe that the overall outcome was good, but there's still more that can be modified and improved upon to yield faster speeds and higher accuracies.

Going forward, we would like to experiment with a few more directions of our study. We would like to test the effectiveness of using a Softmax activation function for computing loss as opposed to the ReLU activation function. More recent studies have shown a high effectiveness for multiclass cases using the Softmax activation function. Additionally, we would like to progressively increase complexity to our algorithm to include the potential use of a meta-learner approach to continue to increase the speed at which a CNN can accurately identify PA types. We believe that a combination of meta-learner in addition to some of the aforementioned methods for PA detection could further the possibilities for detection techniques to come.

Acknowledgements We would like to acknowledge the support from the National Science Foundation (NSF).

References

1. Galbally, J., Marcel, S.: Face anti-spoofing based on general image quality assessment. In: 2014 22nd International Conference on Pattern Recognition, pp. 1173–1178. IEEE (2014)
2. Schwartz, W.R., Rocha, A., Pedrini, H.: Face spoofing detection through partial least squares and low-level descriptors. In: IEEE International Joint Conference on Biometrics (IJCB), pp. 1–8 (2011)
3. Menotti, D., et al.: Deep representations for iris, face, and fingerprint spoofing detection. IEEE Trans. Inf. Forensics Secur. **10**(4), 864–879 (2015). https://doi.org/10.1109/tifs.2015.2398817
4. Freitas Pereira, T., Komulainen, J., Anjos, A., De Martino, J., Hadid, A., Pietikainen, M., Marcel, S.: Face liveness detection using dynamic texture. EURASIP J. Image Video Process. (1), 2 (2014)
5. Pinto, A., Pedrini, H., Schwartz, W.R., Rocha, A.: Video-based face spoofing detection through visual rhythm analysis. In: Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 221–228 (2012)
6. Yang, J., Lei, Z., Li., S.: Learn convolutional neural network for face anti-spoofing. arXiv:1408.5601v2 [cs.CV] (2014)
7. Nguyen, D., Pham, T., Baek, N., Park, K.: Combining deep and handcrafted image features for presentation attack detection in face recognition systems using visible-light camera sensors. Sensors **18**(3) (2018)

8. Liu, Y., Jourabloo, A., Liu, X.: Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 389–398 (2018)
9. Liu, Y., Jourabloo, A., Liu X.: Spoof in the wild (SiW) face anti-spoofing database
10. Wen, D.: Mobile face spoofing database (CASIA-MFSD) database (2014)
11. Marcel, S., Anjos, A.: Replay-attack database (2012)

Optimization of Loss Functions for Predictive Soil Mapping

Harsh Kotta, Karan Pardasani, Megha Pandya, and Ranendu Ghosh

Abstract Loss function is an integral part of any machine learning algorithm. Loss function is used to measure the current performance of machine learning model during training. Choosing a relevant loss function is therefore important. A better loss function can drive the model towards a good sub-optimal solution. Machine learning has been used to model relations between variables of complex systems, e.g. soil. Using a good machine learning algorithm, we can predict various soil properties like pH, soil electrical conductivity (SEC), etc. Prediction of these variables is important since this can help in deciding which food crops should be planted. SEC refers to the amount of soluble salt present in the soil. If the soil contains too much salt, then the vegetation of the area suffers. Also, the prediction of soil electrical conductivity is important because it is difficult to visit each site to estimate it. Using machine learning models to predict SEC, without any in-situ analysis, we can get an intuitive idea about the value of SEC of the soil. An attempt to predict SEC using neural network model is done in this study. To train the machine learning model, several loss functions were optimized that are generally used for prediction in machine learning—mean squared error loss (MSE or L2 loss), mean absolute error loss (MAE or L1 loss), Huber loss. All the loss functions—mean squared error loss, mean absolute error and Huber loss along with different optimizers like Adam and stochastic gradient descent (SGD)—were experimented in the model. Many techniques like dropout, batch normalization and adaptive learning rate were also attempted in order to improve the model. The error metric for evaluation of our predictions with the actual values was MSE. After a lot of iteration and optimizations, the best model estimated MSE of 0.029.

H. Kotta (✉) · K. Pardasani · M. Pandya · R. Ghosh
Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar 382007, Gujarat, India
e-mail: harsh_kotta@daiict.ac.in; harshkotta@gmail.com

K. Pardasani
e-mail: karanpardasani6311@gmail.com

M. Pandya
e-mail: megha_p@daiict.ac.in

R. Ghosh
e-mail: ranendu_ghosh@daiict.ac.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_9

Keywords Neural network · Predictive soil mapping · Soil electrical conductivity · Dimensionality reduction

1 Introduction

Farmers practising precision agricultural practices need very specific and detailed information about the soil. Soil EC is one of the most efficient and cheap ways for them to get the idea of the productivity of their soil. It is one of the important soil features that is widely used in the field of environmental studies and is an indicator of the soil health. Conventional methods for measuring the soil EC require in-situ collection of soil samples for the analysis in the laboratory. This is not possible to do in large areas. Various machine learning algorithms like K-nearest neighbours (KNNs), support vector machines (SVMs) [1], random forest tree, kriging with external drift (KED) [2] have been used to predict the soil properties. This study mainly aims to predict the soil EC using a deep neural network using covariates like normalized difference vegetation index (NDVI), rainfall, digital elevation model (DEM), etc. The network also contains modules in order to prevent overfitting, i.e. dropout [3], and optimizers in order to converge optimally in the loss curve. We have done analysis of different loss functions [4] in order to find out which one gives the best performance.

2 Basis of the Proposed Approach

2.1 Neural Network

Neural network [5] is the architecture which is analogous to the human brain that is designed to recognize pattern in a data. They can be used for predicting continuous values or in classification. The reason why they are so popular is because they have the ability to model complex relationships between the input and output. Also, they have a good ability to generalize and respond well to unexpected inputs.

It is composed of many layers which are made up of neurons. These neurons are interconnected with each other through nodes, and the output of one neuron is then found by the formula:

$$y_j^l = f(w_{1j}^l x_1^{l-1} + w_{2j}^l x_2^{l-1} + \cdots + w_{nj}^l x_n^{l-1}) \tag{1}$$

where $w_{1j}^l, w_{2j}^l \ldots w_{nj}^l$ are weights associated with the connection of neuron j which is in layer l with neuron of the previous layer and $x_1^{l-1}, x_2^{l-1} \ldots x_n^{l-1}$ are the inputs from the previous layer $(l-1)$ [6].

Each of the connections between the neurons is associated with some weight. These weights are the basis by which the neural network learns. Learning process involves the update of these weights so as to minimize the loss in prediction/classification. The process of this update is called back-propagation. The loss at the final layer of the network is passed back through the network in the form of derivatives with the respective weights in order to update them.

There are hidden layers present between the input and output layers which consist of a group of neurons. These help in fine-tuning the weights through which the networks learn and also capture the salient features of the input which can improve the performance.

2.2 Batch Normalization and Dropout

Deep neural networks with a lot neurons often become prone to overfitting. Overfitting is a process where the network gets limited to a set of data points. It strictly learns the outputs for given inputs and therefore is not able to generalize well. Dropout is a technique to tackle this problem. The way it works is that when there is forward propagation in the neural network along the hidden layers, some of the neurons in them turn off and result in zero contribution from them to the layers ahead. Because of this, during back-propagation, the neurons become less dependent on the weights of other nodes and learn independently. Also since the turning off of neurons is random, every time we get a different network connection. It basically means that we train the data points in a sub-part of the network every time and the final learning is the contribution from all the sub-networks.

Batch normalization involves normalizing the input layer by adjusting and scaling the activations. It reduces the amount by which the values in hidden layers shift and helps in speeding up the learning. It basically helps each layer to learn by itself independent of other layers.

2.3 Adaptive Learning Rate

In order to make the model converge optimally towards minima in the loss curve, we have used adaptive learning rate. The idea of adaptive learning rate is that the learning rate decays after certain number of epochs by a multiplicative factor (which is given as an input). So, as the number of epoch increases, the learning rate decreases, which means the gradient update value decreases. This means that now we take little steps towards the minima of the loss curve. Overall, this leads to better convergence and reduces the chance of crossing the minima.

2.4 Optimizer

An optimizer helps in minimizing/maximizing the objective function $E(x)$ where x is the input data, in order to increase the accuracy of predicting the output from the given input values. The internal model parameters which form the basis of learning of the network are updated using the optimization algorithms. Appropriate optimization algorithms impact the performance of the network and help in its effective training. Some of the popular optimization algorithms are stochastic gradient descent (SGD), Adam, Nesterov accelerated gradient (NAG), etc. We mainly focus on SGD and Adam optimizer in this paper.

2.5 Loss Function

It is one of the most important parameters which helps in giving the feedback to the network regarding its learning. It measures the irregularity in the predicted and actual value. It helps model to train better by controlling the update of its parameters. We have done a comprehensive study on various loss functions and chose the optimal one for our network.

Mean squared error (MSE) loss is calculated by the mean of square of the differences between actual and predicted values across the training examples.

$$MSE = \frac{1}{2n} \sum (y_{\text{actual}} - y_{\text{pred}})^2, \tag{2}$$

where n is the number of training examples, y_{actual} are the actual values corresponding to the inputs and y_{pred} are the predicted output values for these inputs.

Mean absolute error (MAE) is another kind of regression loss function used to update the weights. It is defined as the mean of the summation of absolute difference between the predicted and the actual target values across the input data. It ranges from 0 to infinity.

$$MAE = \frac{\sum |y_{\text{actual}} - y_{\text{pred}}|}{n}, \tag{3}$$

where n is the number of training examples, y_{actual} are the actual values corresponding to the inputs and y_{pred} are the predicted output values for these inputs. There are pros and cons in using these loss functions.

Since MSE involves squaring, the error increases a lot when error >1. If there is an outlier in the data, the error will be very large for MSE (because of the squaring) as compared to MAE. This will reduce the accuracy because the update of the weights depends on the gradient of the error function. In MSE, the gradient is directly proportional to the difference between predicted and target values. In MAE, the gradient is either $+1$ or -1. There will be a large weight update due to an outlier when MSE is used. Therefore, model would focus largely in predicting the outlier correctly and pay less attention to other input data, whereas in case of MAE there will be no large weight update due to an outlier. So, MAE is resistant to outliers. The advantage that MSE has over MAE is that the gradient decreases as we move closer to minima, whereas for MAE the gradient is large even for small values of loss. This does not result in correct convergence to the minima of the loss function when MAE is used.

There is another kind of loss function which is Huber loss. It solves the problem of mean squared error by being resistant to outliers. It is basically the absolute error, which becomes quadratic when error is small. The formula for Huber loss is:

$$L(y, f(x)) = \left\{ \begin{array}{ll} \frac{1}{2}(y - h(x))^2, & \text{for } |y - h(x)| \leq \delta \\ \delta|y - h(x)| - \frac{1}{2}\delta^2, & \text{otherwise} \end{array} \right\}$$

where y is the actual value, $h(x)$ is the predicted value for input x and δ is a hyperparameter.

The error which results in this transition is the hyperparameter(δ). As $\delta \approx 0$, the loss approaches MAE, and when $\delta \approx \infty$, it approaches to MSE. But the choice of delta value is important because it determines what to consider as an outlier. The problem with MAE was that it results in large gradient while training. This can result in missing the minima as we go down the loss curve. The way Huber loss solves this is by acting as MSE when there is low error (closer to minima) and so the gradient decreases as we go towards the minima. This helps in reaching the minima efficiently. But the problem with the Huber loss is that it is tough to find the optimal value of the hyperparameter δ.

3 Experimental Set-Up

3.1 Database Used

The data set used in this study is obtained from the region of North Gujarat. The features chosen are based on the literature study. It has latitude, longitude, NDVI (for all 12 months) obtained from Landsat, rainfall (for all 12 months) obtained from Tropical Rainfall Measuring Mission (TRMM) and DEM obtained from Cartosat as the attributes. These inputs were extracted from satellite images using Google Earth Engine (Table 1). The MODIS data is generated after every 16 days. Hence, there are total 23 columns indicating NDVI values for each month. TRMM-3B43 executes 3B43 algorithm once in a single month to generate best estimation of the precipitation rate. Hence, there are 12 columns indicating measurement of rainfall over a region. Also, three columns corresponding to DEM slope, longitude and latitude are taken as input features. Hence, there are total 38 features for each location. After extracting the data set, it was carefully put together and pre-processed. The prediction attribute

Table 1 Details about the input data

Source	Feature
Google Earth Engine	NDVI (Landsat)
	Rainfall (TRMM)
	DEM (Cartosat)
Ground truth value	SEC (soil healthcard)

is SEC which is an empirical evidence obtained from the ground level. The data set was split randomly in 80–20% train–test ratio to evaluate the model's performance.

3.2 Proposed Model

Our model is made of two hidden layers, one input layer and one output layer. The input feature dimension is 38. The hidden layers comprise 20 and 15 neurons, respectively. Identity function is applied in the output layer which consists of one neuron which indicated the predicted EC value. All the other layers make the use of ReLU activation function in order to bring nonlinearity and improve the predictive performance. Hidden layers are also equipped with dropout to prevent overfitting. They each contain dropout with probability of 0.3 or 0.4. We have tried out different optimizers and loss functions in order to get the best precision in predicting the values. The final predictions are evaluated with mean square error (MSE). Batch normalization was also applied in the model. An adaptive learning rate was used during training to improve the model. Different values of dropout probability and per cent reduction in learning rate after certain epochs were experimented in order to set the hyperparameters properly. Using too low dropout leads to overfitting and large value leads to underfitting. The adaptive learning rate was set such that it decays by a factor of 0.1 after 7 epochs (Fig. 1).

4 Experimental Results

After carefully analysing all the loss functions, our network used MSE and Huber losses for training. Adam and SGD optimizers were also used. Analysis was done for different epochs. The results from all the possible combinations of the loss function, number of epochs, optimizers and other performance improving techniques were obtained and then compared. The best result was obtained for MSE loss with SGD as the optimizer (Table 2). This is probably because of the property of MSE loss and benefit of SGD optimizer which is explained later.

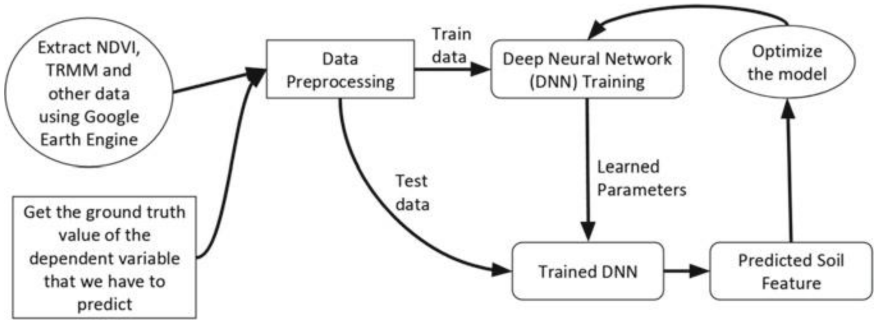


Fig. 1 Flow chart of the methodology

Table 2 Minimum MSE obtained for each of the combinations of loss function and optimizer

Combination	MSE
Adam + MSE	0.06
Adam + Huber	0.063
SGD + MSE	**0.055**
SGD + Huber	0.057

5 Experimental Analysis

5.1 Evaluating the Loss Functions

L1 loss is not preferable as it gives larger gradient values when error is small. This is not good for learning as this will not help the loss function optimally converge to the minima of the curve. Regarding MSE (L2) loss, it can create a problem because it is sensitive to outliers. Even if it does better than L1 loss by giving low gradient (leading to better convergence), its loss value shoots up when the loss is large (because of the squaring term). This will lead to model giving more importance to outliers and hinder the process of learning. But we do not have this problem since we have already removed the outliers from the data. Huber loss could also perform because it captures the useful property of both L1 and L2 losses. For low values of loss, it behaves like L2 loss and gives low gradient updates to the model for good convergence and when the loss value is large, it behaves like L1 loss, being insensitive to outliers that can hamper the learning of the network.

After the experiment, we found out that MSE loss worked out the best. Huber loss did not work out because it is difficult to choose the δ value such that the model converges to minima. Also using the adaptive learning rate (Fig. 2) did not give positive results since it could be making it slow to reach the desired minima in given number of epochs. Even after increasing the epochs, there was not much change in the MSE obtained when using without it. Therefore, for improving computational use, it is better to avoid it.

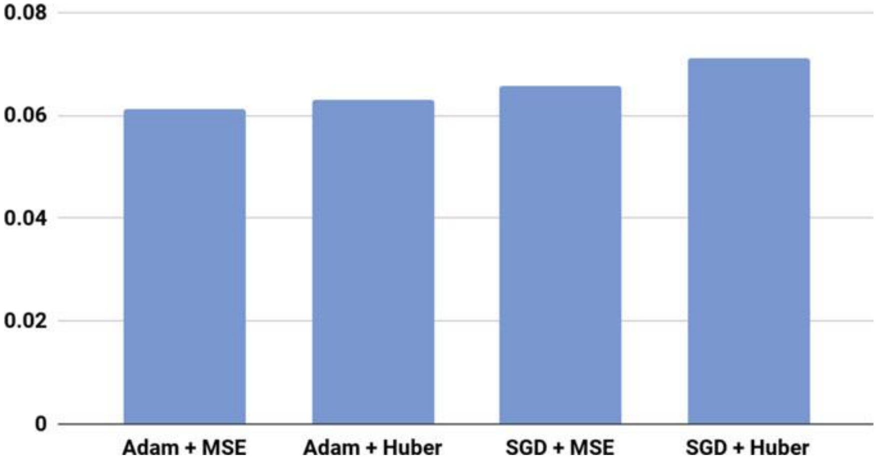


Fig. 2 Minimum MSE with adaptive learning rate

5.2 Why SGD Optimizer?

The analysis of stochastic gradient descent (SGD) optimizer as to why it performed well is as follows:

It can be observed that unlike conventional gradient descent method which takes in all the input values to update the weights of the network, it uses every training example in order to update the parameters of the model in one iteration. This helps in better convergence. Adam optimizer is an addition to SGD that uses different adaptive learning rates for the parameters in order to update them properly. It maintains a learning rate for each of the parameters which updates depending upon the average of the values of the gradients for learning of the model. In the data set, since there are a less number of input parameters, using Adam would make it learn different learning rates for each of the parameters; thereby, it can lead to overfitting. So, Adam leads to better convergence, but SGD helps in better generalization; therefore, using SGD optimizer makes perfect sense. Adam can work better in case we have more number of input parameters. Without dropout, the MSE value obtained was 0.055 using Adam optimizer.

5.3 Analysis of Dropout and Number of Hidden Layers and Neurons

The model worked well when there was no dropout probably because as the number of input parameters is less, using dropout can lead to underfitting. Also since the input data was batch normalized when training, using dropout along with it makes

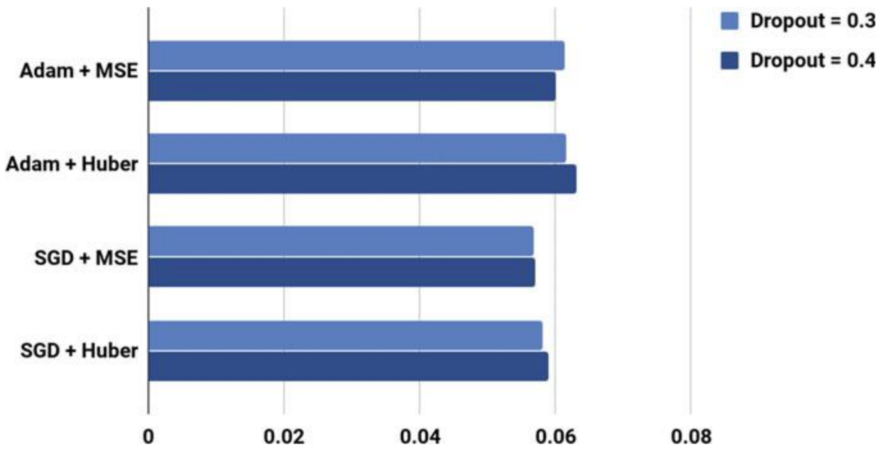


Fig. 3 Minimum MSE with different dropout values

the model fit strictly on training data, failing it to generalize. The MSE value obtained using dropout as 0.3 and 0.4 was 0.0567 and 0.057 (Fig. 3), respectively.

Several combinations of possible number of hidden layers and different numbers of neurons were tried. Care was taken to reduce the overfitting and make the model learn important features. The algorithm was also tested for predicting pH given the same soil features as input. The results indicate that we can also apply the same algorithm for predicting different features having similar distribution as soil electrical conductivity.

5.4 Dimensionality Reduction

Dimensionality reduction is the process of reducing the number of features in the input set. It is done to eliminate the redundancy in the data and convey the information about the data set concisely and effectively. This also leads to computation efficiency, preventing overfitting and therefore leading to a better learning model. We used the technique of permutation test in the best model obtained earlier in the analysis, i.e. SGD + MSE. We have considered three techniques:

Permutation Test: Permutation test was conducted to select the relevant features from the input data. It involves the feature selection after the training of the model. After the model is fitted, a reference testing data set is considered. For each feature, its values are permuted, keeping the values of other features unchanged, and on this data set the trained model is evaluated. The important features give worst performance since when the structure of that feature is changed, the neural network performs worst. In the experiment, 7 features were removed and hence in this case 31 features were used.

After using the reduced number of features in our best model, we obtained a minimum MSE of **0.029** on the test data, which is a decrease of 47.2% from the earlier least MSE value obtained, i.e. 0.055.

6 Summary and Conclusion

Hence, we have proposed a deep neural network model which can be used to predict the soil EC. All the optimization experiments have been carefully performed, and the final results obtained are explained with proper reasoning as to why they should be the one. The model also takes care of preventing overfitting by including batch normalization. We also incorporated the dimensionality reduction techniques in order to achieve 0.029 MSE loss in the test data. Much improvement to the model can be done as a future task, deepening the network, using an algorithm to optimize the neural network such as network in network (NIN) [7]. This can be explored in future.

Acknowledgements We are grateful to Sardarkrushinagar Dantiwada Agricultural University for providing data to carry out the research. We are also highly indebted to Dr. Kankar Shubra Dasgupta, Director of DA-IICT, for his constant support during this study.

References

1. Kovaevic, M., Bajat, B., Gaji, B.: Soil type classication and estimation of soil properties using support vector machines. Geoderma **154**, 340–347 (2010)
2. Vermeulen, D., Van Niekerk, A.: Machine learning performance for predicting soil salinity using different combinations of geomorphometric covariates. Geoderma **299**, 1–12 (2017)
3. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
4. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016)
5. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436 (2015)
6. Yegnanarayana, B.: Artificial Neural Networks. PHI Learning Pvt., Ltd. (2009)
7. Lin, M., Chen, Q., Yan, S.: Network in network (2013). arXiv:1312.4400

Natural Language Information Extraction Through Non-Factoid Question and Answering System (NLIEQA Non-Factoid)

Partha Sarathy Banerjee, Abhijay Ghosh, Aditya Gupta, and Baisakhi Chakraborty

Abstract Over the years, the retrieval of information from unstructured data has increased significantly. The availability of unstructured data, or to be precise, data in the form of natural language is available in abundance. The main role of a Question and Answering System is to process the natural language query and generate a concise answer to it. There are many works done in recent times, which have given a Question and Answering System that helps in answering factoid or list-type queries asked in natural language, but most of them have made use of structured data. The proposed Natural Language Information Extraction through Question and Answering for Non-Factoid cases (NLIEQA Non-Factoid) accepts data and the query fired by the user in the form of natural language text, processes them, and produces the desirable answer. It avoids training the system and the use of Structured Query Language (SQL) for storage and processing. Another advantage of the model is that it can handle complex queries. The model has a strong use of Named Entity Recognition (NER) for classification and extraction of the answers. It also makes use of Stanford's Natural Language Tool Kit (NLTK) for tokenizing, tagging, and chunking of the text.

Keywords NLTK · Information retrieval · NER · Question and answering · Unstructured data · Non-Factoid · Complex queries

P. S. Banerjee · B. Chakraborty
Department of Computer Science and Engineering, National Institute of Technology Durgapur, Durgapur, India
e-mail: partha1010@gmail.com

B. Chakraborty
e-mail: baisakhichak@yahoo.com

A. Ghosh (✉) · A. Gupta
Department of Computer Science and Engineering, Jaypee University of Engineering and Technology, Guna, India
e-mail: abhijayghosh@gmail.com

A. Gupta
e-mail: adityagpt75@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_10

1 Introduction

Question and Answering is an important aspect of Natural Language Processing. Obtaining concise answers to queries asked in natural language is the need of today's world. Especially, if the system can answer complex and compound queries, it will help resolve many problems in various fields. Various fields like that of defense, Central Bureau of Investigation (CBI), biomedical sciences, robotics, chatbots, etc. are where there is a possibility that the people in charge may fire a compound query to fetch Non-Factoid answers. Generating concise answers to the queries being put up by the users is the demand. Question and Answering Systems is very different from mere information retrieval since information retrieval is only fetching related documents based on the query the user has entered, while Q&A systems scan through the documents fetching out only the required answer. Again, Q&A systems are of various types. There can be many constraints put on any Question and Answering System such as being closed domain, using structured data to retrieve information, being dependent on training the machine first. Therefore, dealing with a Q&A system, which implements machine learning, requires a good amount of data in the supervised learning phase. The collection of such enormous data can be a tedious task. If the model is closed domain, its applicability gets restricted to a specific field. The usage of a knowledge base also restricts the variety of queries the Q&A system can handle. In addition to work with structured query language, the user must be skilled enough to be able to learn the query language. The proposed model is far better than the conventional works; since the user is free to input the query in natural language, training the system is not required, the data processing is not dependent on structured data, and it avoids the use of Structured Query Language to fetch the final answers. The Natural Language Information Extraction through Question and Answering (NLIEQA) is capable of processing and producing outputs for complex and compound queries. The system can generate Non-Factoid answers for questions that demand such results. "Where", "Who", "What", "When" can produce Non-Factoid answers. It can deal with various types of date formats also. The system produces only non-descriptive answers in the case of "What" type questions. The system is also capable to generate output for simple "Is" type questions.

2 Utility

The system finds its utility where large chunk of text-heavy information is to be processed and information is to be extracted. The proposed NLIEQA Non-Factoid is specifically designed to handle queries involving Non-Factoid questions and answers. Majority of the work has been done on question and answering involving factoid answers. Areas such as biomedical sciences, where a practitioner might go through several documents to search a certain record related to a patient's details or certain medication, the NLIEQA system can serve the purpose. Even in Defense Agencies, the system has a good scope to retrieve information about the data they want to fetch. It can also find its application major fields like that of AI Chatbots, which are meant to give proper responses to user queries. This system can preferably be used as a package along with any major software/application to help serve the above fields.

3 Advantages

There are several Q&A models that have been proposed prior to the NLIEQA, which have attained commendable level of efficiency. Various tools and techniques have been used in them. Many systems have made use of structured data, knowledge bases, and machine learning techniques to produce results to the queries put up by the users. As a result, many of them are confined to a specific domain. There are systems that are open domain but have made use of a knowledge base to fetch results for user queries. The type of result generated by the system is an important aspect. Most of the systems are generating factoid results and only a few produce a Non-Factoid answer.

The NLIEQA system is an open domain Q&A system that receives both the query and the corpus in natural language as user input. It abstains from the use of Structured Query Language (SQL). The system does not require any prior training. Apart from this, the NLIEQA does not use any storage for storing the input and query, which in turn makes it faster as compared to previous models. This helps to reduce the memory space consumed by the system as well as improves the searching efficiency. It makes use of Named Entity Recognition (NER), which in turn can be improvised if more test cases are dealt in future. It also is capable to deal with "Is" type questions, which can produce decisive and factual results to queries as per the corpus provided. The system first produces a list of approximate answers and then among those statements it produces the final answer. This reduces the number of searches and as a result reduces the searching time.

Furthermore, the system can handle compound and complex queries and produce Non-Factoid results. Most of the previous works could not produce results when the question asked about the date. This system can deal with various date formats and generate result when needed.

4 Related Works

Proposing a new Question and Answering system, the authors of work [1] stated that building a Q&A system can be challenging when the data provided is open domain and unstructured, since the query asked by the user may not contain exactly the same set of words in the text. The work [2] has given an open domain Q&A system which generates Non-Factoid answers to user queries. It has used the Web to extract the required information and then uses Named Entity Recognition (NER) to extract the factoid answer. Another type of Q&A system has been illustrated in work [3], which uses a pre-structured database to answer questions asked in natural language related to the field of education. The authors in work [4] have described the various methods to enhance the performance of Named Entity Recognition (NER) with the help of various techniques like use of machine learning and some rule-based approaches. Work [5] implements Hidden Markov Model to design a Q&A system which can accept queries both using text and speech. The system has a success rate of 83.31%. The use of neural network and knowledge base to generate Non-Factoid answers is seen in Q&A system given in work [6]. Keeping in mind the enormous amount of data available on the net today and the ease of user to access this information, work [7] proposes a Q&A system that makes use of structured data and a knowledge base. It combines the works from fields of information retrieval and Natural Language Processing. A Web-based Q&A system has been used in work [8] to depict the importance and usage of various types of distributional semantic models for ranking the Q&A systems' answers. It was inferred that the random indexing model is appropriate model for being used in the real-time application of the Q&A system. As illustrated in work [9], Q&A systems can make use of a machine learning approach to improve the results generated to the queries put up by the user in natural language. This work uses a knowledge-based approach to evaluate the accuracy of the answers obtained for "Why" type questions. Another Q&A system based on machine learning and using a database is been presented in work [10], which is designed to help the people in the field of health care. Since a large volume of data is generated in this field, the work proposes to handle complex queries and generate Non-Factoid answers with the help of its database. The authors of work [11] have studied 130 out of 1842 papers that deal with various approaches to a Q&A system in order to brief about the various commonly used techniques, efficiency of the system using that technique, etc. They found many of the works to be open domain where researchers have focused on knowledge base and information retrieval paradigms. Various technical aspects like accuracy, domain of application, and language of application were laid emphasis to draw proper inferences. Work [12] illustrates the implementation of a Q&A system, which helps processing queries primarily related to the field of biomedical science. It can generate both factoid and list-type answers for the queries fired by the user. It deals with multi-label question classification by improving the question classification dataset to enhance the performance of list-type questions. The Q&A model proposed in work [13] generates factoid answers to queries with the help of Named Entity Recognition (NER) and various functions of Stanford's Natural Language Toolkit (NLTK). The advantage of this work is that it avoids the use of any system training and use of any kind of database. It uses a set of rules to extract information from the text. It is also the comparative work for the proposed model. Exploring the hierarchy organized type system or the type taxonomy of the entities, work [14] is able to achieve a good level of efficiency in retrieving ad hoc entities. It makes use of structured data and performs analysis based on the type-based and term-based similarity in the retrieval system apart from the above parameter. Another excellent idea of a Q&A system has been presented in work [15], which is meant to serve the biomedical field. It exploits neural networks using machine learning-based answer approach. It has shown that the use of word embedding along with features obtained from input textual sources has improved the performance of the system. Emphasizing the importance of more precise answer retrieval instead of returning a list of relevant documents, the authors of work [16] examined different fine-grained answer presentation methods for Non-Factoid results and proposed an interactive Q&A setting using the Amazon Mechanical Turk.

5 System Architecture

The following figure (please refer to Fig. 1) illustrates the workflow of the proposed NLIEQA model
The model consists of two important functions:

(i) Approximate Answer Generator and
(ii) Query-Type Classifier

First, the user provides the input to the system, i.e., the corpus and the query in natural language. The query is sent to the query_analyzer function to dismantle it into tokenized form to further tag the words present in it with the respective parts of speech, separate the important entities like PERSON, GEO-POLITICAL ENTITY, and ORGANIZATIONS. On the other hand, the text is tokenized in the form of sentences.

Then, these sentences along with the various verbs and entities obtained from the query_analyzer are sent to Approximate_Answer_Generator function. The entities are searched in a taxonomical manner in order to increase the relevancy of the approximate answers obtained. The approximate answers obtained are then printed.

The approximate answers obtained in the above step are now split on the basis of the POS lists obtained from the query_analyzer. Also, the approximate answers obtained are sent to the statement_analyzer for the analysis in the approximate answers the way the query was analyzed.

The intermediate results obtained until now, i.e., the POS lists from query_analyzer, statement_analyzer and split form of approximate answers are sent to the Query-Type Classifier function to classify the type of “Wh” or “Is” question. Each type of question is dealt with a different set of rules, and if the sufficient data is been provided in the query and in the text, the Query-Type Classifier function will yield desired factoid or Non-Factoid answers.

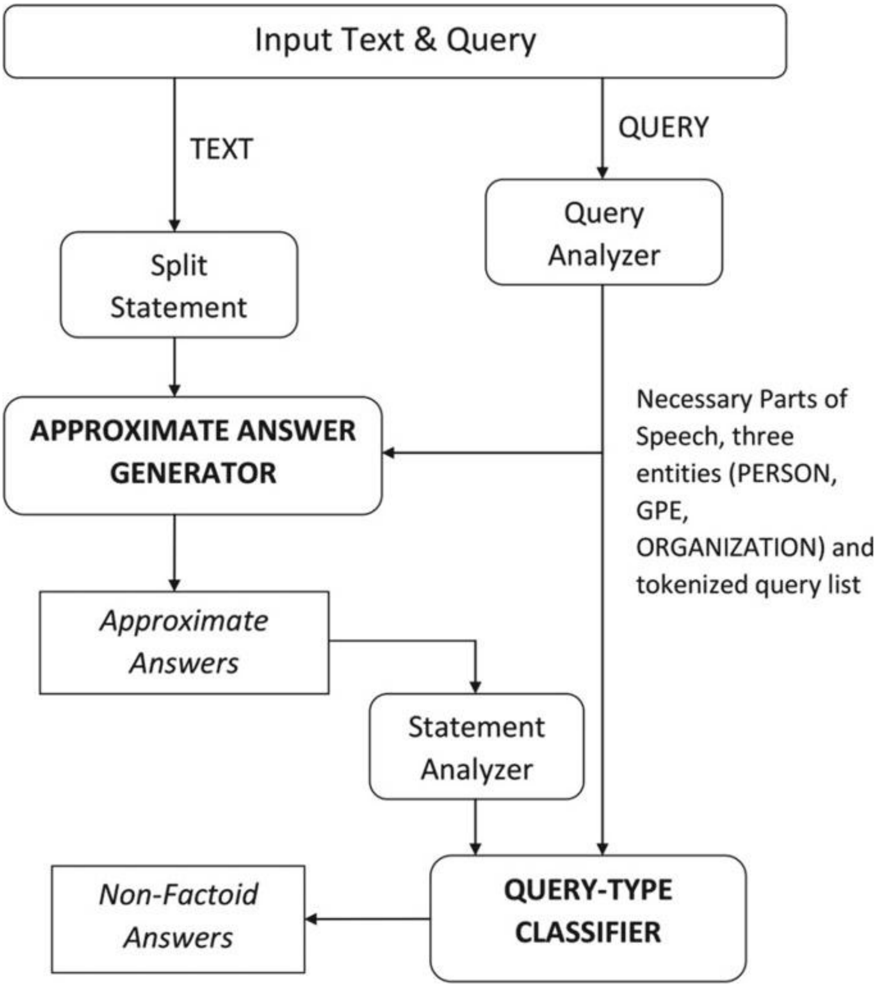


Fig. 1 System architecture of the proposed model

6 Prerequisites About the Algorithm

Since the various "Wh" type and "Is" type questions are solved using a combination of certain number of entities, which are decided based on the possible in which ways different types of questions can be asked. Table 1 shows the list of the entities that are being used for searching the Non-Factoid answers.

If a "Who" type question is detected, then 18 possible combinations of some of the above entities are made to fetch the Non-Factoid answer. Similarly, for "Where" and "When" type questions, we prepared 24 and 25 combinations, respectively.

For instance, for "**Who**" type questions, some of the combinations are

Table 1 Entities and POS tags used

q_gpe	q_person	q_org
q_cardinal	q_verb	q_prepo
q_adjective	q_noun_plural	q_noun

O1 = [q_verb, q_noun, q_cardinal, q_org/q_gpe, q_prepo, q_person]
O2 = [q_verb, q_noun, q_cardinal, q_org/q_gpe]
O3 = [q_verb, q_noun, q_cardinal, q_prepo, q_person]
O4 = [q_verb, q_prepo, q_person]
For “**Where**” type questions, some of the combinations are
E1 = [q_verb, q_person, q_gpe/q_org, q_noun/q_noun_plural, q_prepo]
E2 = [q_gpe, q_org]
E3 = [q_gpe]
E4 = [q_adjective, q_person, q_noun/q_noun_plural, q_gpe/q_org]
For “**When**” type questions, some of the combinations are
N1 = [q_verb, q_person, q_gpe/q_org, q_noun/q_noun_plural]
N2 = [q_verb]
N3 = [q_adjective, q_person, q_noun/q_noun_plural, q_gpe/q_org]
N4 = [q_adjective, q_person, q_noun/q_noun_plural, q_gpe/q_org]
For “**What**” type questions, some of the combinations are
T1 = [q_verb, q_person, q_gpe/q_org, q_noun/q_noun_plural]
T2 = [q_verb]
T3 = [q_cardinal, q_adjective, q_person, q_gpe/q_org, q_noun/q_noun_plural]
T4 = [q_cardinal, q_org, q_prepo]
For “**Is**” type questions, some of the combinations are
S1 = [q_person, q_verb, q_noun, q_adjective]
S2 = [q_person, q_verb, q_noun]
S3 = [q_person, q_noun]
S4 = [q_noun, q_verb]

The combinations are selected by putting the length function around each entity in the combination. It is then checked all the entities one by one (in the particular question type) that if any of the combinations has all the entities with length “1”, then Non-Factoid answer is searched according to that entity only. If none of them satisfies, no result is produced.

7 Algorithm

The following is the description of working of the system.

1. The user provides the corpus and query which is stored in variables X and Y, respectively.
2. Query_Analyser

   a. A word_tok_list stores the tokenized form of the query.
   b. word_tok_list is POS tagged and chunked.
   c. Various entities that emerged out are stored in separate lists.
   d. It returns all the lists of the entities.

3. The text is split into statements
4. Approximate_Answer_Generator

   a. The split statements and the verb (various forms), PERSON, GPE, and ORGANIZATION entities obtained from Step-2, act as the parameters for this function.
   b. The verb and (PER, GPE, ORG) entities are put into different lists.
   c. The split statements are searched in, for the presence of these entities. The relevant statements found are stored in an app_ans list.
   d. The app_ans list contains all the approximate answers, so it is printed.

5. Statement_Analyser

   a. The approximate answers' list is the parameter for this function.
   b. The statements in the list are split on the basis of the stop word "and".
   c. Now, all the sentences are word tokenized, POS tagged, and chunked.
   d. It returns all the lists of the entities.

6. Query-Type Classifier

   a. The approximate answers' list, lists of entities obtained from both Step-2 and Step-5, are passed as the parameters for this function.
   b. Now, the category of the question is identified, whether it is "Who", "What", "When", "Where", or "Is".
   c. According to the category of the question, the flow of the program jumps onto that case.
   d. As stated earlier, under each type of question, each combination's entities' lengths are checked. If any of the combinations has all its entities of length "1", then Non-Factoid answer is obtained.

The Non-Factoid answer is then printed.

8 Results

Figure 2 (given below) represents the data flow diagram of the NLIEQA system, where the order of invocation of the methods is specified along with the name of the method.

Let the input query be "when will Tommy arrive and Harry begin his speech" (Please refer to Fig. 3). As per the system flow, the query is broken down into different parts of speech and the entities: PERSON, GPE, and ORGANIZATION are also separated out.

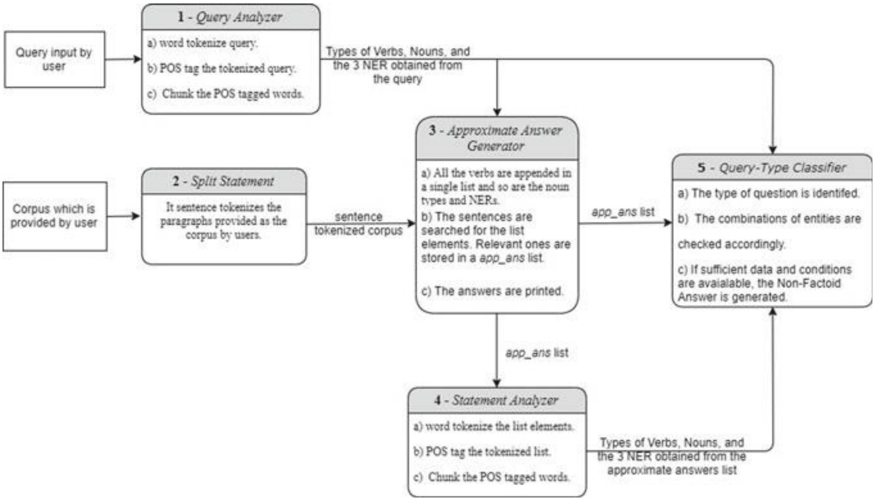


Fig. 2 A data flow diagram depicting the process flow of the system

```
question is:
 when will Tommy arrive and Harry begin his speech

The person in the question is:  ['Tommy', 'Harry']

The person in approximate answere  is:  ['Tommy', 'David', 'Texas', 'Harry', 'Tom']

The geopoliticial Entity and Organisation listed in question are:  []

The geopoliticial Entity and Organisation listed in approximate answere are:  []

Various objects in the question are:  ['speech']

Various objects in approximate answere are:  ['minutes', 'tomorrow', 'match', 'performance', 'noodles', 'home', 'car'
, 'speech', 'guests', 'tonight']

Approximate answere is:  ['The guests will arrive tomorrow', 'Harry will begin his speech tonight', 'David was eating
noodles while Harry drove the car', 'Tommy will then return home on 15', 'Tom  will begin his performance in 3 minute
s', 'Harry will arrive home on 26-09-2019', 'Tommy will arrive Texas at 1am', 'The match will begin at 5pm', 'Tommy w
ill start playing from 15-Aug-2020', 'Mary her begin their performance in 4 minutes']

Various actions performed in approximate answere are ['drove', 'eating', 'return', 'playing', 'begin', 'arrive', 'sta
rt']

1am
['2019-09-19 01:00:00']
tonight
the above is the non-factoid answere of the query provided
>>> |
```

Fig. 3 An illustration for the query "when will Tommy arrive and Harry begin his speech"

Hence, the query contains only PERSON = ['Tommy', 'Harry'] and Object = ['speech'].

The verbs obtained from the query go into the approximate answer generator along with the corpus (split form) and the entities. The entities and the verbs are searched in the sentences in a particular order to improve the results obtained from this function. After the approximate answers are found, they are printed, for instance

["The guests will arrive tomorrow", "Harry will begin his speech tonight"… "Tommy will return home on 15" … "Tommy will arrive in Texas at 1am"….]

This approximate answer now is sent to the statement analysis function for getting broken down into different parts of speech and the entities are also extracted like that of the query.

Inside the *Query-Type Classifier* function, the type of question is identified first. In this case, the type is "when". It is checked initially if it is a complex query, whether the gerund available in the query is common for the whole question or there are multiple gerunds. The query is then split accordingly. The search for "when" type begins with searching of the main entities, if either is missing, checking for the parts of speech according to certain rules. If relevancy is found the Non-Factoid, result is generated. For this case, our results generated are

1am
[YYYY – MM - DD 01:00:00] } Answer to "**when will Tommy arrive**"

tonight } Answer to "**when will Harry begin his speech**"

[Note: If either the date or time is not mentioned, the system prints the current date and time.]

Taking up another case which deals with "Who" type question (Please refer to Fig. 4).

Input = "who is playing cricket in America and eating mango at Paris"

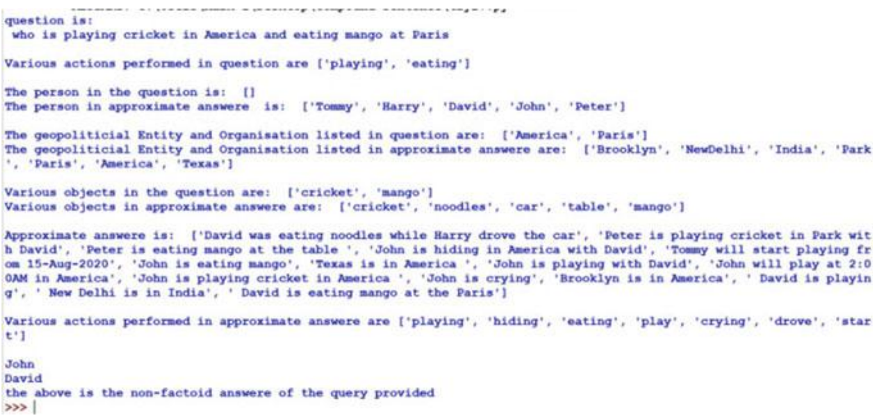
Here, the various entities inferred from the query are

Verbs = ['playing', 'eating']

GPE = ['America', 'Paris']

Noun = ['cricket', 'mango']

On the basis of these parts of speech and entities, the approximate answers obtained are:


```
question is:
 who is playing cricket in America and eating mango at Paris

Various actions performed in question are ['playing', 'eating']

The person in the question is:  []
The person in approximate answere  is:  ['Tommy', 'Harry', 'David', 'John', 'Peter']

The geopoliticial Entity and Organisation listed in question are:  ['America', 'Paris']
The geopoliticial Entity and Organisation listed in approximate answere are:  ['Brooklyn', 'NewDelhi', 'India', 'Park
', 'Paris', 'America', 'Texas']

Various objects in the question are:  ['cricket', 'mango']
Various objects in approximate answere are:  ['cricket', 'noodles', 'car', 'table', 'mango']

Approximate answere is:  ['David was eating noodles while Harry drove the car', 'Peter is playing cricket in Park wit
h David', 'Peter is eating mango at the table ', 'John is hiding in America with David', 'Tommy will start playing fr
om 15-Aug-2020', 'John is eating mango', 'Texas is in America ', 'John is playing with David', 'John will play at 2:0
0AM in America', 'John is playing cricket in America ', 'John is crying', 'Brooklyn is in America', ' David is playin
g', ' New Delhi is in India', ' David is eating mango at the Paris']

Various actions performed in approximate answere are ['playing', 'hiding', 'eating', 'play', 'crying', 'drove', 'star
t']

John
David
the above is the non-factoid answere of the query provided
>>> |
```

Fig. 4 An illustration for the query "who is playing cricket in America and eating mango at Paris"

[“David was eating noodles while Harry drove the car”, “Peter is playing cricket in Park with David”, “Peter is eating mango at the table”, “John is hiding in America with David”, “Tommy will start playing from 15-Aug-2020”, “John is eating mango”….. “John is playing cricket in America”…. “New Delhi is in India” …. “David is eating mango at Paris”]

Now, after the approximate answer is processed along with the POS and entity tags from query_analyzer and statement_analyzer, the result generated for this is

John Answer to “**who is playing cricket in America**”

David Answer to “**who is eating mango in Paris**”

9 Limitations

Since the NLIEQA works with the help of Stanford’s Natural Language Tool Kit (NLTK), it unable to tag and chunk regional names and words. NLTK also demands the presence of a noun just after any “ing” verb to identify that verb as a gerund in some cases while working with the query.

For example,

“Who is playing”	(Here the POS tagger tags “playing” as a VBG)
“Who is eating and playing”	(Here framing of the question has to be like “Who is eating mango and playing football”)

The system cannot deal with cases like, if the corpus contains a sentence “John was **having** sandwich in his breakfast today with David.” and a query is fired “what was David **eating** in his breakfast”.

If the statement/query has a negation in it or if co-referencing is used, the system is unable to produce the expected answer. For example, if the texts are “John was not playing cricket in the park with David,” “David is the tallest boy in this class. He is 5 foot tall.” and queries corresponding to the texts are “Who is playing with David in the park” and “who is 5 foot tall in the class” respectively, the system will print “John” for the first query and no result for the second query.

10 Conclusion and Future Work

The proposed NLIEQA system uses 3-Class NER to generate Non-Factoid answers. It allows the user to retrieve answers to simple as well as compound and complex queries from the corpus by user himself. The system can handle “Who”, “When”, “Where”, “What” and “Is” category questions. For “When” type questions, the system can produce answers which may contain date, time in any format apart from the regular answers like "today", "tomorrow", etc. In the "Is" type of question, if the question asked is decisive, the system will produce an output YES or NO, else if the question is factual, it will generate the answer as per given in the corpus. The system does not perform so well if the category of "Wh" question is "What" and "When". The reason is that these categories can demand for narrative type answers and our system does not yield such answers.

In future research, the concept of natural language generation would be adopted for achieving better results like producing narrative answers. In addition, we expect to work on designing our own package to recognize many regional names that would further improve its usability.

Acknowledgements This publication is an outcome of the R&D work undertaken project under the Visvesvaraya Ph.D. Scheme of Ministry of Electronics and Information Technology, MeitY, Government of India, being implemented by Digital India Corporation. This research work has been done at Research Project Lab of National Institute of Technology (NIT), Durgapur, India. Financial support was received from Visvesvaraya Ph.D. Scheme, Deity, Government of India (Order Number: PHD-MLA/4 (29)/2014_2015 Dated-27/4/2015) to carry out this research work. The authors would like to thank the Department of Computer Science and Engineering, NIT, Durgapur, for academically supporting this research work. The authors would also like to thank the Department of Computer Science and Engineering, Jaypee University of Engineering and Technology, Guna MP.

References

1. Angeli, G., Nayak, N., Manning, C.: Combining natural logic and shallow reasoning for question answering, pp. 442–452 (2016). https://doi.org/10.18653/v1/P16-1042
2. Ranjan, P., Rakesh, C.: Question answering system for factoid based question, pp. 221–224 (2016). https://doi.org/10.1109/IC3I.2016.7917964
3. Lende, S.P., Raghuwanshi, M.M.: Question answering system on education acts using NLP techniques. In: 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave), Coimbatore, pp. 1–6 (2016). https://doi.org/10.1109/STARTUP.2016.7583963
4. Wongso, R., Suhartono, D.: A literature review of question answering system using named entity recognition. In: 2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, pp. 274–277 (2016). https://doi.org/10.1109/ICITACEE.2016.7892454
5. Ho, H., Mawardi, V., Dharmawan, A.: Question answering system with hidden Markov model speech recognition. In: 2017 3rd International Conference on Science in Information Technology (ICSITech), Bandung, pp. 257–262 (2017). https://doi.org/10.1109/ICSITech.2017.8257121
6. Ma, R., Zhang, J., Li, M., Chen, L., Gao, J.: Hybrid answer selection model for non-factoid question answering. In: 2017 International Conference on Asian Language Processing (IALP), Singapore, pp. 371–373 (2017). https://doi.org/10.1109/IALP.2017.8300620
7. Chandurkar, A., Bansal, A.: Information retrieval from a structured knowledgebase. In: 2017 IEEE 11th International Conference on Semantic Computing (ICSC), San Diego, CA, pp. 407–412 (2017). https://doi.org/10.1109/ICSC.2017.95
8. Mahajan, R.S., Zaver, M.A.: Novel answer ranking approach in question answering system using compositional distributional semantic model. In: 2018 3rd International Conference for

Convergence in Technology (I2CT), Pune, pp. 1–5 (2018). https://doi.org/10.1109/I2CT.2018.8529741
9. Pundge, A., Mahender, C.: Evaluating reasoning in factoid based question answering system by using machine learning approach. In: 2018 3rd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2018, pp. 821–825 (2018). https://doi.org/10.1109/CESYS.2018.8724085
10. Yeo, H.: A machine learning based natural language question and answering system for healthcare data search using complex queries. In: 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 2467–2474 (2018). https://doi.org/10.1109/BigData.2018.8622448
11. Calijorne Soares, M.A., Parreiras, F.S.: A literature review on question answering techniques, paradigms and systems. J. King Saud Univ. Comput. Inf. Sci. (2018). https://doi.org/10.1016/j.jksuci.2018.08.005
12. Wasim, M., Mahmood, W., Asim, M.N., Khan, M.U.: Multi-label question classification for factoid and list type questions in biomedical question answering. IEEE Access **7**, 3882–3896 (2019). https://doi.org/10.1109/access.2018.2887165
13. Banerjee, P.S., Chakraborty, B., Tripathi, D.: A information retrieval based on question and answering and NER for unstructured information without using SQL. Wirel. Pers. Commun. **108**, 1909 (2019). https://doi.org/10.1007/s11277-019-06501-z
14. Garigliotti, D., Hasibi, F., Balog, K.: Identifying and exploiting target entity type information for ad hoc entity retrieval. Inf. Retr. J. **22**, 285 (2019). https://doi.org/10.1007/s10791-018-9346-x
15. Dimitriadis, D., Tsoumakas, G.: Word embeddings and external resources for answer processing in biomedical factoid question answering. J. Biomed. Inform. (2019). https://doi.org/10.1016/j.jbi.2019.103118
16. Qu, C., Yang, L., Croft, W.B., Scholer, F., Zhang, Y.: Answer interaction in non-factoid question answering systems. In: Proceedings of the 2019 Conference on Human Information Interaction and Retrieval—CHIIR '19 (2019). https://doi.org/10.1145/3295750.3298946

An Enhanced Differential Evolution Algorithm with New Environmental-Based Parameters for Solving Optimization Problems

Avjeet Singh, Alok Kumar, and Anoj Kumar

Abstract For solving the complex and nonlinear problems, there are numerous nature-inspired algorithms exist, and differential evolution (DE) is one of the best algorithms. DE is utilized in numerous fields of science and engineering. It suffers some drawbacks like, this algorithm suffers from premature convergence rate and furthermore the stagnation problem. This paper presents a novel approach New Environmental-based Parameter for Differential Evolution (NEPDE). In this article, two new variants of DE (NEPDE-B and NEPDE-CR) have been proposed for solving the stagnation problem and maintaining diversity. This approach improved the performance of the standard differential evolution, and convergence speed is also improved. The candidate's solutions do not converge to a specific point to the search space. We used CEC2015 benchmark functions for measuring the performance of the proposed algorithm, and the results indicate that the proposed approach is able to find effective solutions after comparing the different state-of-the-art algorithms of DE.

Keywords Environmental mutation · Evolutionary algorithms · Differential evolution

1 Introduction

For a given problem, optimization is the technique of picking the best solution under certain circumstances. Differential evolution is a simple, heuristic and population-based optimization technique, initially proposed by the Storn R. et al. [1] in 1996. It is a stochastic and search-based algorithm, which depends on the population size. It

A. Singh (✉) · A. Kumar · A. Kumar
MNNIT Allahabad, Prayagraj, India
e-mail: 2016rcs01@mnnit.ac.in

A. Kumar
e-mail: alokkumar@mnnit.ac.in

A. Kumar
e-mail: anojk@mnnit.ac.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_11

is used to resolve persistent optimization problems that may be constraint or unconstrained. It is a technique for solving the complex problem through iterative manner. DE works only with three operators like mutation, crossover and selection. These operators help to control the algorithm from different and complex problems. These operators only contribute to improving the efficiency of this algorithm; for various engineering problems and other application, it is easy to implement. The rest of the paper is organized as pursues: The fundamental of DE is presented in Sect. 2, Sect. 3 discussed the literature review of different variants of DE. The proposed approach NEPDE is introduced in Sect. 4. The results and the experimental setup with comparative analysis of state of the art with the proposed method are discussed in Sect. 5. At last, Sect. 6 concludes this proposed approach (NEPDE) and also discussed the future direction.

2 Differential Evolution

DE is a population-based technique that is used for solving a wide range of optimization issues. DE encodes the randomized solutions for the ith individual at the generation G, with the help of the following equation.

$$\overrightarrow{X}_i = \left\{\overrightarrow{X}_{1,G}, \overrightarrow{X}_{2,G} \ldots \overrightarrow{X}_{j,G} \ldots \overrightarrow{X}_{D,G}\right\} \tag{1}$$

where $i = 1, 2,\ldots j\ldots, D$, for each vector D represents the total number of dimensions in solution space and G represents the total number of generations for search space. So, the population contains D, parameter vectors. Algorithm 1 and Algorithm 2 describe the DE/best/1 and DE/current to best/1 with NEPDE variants. Mainly, DE works with four steps for searching a desired solution like initialization, mutation, crossover and selection.

Mutation: After generating the initial population and doing the fitness evolution from the initial population, to perform the mutation for each vector, we usually select three solutions from the population randomly, so the three vectors are chosen randomly.

$$\text{DE/rand/1} \quad {}_i^k\overrightarrow{V}_G = \overrightarrow{X}_{R_1^i,G} + F_1^i\left(\overrightarrow{X}_{R_2^iG} - \overrightarrow{X}_{R_3^iG}\right) \tag{2}$$

where i represented the index of the population member and index i run from 1 to population size. R_1, R_2 and R_3 are three different solution members taken from the population, so R_1, R_2 and R_3 *are* different from ith member. These members randomly selected from the remaining population. First one is the base vector, and two others are the separated randomly chosen to form the difference. F is the scale factor; initially, it was constant, and so many authors used it like adaption or them. DE has an advantage compared to, for example, evolution strategy. DE/rand/1 mutation strategy is introduced by the Storn et al. in 1996. The advantage of the difference vector is that suppose having a search space and randomly initializing the member all over the search space so, in the initial stage, the population may be well separated. When the population members are well separated, the difference vector will have to resemble large magnitude because the population is separated. Evolutionary algorithms evolve over a generation, the population member usually tends to move better solutions towards the good region and the solution members towards the good region, and they become cross each other. After crossing each other, the magnitude of the difference vector will naturally become smaller. DE is usually altering the search behaviour from exploration to exploitation as the population evolves.

Crossover: The higher crossover value will generate a better solution more frequently. After successful completion of the mutation phase, a trail vector is generated for each pair of the target vector. Crossover operation is done between the *i*th population member and the mutation vector.

Selection: After completing the crossover operation, apply better selection scheme. We calculate and select the offspring if the objective value of the offspring is lower or similar.

$$ {}_{i}^{k+1}\overrightarrow{X}_{i} = \begin{cases} \overrightarrow{Y}_{i,G+1}, & \text{if } f\left(\overrightarrow{Y}_{i,G+1}\right) \leq \text{target function} \\ {}_{i}^{k}\overrightarrow{X}_{i} & \text{if } f\left(\overrightarrow{Y}_{i,G+1}\right) > \text{target function} \end{cases} $$

Here, $f\ (\overrightarrow{X}_{i,G})$ is the target function if the objective value of offspring is higher than the target function, keep the parents. The target value is set $\overrightarrow{X}_{i,G}$ in place of the trail vector if the $\overrightarrow{Y}_{i,G+1}$ is a superior cost function value from the target function otherwise remains the old value of $\overrightarrow{Y}_{i,G+1}$.

Singh et al. [2] proposed dynamically selection of the parameters called DSPDE. The algorithm proposed by using DE's parameters dynamically has helped in increasing the convergence speed and its performance. DSPDE was compared with the different variants of DE and state-of-the-art algorithms and tested in 24 benchmark functions of CEC2015. The results show that DSPDE comparatively gets the best results with state-of-the-art algorithms.

3 Related Work

For solving a nonlinear optimization problem, a metaheuristic, stochastic method approach is introduced by the Storn R. et al. which is called differential evolution [1]. It performs very well with non-convex, multi-objective and dynamic components. Whenever initializing the population (initialize members randomly), there may be a possibility for the solution which is very good. But if this good member solution allows in the initial population, it influences all others. It has the potential to attract all others near it. But the initial solution is far away from the global solution because they randomly initialize all of them.

Mohamed, A. W. [3] proposed a novel approach called NDE; in this approach, a triangular mutation is introduced with DE for solving the constraint optimization problems. NDE is defined by in terms of best, better and worst vectors which are based on the convex combination vector. The proposed approach enhances the convergence rate of the previous algorithm. It is searching the better solutions between global and the local exploitation. The introduced method is tested on 24 benchmark functions with constrained optimization and other mechanical design problems. NDE demonstrates a superior after comparing with the other recent DE variants.

For improving the performance of DE, Mohamed et al. [4] proposed new mutation strategies like ord_best and ord_pbest. This approach introduced four work contributions for enhancing the exploration capability:

i. Purposed a less greedy mutation approaches called DE/current-to-ord_best/1 and DE/current-to-ord_pbest/1
ii. Purposed the new approaches with novel mutation for enhancing the performance of the DE called EDE, EBDE, ESHADE, EBSHADE, ELSHADE and EBLSHADE.
iii. Author purposed the hybrid framework (ESHADE and ELSHADE frameworks).

This approach is tested on the CEC2013 and CEC2017 benchmark functions and shows that the purposed method finds the best experimental results with increasing the dimensions as compared with the other DE variants. For large-scale global numerical optimization, the performance was also evaluated on the CEC2010 benchmark function. For solving global numerical optimization problems, Cui et al. [5] purposed novel approaches for adaptive differential evolution algorithm called MPADE. Purposed three novel DE approaches tested for checking with exploitation and exploration. MPADE was tested on 55 benchmark functions and 15 real-world problems. The experimental results incorporate that the purposed approach is better than the other state-of-the-art algorithms. A. K. Qin et al. [6] proposed a strategy of adaption-based differential evolution called SaDE. Here, a trial vector generation assign to the control parameters for every target vectors in an initial population.

4 The Proposed Work

Differential evolution is entirely depended on its control parameters to improving the enforcement. These control parameters need to vary at a time. The main problems have to face to change the control parameters when we moved to a higher dimension or increase the different dimensions. The control parameters decided the performance of DE algorithm at any time. In this section, a novel DE called new environmental-based parameter for differential evolution algorithm had been introduced.

4.1 The NEPDE Algorithm

Initially, DE/rand/1 mutation strategy is introduced by the Storn R. et al. in 1996 [1]. Three vectors are chosen randomly from the population, and one of the base vectors which is selected randomly is added with the difference of others two vectors. To achieve the best result and to improve the performance, tuning with the parameters is very important. Therefore, it can keep up population diversity and the search ability in locally and globally with no inclination to a specific domain. The functionality of the metaheuristic-based DEA depends on the mutation, crossover and population. For maintaining the convergence speed and solving the stagnation problems, there are many other greedy strategies proposed like DE/rand/2, DE/best/1, DE/current to best/1 and DE/best/2 which incorporate the information of the optimal solutions. DE/rand/2 is the same as a DE/rand/1 but with the addition of two more vectors and the difference between these two vectors are added with them. Numerous techniques have been proposed to now to take care of this stagnation problem. For improving the performance of the existing DE algorithm, Zhang, J. et al. [7] proposed a new mutation strategy in 2009 called DE/current to p_best. Control parameters are updated adaptively, and DE/current-to-p_best with external archive in this approach. For choosing the mutant vector randomly, the author makes the union of set of achieving interior solutions and the current population size. To avoiding the computation overhead, author used to perform operation because it helps to improve the diversity of the population. JADE tested with the Dixon–Szego function in 30-D and 100-D, and it's comparing with others state-of -the-art algorithms like JDE with achieving, without achieving, SADE and PSO.

Subsequently, this article utilizes another environmental-based mutation rule with a perspective on adjusting global exploration capacity and local exploitation inclination. NEPDE maintains the diversity during initial generation, which is achieved with an environmental-based mutation from new environmental-based vectors NEPvs. These vectors are created to maintain the diversity and to improve the convergence rate when the solution is faced the stagnation problem in local search space. This approach has used in mutation operator of DE, new vectors are:

$$\text{vector}_1 = \sum_{i=1}^{\text{NP}} \overrightarrow{X}_{R_1^i G} * \text{NEPv1}, \ \text{vector}_2 = \sum_{i=1}^{\text{NP}} \overrightarrow{X}_{R_1^i G} * \text{NEPv2}$$

where NEPv1 and NEPv2 are the new environmental-based vector vectors depending concerning the nature of the best arrangement regarding best look through space. The new environmental-based mutation is created as:

New donor vector is:-

$$ {}_i^k\overrightarrow{V}_G = \overrightarrow{X}_{\text{best}} + F_1^1 \left(\sum_{i=1}^{\text{NP}} \overrightarrow{X}_{R_1^i G} * \text{NEPV1} - \sum_{i=1}^{\text{NP}} \overrightarrow{X}_{R_2^i G} * \text{NEPV2} \right) \tag{3}$$

where $\overrightarrow{X}_{\text{best}}$ represent the best vector and $\overrightarrow{X}_{R_1^i G}$, $\overrightarrow{X}_{R_2^i G}$ will be generated from the possible solution space. Algorithm 1 and algorithm 2 will discuss two different variants of DE called NEPDE-B and NEPDE-CR, which is based on DE/best/1 and DE/current to best/1 variants. Proposed approach that follows the concept of NEPDE-B (/best/1), this procedure explains in Algorithm 1.

Algorithm 1 NEPDE-B (/ best /1)

Parameter definition
F *mutation (scale factor)*
Cr *crossover*
G *number of generation*
1: Start NEPDE-B
2: Population initialization.
4: F ∈ [0.1 to 2], Cr ∈ [0.1 to 1]
5: Evaluate the fitness value.
6: Apply mutation technique:
$^{k}_{i}\vec{V}_{G} = \vec{X}_{best} + F_1^{\ 1} (\sum_{i=1}^{NP} \vec{X}_{R_1^i\ G} * NEPV1 - \sum_{i=1}^{NP} \vec{X}_{R_2^i\ G} * NEPV2)$
Else $^{k}_{i}\vec{V}_{G} = X_i^G$ **end if**
7. If updated value $\leq \overrightarrow{\mathbf{X}}_{\mathbf{1,G}}$, keep the previous value
Else updated new value $X_i^{G+1} = \vec{X}_{\mathbf{1},G}$ **end if**
8. Choose best value and updated, $X_i^{G+1} + 1$
If conditions met, then termination otherwise $G^i = G+1$
end procedure

Proposed approach follows the concept of NEPDE-CB; this procedure is explained in Algorithm 2.

Algorithm 2 NEPDE-CB (/current-to-best /1)

1: Start NEPDE-CR
2: Population initialization.
4: F ∈ [0.1 to 2]
5: Cr ∈ [0.1 to 1]
6: Evaluate the fitness value.
7: Apply mutation strategy:
$^{k}_{i}\vec{V}_{G} = \vec{X}_{R_1^i,G} + F_1^i (\vec{X}_{best} - \vec{X}_{R_2^i\ G}) + F_2^i (\sum_{i=1}^{NP} \vec{X}_{R_3^i\ G} * NEPV1 - \sum_{i=1}^{NP} \vec{X}_{R_4^i\ G} * NEPV2)$
Else $^{k}_{i}\vec{V}_{G} = X_i^G$ **end if**
8. If updated value $\leq \overrightarrow{\mathrm{X}}_{\mathbf{1},G}$, keep the last value
Else updated new value $X_i^{G+1} = \vec{X}_{\mathbf{1},G}$ **end if**
9. Choose best value and updated, $X_i^{G+1} + 1$
If conditions met, then termination otherwise $G^i = G+1$
end procedure

5 Experimental Result

Comparing and evaluating the performance of optimization algorithms is a troublesome and dreary task to perform. The COCO framework provides the platform of Black-Box Optimization Benchmarking (BBOB) to facilitate the procedure of single-objective optimization [8]. BBOB evaluates the performance of the algorithms with 24 noiseless text functions. NEPDE variants are tested in COCO framework as forgiven criteria: population size is 50, scale factor (F) [0–2], crossover rate [0–1], dimension low [2, 10] and dimension high [10, 20 and 40]. Proposed variants (NEPDE-R, NEPDE-CR) have been compared with other DE variants as PaDE, NDE, JADEb, MVDE, uBBDE, R-DE-10e2 and RL-SHADE-10e2 on the 2, 10, 20 and 40 dimensions with all functions (f_1–f_{24}).

Testing Framework and Benchmark Functions

BBOB approaches drive perturbation direction from an initial population of candidate solution. The standard of the optimization algorithms is critical to evaluate the performance of optimizers, comprehends the shortcomings and qualities of every algorithm and is a necessary way to assess new algorithm structures. The search space is drive −5 to 5 like $[-5. 5]^D$. Most of the benchmark functions having optima in [−4, 4] search domain. Fifteen instances of each function were taken out in each dimension. Termination condition is done when the arrangement with the accuracy of more than 10–8 fitness is acquired.

At whatever point accomplishes the most significant number of function evaluations, the termination is done. The target function is achieved as $F_{\text{target}} = \partial F + F_{\text{opt}}$, where F_{opt} is the optimal solution and ∂F is required accuracy for each benchmark function individually. The experimental analyses were done on MATLAB 2017 using machine with Intel Lenovo Core i5 CPU with a speed of 4.1 GHz, (RAM) 12 GB, operating system Windows 10 Pro 64-bit-based processor.

Result Analysis: The experimental results were performing in MATLAB 2017 and tested with the help of COCO framework. Proposed variant is compared with standard DE algorithms like PaDE, NDE, JADEb, MVDE, uBBDE, R-DE-10e2 and RL-SHADE-10e2 on the dimensions 2D, 10D, 20D and 40D for all functions (f_1–f_{24}). Table 1 describes the rank of the NEPDE variants along with state-of-the-art algorithm in terms of the different dimensions (2D, 10D, 20D and 40D). The datasets of these algorithms are available in https://coco.gforge.inria.fr/doku.php?

Table 1 Position of NEPDE variants with state-of-the-art algorithms over BBOB benchmark functions results

Dim.	Spar (f_1–f_5)	lcond (f_6–f_9)	hcond (f_{10}–f_{14})	multi (f_{15}–f_{19})	multi2 (f_{20-24})	All functions
2D	2-NEPDE-CR	2-NEPDE-CR	4-NEPDE-B	4-NEPDE-B	1-NEPDE-B	1-NEPDE-B
10D	1-NEPDE-B	2-NEPDE-CR	4-NEPDE-CR	3-NEPDE-B	1-NEPDE-CR	1-NEPDE-CR
20D	1-NEPDE-B	1-NEPDE-CR	4-NEPDE-CR	3-NEPDE-B	1-NEPDE-CR	1-NEPDE-CR
40D	1-NEPDE-CR	1-NEPDE-CR	3-NEPDE-CR	4-NEPDE-CR	1-NEPDE-CR	1-NEPDE-CR

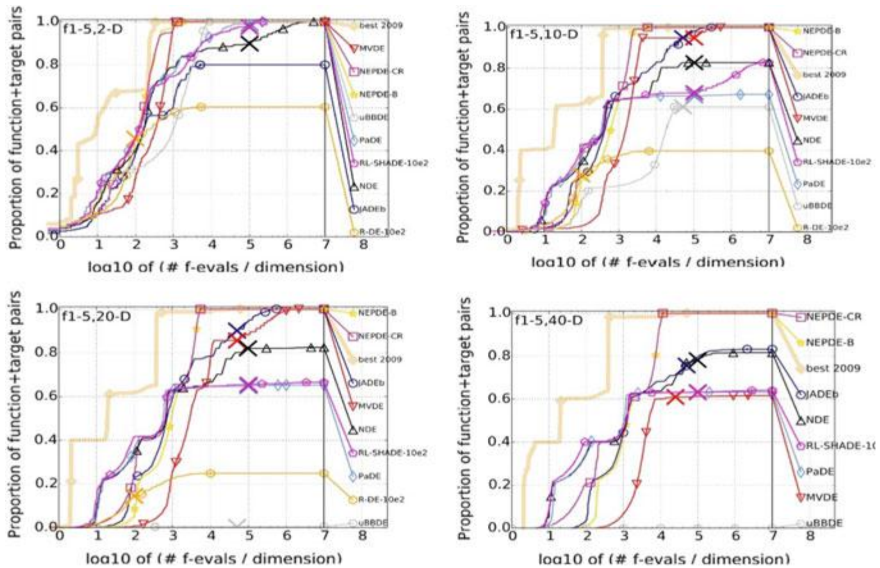


Fig. 1 Comparison of NEPDE with other DE variants on separable function

id=algorithms-bbob [8]. A total of 24 noiseless text functions are divided as (f_1–f_5) is separable functions; (f_6–f_9) is low or moderate conditioning; (f_{10}–f_{14}) is high conditioning and unimodal; (f_{15}–f_{19}) is multi-modal with adequate global structure; (f_{20}–f_{24}) is multi-modal with weak global structure. Comparative analysis of NEPDE variants done with state-of-the-art DE variants are shown in Figs. 1, 2, 3, 4, 5 and 6, and the analysis has been done within 2-D, 10-D, 20-D and 40-D. Table 1 shows the rank of the NEPDE variants in different dimensions, which are measured by the result of Figs. 1, 2, 3, 4, 5 and 6.

6 Conclusion

In this paper, a NEPDE is challenging to improve the convergence rate and the performance in terms of the environment-based mutation, which enhances the convergence speed and maintains the diversity of the DE algorithm. This article proposed new environmental-based parameters as NEPDE-B and NEPDE-CR for DE, which helps to improve the performance and solve the stagnation problem in the local search space. In the DE algorithm, start the maximum generation of function evaluation then apply the new environmental-based mutation for exploration and exploitation. Whenever it goes to higher dimensions in the search space, this algorithm gives better results and improves the convergence rate. Comparative analysis is done with COCO framework for real-parameter global optimizers with 24 benchmark functions of CEC2015 and finds a better result compared with the state-of-the-art algorithms

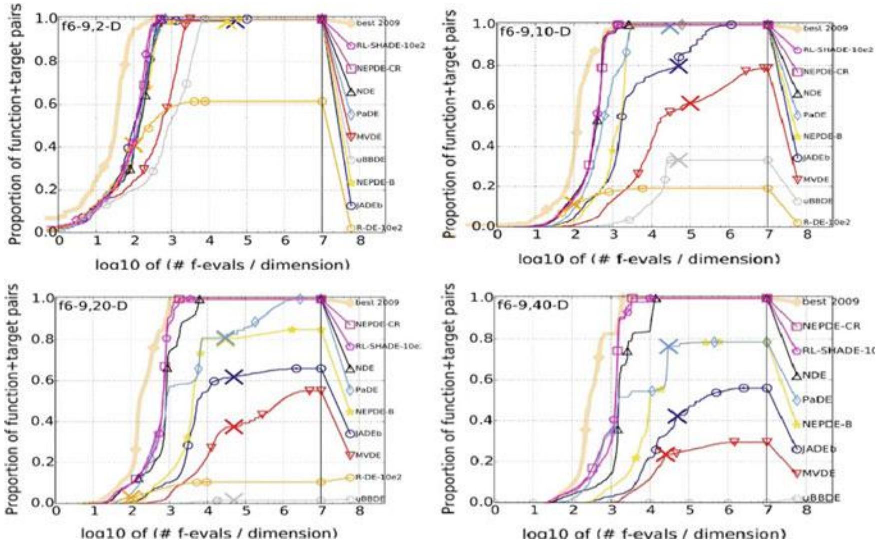


Fig. 2 Comparison of NEPDE with other DE variants on low or moderate conditioning function

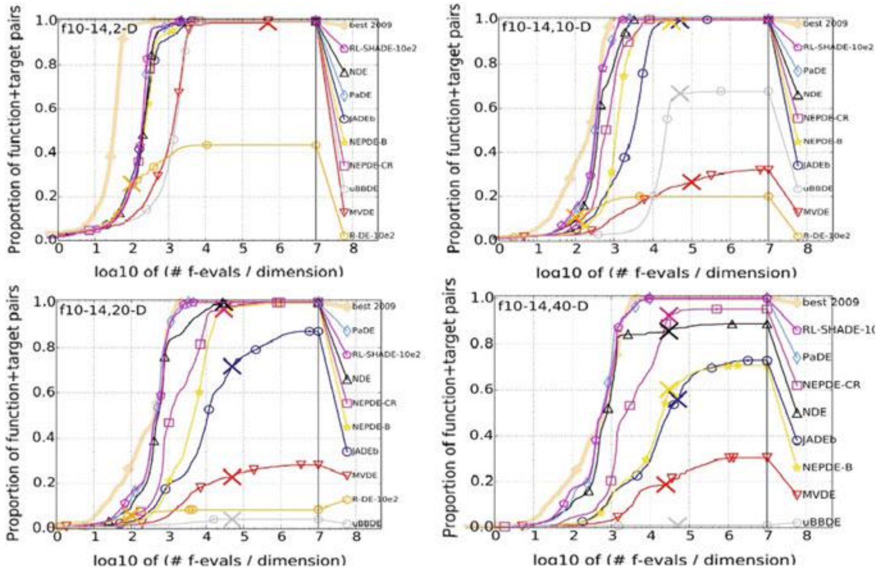


Fig. 3 Comparison of NEPDE with other DE variants on high conditioning and unimodal functions

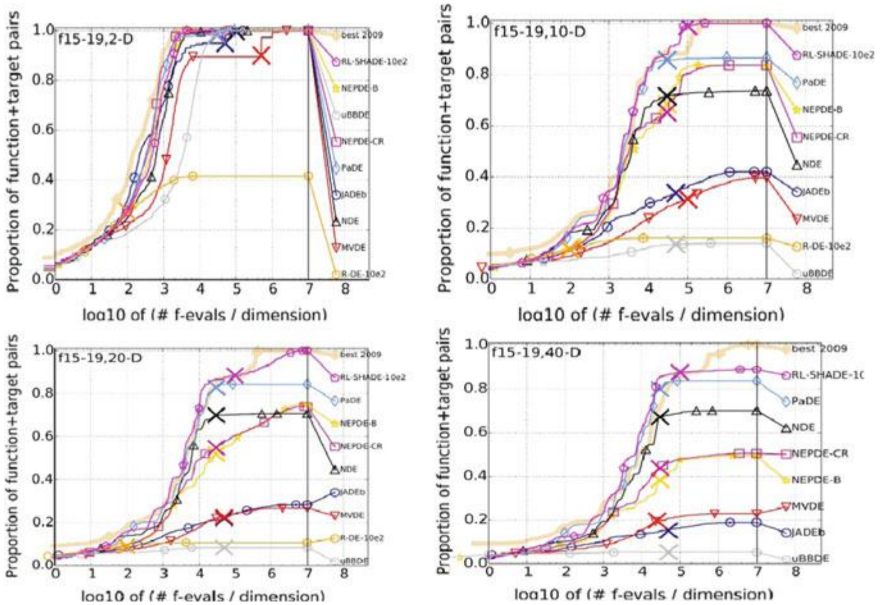


Fig. 4 Comparison of NEPDE with other DE variants on multi-modal with adequate global structure

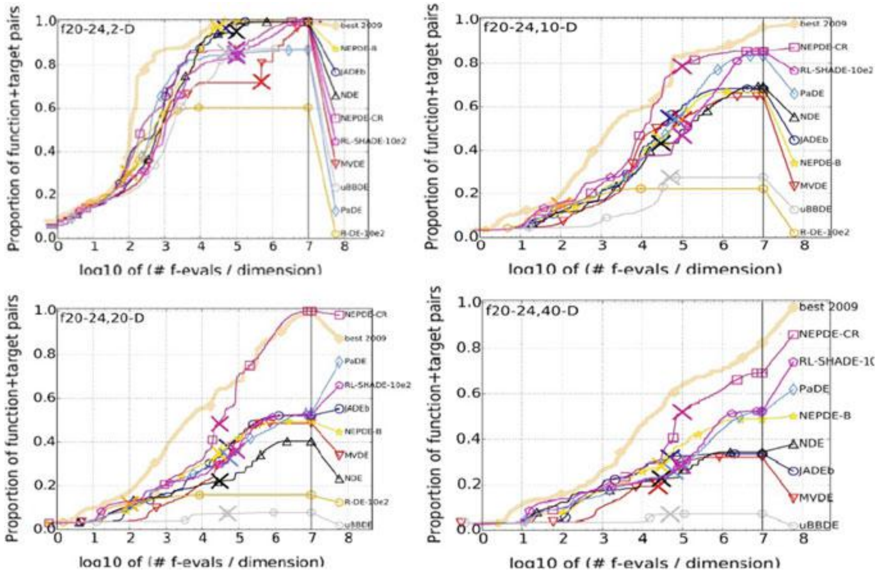


Fig. 5 Comparative analysis of NEPDE variants on multi-modal with weak global structure

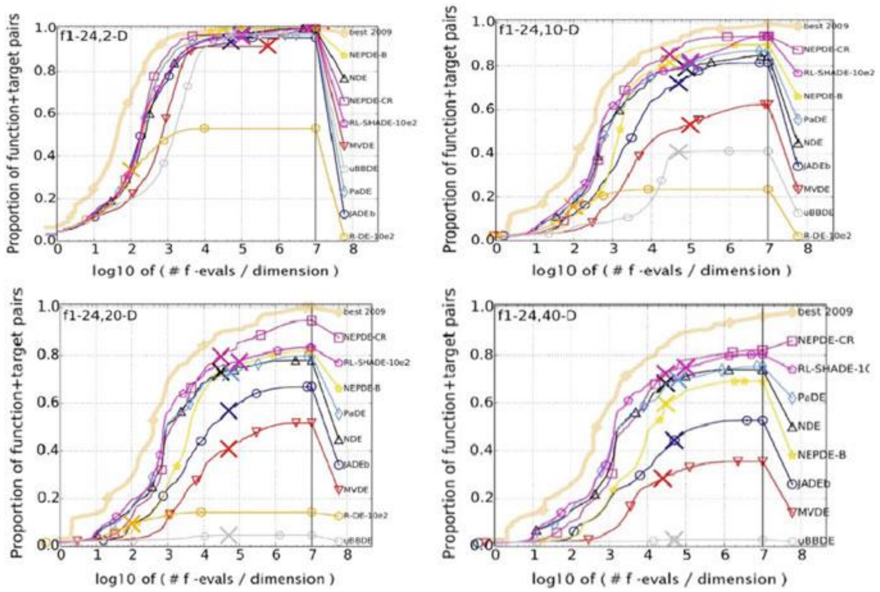


Fig. 6 Comparative analysis of NEPDE variants on all functions

(PaDE, NDE, JADEb, MVDE, uBBDE, R-DE-10e2 and RL-SHADE-10e2). This article maintains the diversity in local and global search spaces.

References

1. Storn, R., Price, K.: Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Optim. **11**(4), 341–359 (1997)
2. Singh, A., Kumar, A.: Dynamic selection of parameters in differential evolution algorithm. In: 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE (2018)
3. Mohamed, A.W.: A novel differential evolution algorithm for solving constrained engineering optimization problems. J. Intell. Manuf. **29**(3), 659–692 (2018)
4. Mohamed, A.W., Hadi, A.A., Jambi, K.M.: Novel mutation strategy for enhancing SHADE and LSHADE algorithms for global numerical optimization. Swarm Evol. Comput. (2018)
5. Cui, L., et al.: Adaptive differential evolution algorithm with novel mutation strategies in multiple sub-populations. Comput. Oper. Res. **67**, 155–173 (2016)
6. Qin, A.K., Huang, V.L., Suganthan, P.N.: Differential evolution algorithm with strategy adaptation for global numerical optimization. IEEE Trans. Evol. Comput. **13**(2), 398–417 (2008)
7. Zhang, J., Sanderson, A.C.: JADE: adaptive differential evolution with optional external archive. IEEE Trans. Evol. Comput. **13**(5), 945–958 (2009)
8. CEC-BBOB Homepage. https://coco.gforge.inria.fr/doku.php?id=algorithms-bbob. Accessed 27 Sept 2019

Reactive Power Optimization Approach Based on Chaotic Particle Swarm Optimization

Fengqiang Li, Lianjun Song, and Bo Cong

Abstract Reactive optimization is an important measure to ensure the reliable operation of the system. In view of the characteristics of reactive optimization of power grid, a method of reactive optimization of the distribution network based on the combination of local voltage stability index partition and improved particle group algorithm is proposed. First, the local voltage stabilization index of the system load node is calculated, the load node is sorted according to the size of the voltage stabilization index, the load node is selected as the collection of candidate compensation points and the electrical distance is combined with the electrical distance to partition it, and then, the system's optimal compensation point position and reactive compensation amount are obtained by improving the particle group algorithm. The method combined with the local voltage stabilization index and the electrical distance can narrow the range of the search, obtain the reasonable and effective candidate compensation area, improve the particle group algorithm to initialize particle diversity is better, with faster convergence speed.

Keywords Reactive power compensation · Electrical power system · Particle swarm optimization

1 Introduction

Particle swarm algorithms mimic the cluster behavior of birds, which seek food in a collaborative way [1]. The reactive compensation improves the operating voltage of the system, reduces the loss of the network, and improves the stable level of the system [2]. After the power system is reactively optimized, the voltage level can be changed, and the voltage is an important index of power quality. Reactive compensation and reactive balance of power system are the basic conditions to ensure voltage quality, effective control, and reasonable reactive compensation, not only to ensure voltage

F. Li (✉) · L. Song · B. Cong
Department of Marking, State Grid Liaoning Electric Power Supply Co. Ltd., Shenyang 110004, China
e-mail: fengqingliby@163.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_12

quality, reduce power loss of the power grid, but also can effectively improve the stability and safety of power system operation and give full play to the economic benefits of power grid operation. PSO is an evolutionary algorithm, but it is not used to update the genetic chromosome number of a sub-gene, but is similar to the gradient descent algorithm, so each chromosome plays the highest role in the population [3].

Therefore, reactive optimization is a hybrid nonlinear planning problem of multivariable, multi-constraint, and multi-objective [4, 5]. The algorithm studies the captured behavior of birds or fish by simulating a simple social system and continues to evolve [6, 7]. The alternating decoupling strategy is adopted in the power system protection and control optimization. The nonlinear in point algorithm is combined with PSO and has good optimization ability and running speed [8, 9].

The PSO algorithm clusters too fast, and it is light to obtain local assembly, resulting in low assembly accuracy [10]. The weight coefficient and inactive particles of the algorithm are improved. At the same time, the particle group is optimized by the slice method, which makes the overall optimization speed and effect better [11].

2 Particle Swarm Optimization

Consistent with the definition of the function, measuring the range of each particle is a superior solution. The best solution for the entire spatial search based on the purpose of each particle and other particles. Parallel computing uses feedback principles and search techniques to handle nonlinear problems [12].

PSO is a group-based search algorithm. It assumes that particle m and particle S are filled in the n-dimensional search space. The general expression is $S = (Y_1, Y_2, Y_3, \ldots, Y_m)$, and the particle $Y_i = (Y_{i1}, Y_{i2}, Y_{i3}, \ldots, Y_{in})$ is $Y_i = (Y_{i1}, Y_{i2}, Y_{i3}, \ldots, Y_{in})$ and $i = 1, 2, 3, \ldots, m$. This kind of particle can be calculated by inputting the objective function. This institution has gone through the best position. It is believed that the position is the best position to accommodate the values $G_i = (G_{i1}, G_{i2}, G_{i3}, \ldots, G_{in})$ and $i = 1, 2, 3, \ldots, m$. For each generation, the particle speed and location are based on the following formula:

$$v_{id}^{k+1} = v_{id}^k + c_1 r_1 \left(G_{id}^k - Y_{id}^k\right) + c_2 r_2 \left(G_{gd}^k - Y_{id}^k\right) \tag{1}$$

$$Y_{id}^{k+1} = Y_{id}^k + v_{id}^{k+1} \tag{2}$$

The particle speed update (1) has three compositions: The first composition is the particle, called the memory item, which reflects the current velocity of the particle memory capacity. The second composition is the current location of the particles, which are the best thing in history. It is known as the distance of the self-conscious item itself. It is refreshed to reflect the impact of learning and particle motion on information as part of the resulting particle behavior. The third composition is the state of the particles in the crowd, called the collective cognition project, which is reflected in some cooperation between information sharing and particles.

The model of reactive power optimization generally includes an objective function, a powerful constraint equation, and a variable constraint condition.

$$\begin{cases} \min f(u,x) \\ s.t.g(u,x) = 0 \\ h(u,x) \le 0 \end{cases} \tag{3}$$

From an economic perspective, the system's operational network loss as a target function is minimized. Consider the functions of transformer tapping and minimum capacitance switching as well as empirical functions that minimize power loss and dynamic reactive power optimization [13].

In general, select the target function with the smallest loss of the net:

$$F = \min \sum_{k=1}^{n} P_{k\mathrm{loss}} = \min \sum_{k=1}^{n} G_{k(i,j)} \left[U_i^2 + U_j^2 - 2U_iU_j \cos(\delta_i - \delta_j) \right] \tag{4}$$

It should satisfy the power constraint of the power equation according to the reactive optimization model:

$$\begin{cases} P_i = U_i \sum\limits_{j=1}^{n} U_j \left(G_{ij} \cos \delta_{ij} + B_{ij} \sin \delta_{ij} \right) \\ Q_i = U_i \sum\limits_{j=1}^{n} U_j \left(G_{ij} \cos \delta_{ij} + B_{ij} \sin \delta_{ij} \right) \end{cases} \tag{5}$$

Inequality constraints of control variables include:

$$\begin{cases} U_{Gi,}\min \le U_{Gi} \le U_{Gi,}\max & i \in N_G \\ T_{Tj,}\min \le T_{Tj} \le T_{Tj,}\max & j \in N_T \\ Q_{Ck,}\min \le Q_{Ck} \le Q_{Ck,}\max & k \in N_C \end{cases} \tag{6}$$

The inequality constraint state variable is:

$$\begin{cases} Q_{Gi,}\min \le Q_{Gi} \le Q_{Gi,}\max \, i \in N_G \\ U_{Dj,}\min \le U_{Dj} \le U_{Dj,}\max \, j \in N_D \end{cases} \tag{7}$$

3 Improved Particle Swarm Optimization

3.1 Chaotic Initialization Based on Improved Ten Mapping

The basic particle group algorithm adopts the method of random initialization for the initialization of population position and velocity, which can sometimes not guarantee the traversal and diversity of particles, and the chaos has the characteristics of traversal, randomness, and regularity, so the population position and velocity are initialized by the chaotic sequence. In addition to ensure the randomness of initialization and the diversity of particle groups, the following expressions are as follows:

$$x_{k+1} = T(x_k) = \begin{cases} 2x_k & 0 < x_k \le 0.5 \\ 2(1-x_k) & 0.5 < x_k \le 1 \end{cases} \tag{8}$$

3.2 Standardized Empowerment Speed Update Formula

Set p_{best}^k which indicates the optimal position selected by the particle in each iteration, gives a reasonable weight ω_k $(0 < \omega_k < 1, \sum \omega_k = 1)$ to the selected optimal position of each particle during each evolution, weights all particles $\bar{p}_{\text{best}} = \sum_{k=1}^{n} \omega_k p_{\text{best}}^k$, replaces the best position $g_{k\text{best}}(k = 1, 2 \ldots)$ of the particle itself with $\bar{p}_{\text{best}}$, and gets the improved particle velocity update which is defined by Eq. (9).

$$v_i^{k+1} = w \cdot v_i^{k+1} + c_1 r_1 (\bar{P}_{\text{best}} - x_i^k) + c_2 r_2 (g_{\text{best}} - x_i^k) \tag{9}$$

The standard empowerment velocity Eq. (9), the particles in the group in the evolution process are not only to learn from their extremes but to learn from optimal position information for all contemporary molecules, thus enriching the molecules of the source of the learning information in a large way. To avoid falling into local optimal and speed up convergence, the influence of weight in the standardized empowerment proposed method should be considered, and then, the weight value is reasonable that can speed up the convergence and its accuracy. If the weight value is unreasonable, this will cause the convergence speed which is slow or even stop convergence. The standardized empowerment formula is as follows:

$$\begin{cases} a_k = \frac{\max f(p_{\text{best}}^j) + \rho - f(p_{\text{best}}^k)}{\max f(p_{\text{best}}^j) - \min f(p_{\text{best}}^j)} & j = 1, 2 \ldots m \\ \omega_k = \frac{a_k}{a_1 + a_2 + \cdots + a_n} \end{cases} \tag{10}$$

In formula, $f(g_{\text{best}}^j)$ represents the adaptation value of each particle selected for the best position during each iteration, and $\rho \ge 0$, ρ should be taken as small as possible, which can both guarantee $a_k \neq 0$ and ensure that a_k does not gap too much because of the value of ρ.

3.3 Adaptive Variation Operator

The adaptive mutation operator strategy is adopted to maintain the diversity of the population, avoid the local optimality in the early evolution of the population, and ensure the stability and local search ability of the population in the late evolutionary stage. The strategy is to mutate the position of the particle in the process of particle swarm evolution with a certain probability. At the beginning of the evolution, a large scale of variation is adopted, and the evolution process gradually reduces the scale of variation.

$$\begin{cases} \eta(t) = 1 - a^{[1-t/\mathrm{maxgen}]^b} & U(0,1) > p_t \\ x^{k+1} = x^k + \eta \cdot x^k \cdot U(0,1) & r \geq 0.5 \\ x^{k+1} = x^k - \eta \cdot x^k \cdot U(0,1) & r \leq 0.5 \end{cases} \tag{11}$$

In formula P_t, the probability of mutation is 0.5, and random number $U(0, 1)$, and if it is larger than P_t, the population particles will mutate; if it is less than P_t, the population particles will not mutate; $\eta(t)(0 \leq \eta(t) \leq 1)$ is the scale of variation of the population during the second evolution of t; $\alpha \in (0, 1)$; maxgen is the total number of evolutions; b is 2; r is the random number. It can be seen from formula (11) that in the early stage of evolution, the larger $\eta(t)$ value can ensure the search range of the particle swarm and the diversity of the population. In the later stage, $\eta(t)$ becomes smaller, which enhances the local search ability of the particle. The flowchart of the improved PSO algorithm is shown in Fig. 1.

4 Control Parameters of the Algorithm

The particle swarm optimization algorithm needs to consider factors such as the size and speed of the overall position of the boundary. The size of the population depends on the diversity of the individual population. The more people there are, the more likely they are to identify the best solution, but the amount of calculations and time will increase accordingly. The improved particle swarm optimization algorithm improves the search speed and the optimal solution location, but in order to prevent the PSO algorithm from convergence, it must be searched in a certain range.

While you may find the best solution under the influence of demographic information, it is not difficult to search for local optimizations in slightly more complex problems. Without the third part, particles lack social awareness, and there is no information sharing and cooperation between particles, but a search taken into account

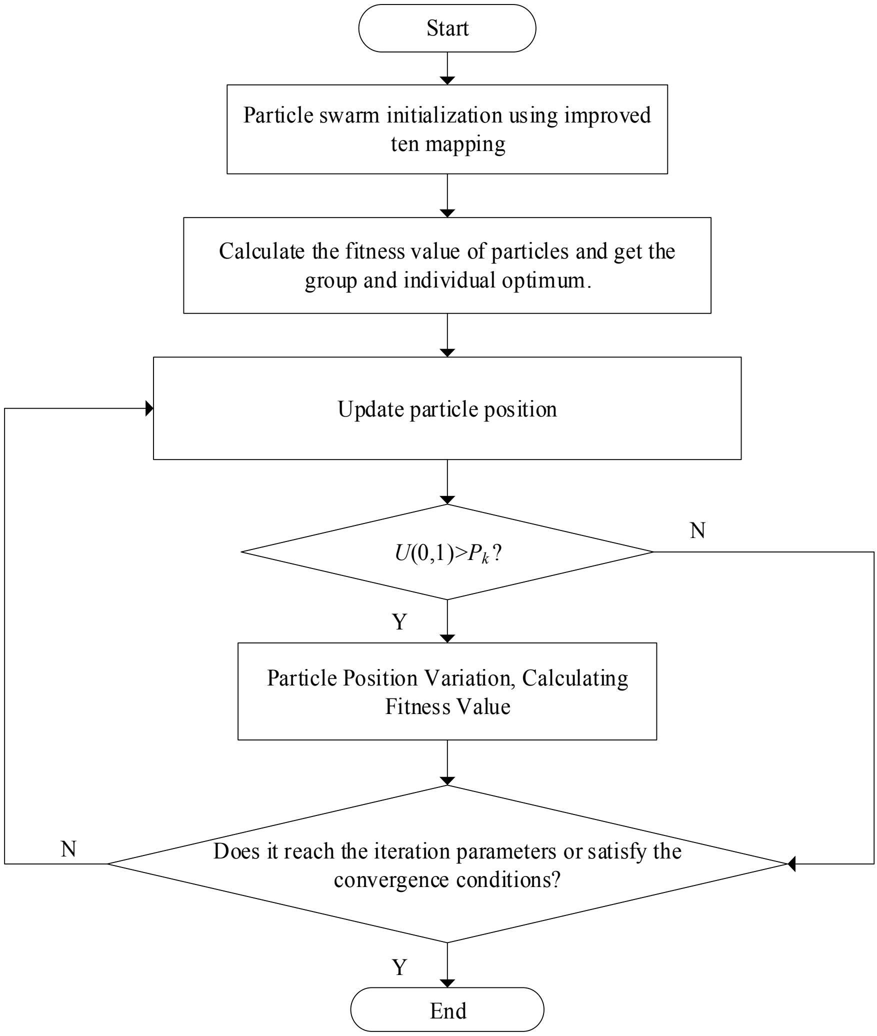


Fig. 1 Flowchart of improved particle swarm optimization algorithm

their personal experience, which is equivalent to the search optimization ability of a single particle.

5 Conclusion

In view of the fact that the results obtained by the step-by-step solution of compensation point and compensation capacity are not necessarily global optimal and the problem of being vulnerable to dimensional mass disaster is solved, a method of reactive optimization of the distribution network combining the local voltage stabilization index partition and the improved particle group algorithm is proposed, and the following conclusions are obtained: (1) The candidate compensation area obtained by combining local voltage index with electrical distance is reasonable and effective, the search space is reduced, as many feasible compensation points are retained, and the system network loss is smaller, and (2) improved particle swarm optimization algorithm initializes particles with better diversity and faster convergence.

References

1. Wang, C.: The managing and scheduling system for the disaster rescue equipment of PAP based on the internet of things. Int. J. Internet Things Big Data **1**(1), 21–28 (2016)
2. Sharma, P.: Prediction of heart disease using 2-tier SVM data mining algorithm. Int. J. Adv. Res. Big Data Manag. Syst. **1**(1), 11–24 (2017)
3. Kong, W.: Analyzing the usage of data mining in spatial database. Int. J. Internet Things Appl. **1**(1), 1–12 (2017)
4. Park, S.B., Kang, S.G., Lee, S.W.: Analysis of dissolution and sterilization effect of water-soluble Na_2O-ZnO-SiO_2 glass system for anti-marine biofouling application. Int. J. Bio-Sci. Bio-Technol. **9**(5), 25–30 (2017)
5. Park, J.S., Lee, S.H., Kim, Y.S.: The impact of HR capability on performance: the roles of Internal firm resources. Int. J. Transp. Logist. Manag. **2**(1), 1–6 (2018)
6. Woo, N.S., Kim, H.J., Han, S.M., Lee, W.D., Huh, S.S., Kim, Y.J.: The design optimization of subsea separator for gas-liquid mixed flow in subsea oil-well. Asia-Pac. J. Model. Simul. Mech. Syst. Des. Anal. **1**(1), 35–40 (2016)
7. Wei, F., Wang, G., Ren, B.Y.: A new feature evaluation and weighting method using three measurement hybrid model and its application to fault diagnosis of rolling element bearing. Asia-Pac. J. Model. Simul. Mech. Syst. Des. Anal. **2**(1), 39–46 (2017)
8. Kang, Y.D., He, Q., Yang, X.H., Huo, W., Yang, F., Mamtimin, A., Zhou, C.S.: Based on the earth's surface layer dust flux field observation and study on collection efficiency. Asia-Pac. J. Model. Simul. Mech. Syst. Des. Anal. **2**(1), 71–80 (2017)
9. Mirzal, A.: Delay compensation using the Smith predictor: a brief review with numerical examples. Int. J. Comput.-Aided Mech. Des. Implement. **3**(1), 1–10 (2017)
10. Iseman, Y.: Self-driving car—a computer will park for you. Int. J. Eng. Technol. Automob. Secur. **1**(1), 9–16 (2017)
11. Yang, Y.G.: Pedestrian navigation system using low cost AHRS for the visually impaired. Int. J. Comput. Graph. **8**(2), 7–11 (2017)
12. Bshwah, V.S., Tomar, G.S., Badauria, S.S.: Performance comparison of microstrip band-pass filters for different dielectric materials. Asia-Pac. J. Adv. Res. Electr. Electron. Eng. **1**(1), 31–40 (2017)
13. Gururani, H., Arya, R.: Review on low voltage and high speed dynamic comparator. Int. J. Energy Technol. Manag. **1**(1), 15–22 (2017)

Computing and Technology Trends

Data Mining Model for Better Admissions in Higher Educational Institutions (HEIs)—A Case Study of Bahrain

Subhashini Sailesh Bhaskaran and Mansoor Al Aali

Abstract Data mining has been used for a variety of objectives for improving the quality of higher education institutions and especially for improving students' performance and institution quality. The use of data mining for assessing prior learning and for improving the admission criteria has not been addressed extensively. Guiding applicants to select the correct and most suitable degree based on their prior learning at their previous institution is of great importance. We present in this paper our approach of using data mining for guiding applicants to decide the correct and most suitable degree based on their prior learning at their previous institution, and the results demonstrate the success of this method and confirm the expected benefits for the students and the institutions. The C4.5 decision tree algorithm is applied on successfully graduated student's prior learning data along with the GPA and programme in HEI in order to predict the programme of new applicants/students of similar prior learning characteristics. The outcome of the decision tree predicted the list of appropriate programmes with the GPA expected if registered in that programme, for the applicants from similar prior learning attributes. The decision rules present a list of choices of programmes to which new students can enrol with a hint of the success level expected in terms of GPA, which gives a forecast/projection on the success level that can be expected at the end of the study tenure. Furthermore, this knowledge can be used by advisors in preventing a student from enrolling to an inappropriate programme which would make the student fail from graduating.

Keywords Admissions · Data mining · Higher Educational Institutions (HEIs)

S. S. Bhaskaran (✉) · M. A. Aali
Ahlia University, Manama, Bahrain
e-mail: sbhaskaran@ahlia.edu.bh

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_13

1 Introduction

Because of the expansion in focused instructive environment, understudies have numerous options accessible to them; consequently factors that empower Higher Educational Institutions (HEIs) to draw and hold students ought to be truly considered. Since there is an absence of chances for the private higher establishments to get any monetary help from the administration, they need to depend more on the correspondence and component of the market. In this way, they have to concentrate more on better training administration to their students to pull in and obtain a few new students to make due in this solid challenge.

With expanded challenge in the part, duty absolutely lies on HEI—Higher Education Institutions to be progressively translucent to their clients when giving data in regard to their quality estimation strategies. Because of the accessibility of the choices, organizations are probably going to be always reviewed by the clients until and except if right confirmation is not given [1–3].

Since fulfilling students' prerequisite and keeping up their place in the challenge is not simple, HEI are getting to be testing nowadays. Administration quality as per the writing has been viewed as the indispensable aggressive edge; in this manner, the presentation of the establishment is appended to the administration given to the students [2]. In this manner, viable frameworks must be actualized in the HEIs to manage understudy execution [4, 5].

Researchers have identified many challenges for Higher Educational Institutions of which prior learning assessment is one among them. PLA could directly affect the student performance and success; hence, it is one of the vital challenges related to student success/performance and in turn to the institution as well. It is also essential to assess prior learning of the student and guide the student during the entry process which is admissions. Generally, PLA is used in HEI as an evaluation and suitability for entry to a university programme or degree for which credits are awarded [6]. As PLA earns credits for the students, and as it is an entry, criteria it becomes the responsibility of the HEI to guide the student to the right degree based on the prior educational background and score of the student resulting in an increased interest in PLA. However, research on PLA and evaluation of the procedures involved are still very limited.

Due to the impact that results of PLA have on the career of the student and of the HEI, this assessment can be regarded as high stake, which in turn makes the study of these procedures particularly important from a validity perspective. Validity refers to the extent to which the proposed interpretations and uses of the assessment are appropriate and plausible [7]. There are a number of challenges related to validity in these procedures, and one of these challenges has got to do with the trustworthiness of decisions made in relation to PLA in HEI. Hence, there is a need to study PLA from a more scientific perspective with the records of graduated students which will guide the new applicants.

From previous researches, it was found that there are some students who come with good scores from high school to the university and are not able to perform better or maintain the same performance as in school mainly because of the reason that they are admitted to the wrong degree programmes which is the first step. As the very first step is wrong, the whole process of the student's education in the university is likely to go wrong in most cases. Hence, this study attempts to give a solution to the above-mentioned problem by recommending degree programmes by ranking them to applicants/prospective students based on its appropriateness to the prior learning background of the applicant/prospective students and that could lead to successful completion of the degree programme, based on graduated students records coming from similar background, during admission process.

2 Related Work

Arsad et al. [8] tried to predict student performance using parameters like educational background, CGPA and gender. They used artificial neural network (ANN) to predict student performance. The researchers found that students with good foundation were able to perform better. This makes educational background as one of the prominent factor in predicting student performance.

Mohsin et al. [9] tried to predict student performance using parameters like academic background, personality type, demographic and grade in programming by using the rough set method. They found that student prior experience in programming before university can contribute to a good grade but it is not guarantee that they can easily score the subject.

Akinola et al. [10] tried to predict student performance using result obtained during matriculation examination using artificial neural networks. They used a tool NeuroShell Classifier.

There have been lot of studies on whether prior learning has any effect on the performance of students in HEI. Specialists like Cortez and Silva [11] have reasoned that students' accomplishment is very much linked with their performance in the previous years, and with other scholarly, social and cultural attributes of the students. Analysts like Ramaswami and Bhaskaran [12] also concur that prior learning appears to add to students achievement. Adelman [13], Berkner et al. [14] and Hoachlander guarantee that understudies' scholarly planning in secondary school is firmly identified with HEI results. Astin [15] found that secondary school evaluations and Scholastic Aptitude Test scores were identified with academic achievements.

From the above-mentioned literature, it can be seen that most researchers consider educational background (prior learning assessment) as one of the major factors that contribute to student performance. It can also be seen from the extensive literature that most of the researches have tried to link PLA with other factors such as family background, demographics and GPA, contributing to student performance, and there are very less studies that have analysed all attributes of prior learning and that focuses on admission point of view. Hence, it can be seen there is a need to focus on different prior learning attributes and how students and institutions can benefit from the prior learning characteristics stored in any Student Information System. This study is different that it does not use prior learning to predict student performance which has already been addressed by many researchers, but it uses prior learning to predict the appropriate programme to which the student should be admitted to, which will give the student success. However, it is at the discretion of the student to take up the advice given by the institution on the programme to be enrolled.

3 Purpose of the Study

The purpose of this study is to identify appropriate programme during admissions that would give success for students based on their PLA using data mining techniques. With regard to PLA, this study is different from other studies because it does not only consider the results of prior learning but also analyses the different attributes of PLA such as the medium or language of study, system of study, branch of study and English certifications like IELTS, TOEFL, etc., and has attempted to find out their order of impact on predicting student programme. From student perspective, this study examines how PLA can be used to predict the best degree programme for the applicant during the admission process in order to achieve high grades or marks. From the institution perspective, this study helps the advisors, academicians and registrars to guide the student to the right programme which could give success to the student.

Research Questions

1. *How can data mining help to have a better understanding of PLA?*
2. *Can this understanding be utilized by the university management to enhance the admission decisions?*
3. *Will patterns emerge from this study of PLA which can be used to predict the programme or degree to which the student has to be admitted based on successful student records with similar background?*

4 Data Mining Methodology

Data mining has made a ton of enthusiasm for the data innovation industry, because of the presence of enormous volume of information which is put away in different configurations like documents, writings, records, pictures, sounds, recordings, logical information and numerous new information designs. There is rising requirement for transforming such colossal information into important data and learning. The

Table 1 Predictors and variables used for the analysis

Variables	Values
Previous education GPA (PEGPA)	High distinction with first-class honours (HDFCH) 3.90–4.00, High distinction (HD) 3.75–3.89, Distinction (D) 3.50–3.74, Very good (VG) 3.00–3.49, Good plus (GP) 2.50–2.99, Good (G) 2.00–2.49
Current education GPA (CTGPA)	High distinction with first-class honours (HDFCH) 3.90–4.00, High distinction (HD) 3.75–3.89, Distinction (D) 3.50–3.74, Very good (VG) 3.00–3.49, Good plus (GP) 2.50–2.99, Good (G) 2.00–2.49
Previous education institute type (STYPE)	{Private, Public}
Medium or language of instruction (MED)	{English, Arabic}
Curriculum or system type (CRTYPE)	{American, British, Local}
Branch (BR)	{Commercial, Industrial, Literary, Science, Technical, General}
English certification (CERT)	{Yes, No}

information gathered from different applications requires a legitimate data mining system to extricate the learning from huge storehouses for basic leadership. Data mining, likewise called Knowledge Discovery in Databases (KDD), is the domain of discovering valuable data from voluminous information [16].

There are rising exploration interests in educational field to utilize data mining. This new rising field, Educational Data Mining (EDM), manages creating techniques that find learning from data arising from educational environment. The investigation performed in EDM research is frequently identified with procedures drawn from assortment of literary works [17]. EDM is an intuitive cycle of theory development, testing and refinement [18].
Data Mining Methodology for Prior Learning Assessment

A. Data Preparations
The data set used in this study was obtained from Bahrain. Graduated bachelor students were taken for the study. The size of the data is 430. In this step, data stored in different tables was joined in a single table after joining process errors were removed.

B. Data selection and transformation

In this step, fields that were required for the study were extracted. Few derived variables were also included. The information for the variables was extracted from the database. All the predictor and response variables which were extracted from the database are given in Table 1 for reference.

Figure 1 shows the detailed steps involved in the process of extracting hidden knowledge in terms of choices of programmes along with the G.

C. Modelling

The authors proposed to use decision tree as it was most suited for the requirement.

A tree where each branch hub speaks to a decision among various choices, and each leaf hub speaks to a choice is called as decision tree. It is predominantly utilized for separating data with the end goal of choice-making. Choice tree begins with a root hub. From the root, clients split every hub iteratively as per decision tree learning algorithm. The last outcome is a choice tree where each branch speaks to a

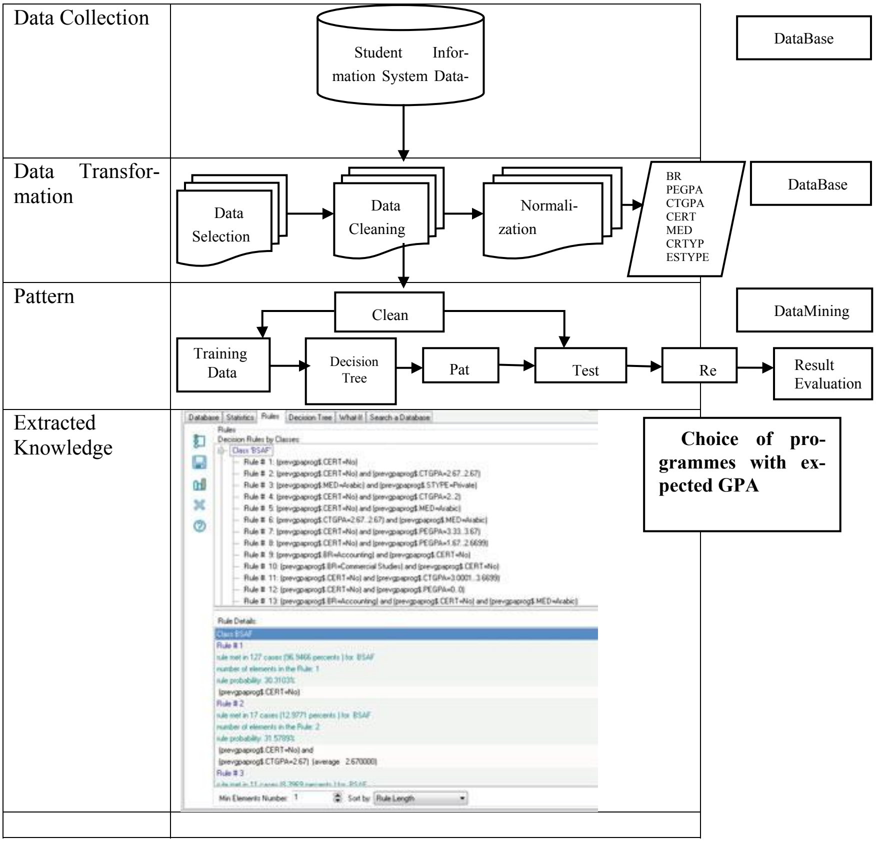


Fig. 1 Steps involved in generating rules for predicting the student programme during admission

conceivable situation of choice and its result. The three generally utilized decision tree learning algorithms are ASSISTANT, ID3 and C4.5.

5 Findings and Discussion

The importance of variables in predicting the programme of the prospective student or applicant is as follows:

Variable	Importance
BR (branch of study/specialization in the prior learning)	100.00
PEGPA (GPA scored in the prior learning)	74.624
MED (medium or language of instruction in prior learning)	4.382
STYPE (school or institution type)	1.343
CRTYPE (curriculum type)	0.722
CERT (English certifications)	0.161

The above table shows that by applying decision tree algorithms, it was found that the branch of study in the prior learning has 100% importance, prior learning GPA has 75% importance which signifies that to predict the current programme, branch of study and prior learning GPA are important and vital factors. BR (branch) has the highest gain; therefore, it is used as the root node.

The generated decision tree is shown as follows.

Figure 2 shows that the generated decision set rules can be defined as

If BR = 'Accounting' and CERT = 'No' and STYPE = 'Public' or 'Private' and CRTYPE = 'Local' or 'British' or 'American' and CTGPA = 2.00 and PEGPA = 2.00 and MED = 'Arabic' then PROG = 'BSAF'
If BR = 'Banking and Finance' and CERT = 'Yes' and STYPE = 'Public' or 'Private' and CRTYPE = 'Local' or 'British' or 'American' and CTGPA = 3.00 and PEGPA = 2.00 and MED = 'Arabic' then PROG = 'BSAF'
If BR = 'Banking and Finance' and CERT = 'No' and STYPE = 'Public' or 'Private' and CRTYPE = 'Local' or 'British' or 'American' and CTGPA = 3.00 and PEGPA = 2.00 and MED = 'Arabic' then PROG = 'BSMIS'
If BR = 'General Studies' and CERT = 'No' and STYPE = 'Public' or 'Private' and CRTYPE = 'Local' or 'British' or 'American' and CTGPA = 3.00 and PEGPA = 2.00 and MED = 'Arabic' then PROG = 'BSAF'
If BR = 'Information Technology' and CERT = 'No' or 'Yes' and STYPE = 'Public' or 'Private' and CRTYPE = 'Local' or 'British' or 'American' and CTGPA = 3.00 and PEGPA = 2.00 and MED = 'Arabic' or 'English' then PROG = 'BSDSM'

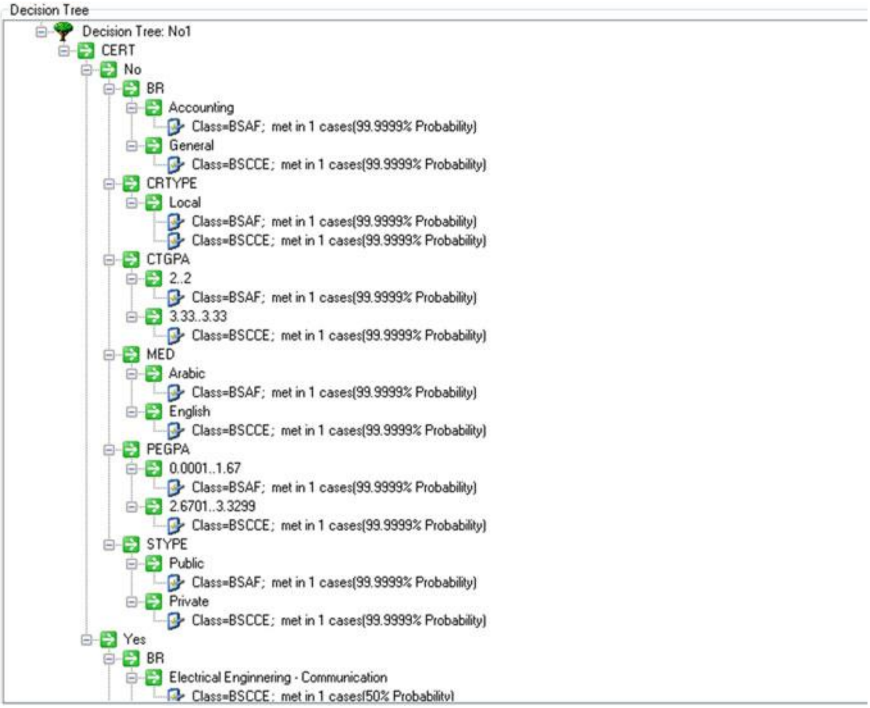


Fig. 2 Generated decision tree

The prediction of programme is shown in Fig. 3.

The following tables of predictions were generated taking one branch of study or specialization as sample, for instance, 'Science', and similar predictions of the programme could be made for other specialization or branch of study using the decision tree algorithm. Also the variables or predictors input could be varied, and different predictions could be obtained which indicates that the algorithm could result in personalized results which are very apt for the student based on the his/her variable inputs.

BR = 'Science' STYPE = 'Public' Prog to be admitted = 'BSPT'

From the generated decision rules, it could be understood that the rankings or priority of predicted programmes vary based on the input prior learning GPA/result, which means that a student who has prior learning result as 2.00, BR = 'Science', STYPE = 'Public' should be advised to take a programme (BSCCE) which is different (BSAF) from that of a student with prior learning result 2.33. As the predictor variables are that of graduated students, the prospective student or applicant can be advised or guided to follow the path of a successful student who came from the similar background.

From the rules generated, it could be understood that the rankings or priority of predicted programmes vary based on the input prior learning GPA/result and also

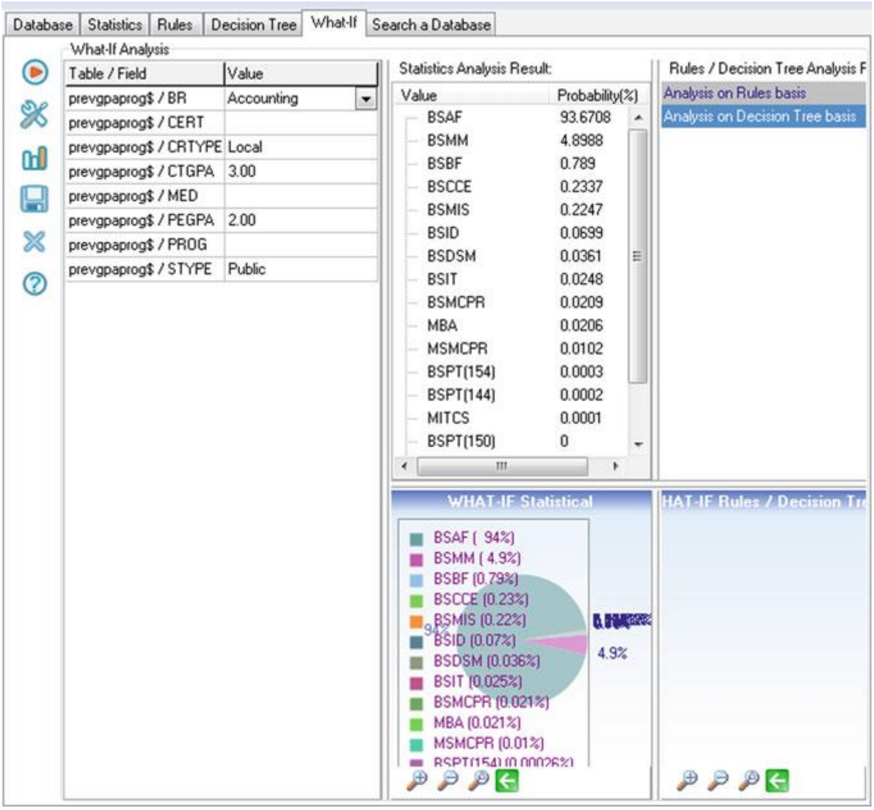


Fig. 3 Degree programme prediction

with that of school type, which means that a student who has prior learning result as 2.00, BR = 'Science', STYPE = 'Private' should be advised to take a programme (BSAF) which is different (BSPT) from that of a student with prior learning result 2.67. Also it can be seen that in some cases, when the school type changes, the predictions also vary even though the students' prior learning result was same which is 2.00 in this case. Hence there is a need to investigate on the dominating prior learning factors that affect the programme to which the student has to be admitted.

6 Conclusion

Using data mining prior learning attributes can be used to predict the right programme for an applicant in Higher Education Institutions. By doing so, we are benefitting from the data present in Student Information System to predict the programme from graduated student data belonging to similar background and prior learning which would be beneficial to advisors and registrars. Our research demonstrates clear benefits and applications of data mining when applied to prior learning in relation to historical data of current and graduated students.

References

1. Koslowski III, F.A.: Quality and assessment in context: a brief review. Qual. Assur. Educ. **14**(3), 277–288 (2006)
2. Hill, F.M.: Managing service quality in higher education: the role of the student as primary consumer. Qual. Assur. Educ. **3**(3), 1–11 (1995)
3. Russell, M.: Marketing education: a review of service quality perceptions among international students. Int. J. Contemp. Hosp. Manag. **17**(1), 65–77 (2005)
4. Mizikaci, F.: A system approach to program Evaluation model for quality in higher education. Qual. Assur. Educ. **14**(1), 37–53 (2006)
5. Snipes, R.L., Thomson, N.L., Oswald, S.L.: Gender bias in customer evaluation of service quality: an empirical investigation. J. Serv. Mark. **24**(4), 274–284 (2006)
6. Andersson, P., Harris, J.: Re-theorising the Recognition of Prior Learning. National Institute of Adult Continuing Education, Leicester, UK (2006)
7. Kane, M.T.: Validation. In: Brennan, R.L. (ed.) Educational Measurement, 4th edn, pp. 17–64. The National Council on Measurement in Education & the American Council on Education, Washington, DC (2006)
8. Arsad, P.M., Buniyamin, N., Manan, J.-L.A., Hamzah, N.: Proposed academic students' performance prediction model: a Malaysian case study. In: 3rd International Congress on Engineering Education (ICEED) (2011)
9. Mohsin, M.F.M., Norwawi, N.M., Hibadullah, C.F., Wahab, M.H.A.: Mining the student programming performance using rough set. In: Paper Presented at the International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 15–16 Nov 2010
10. Akinola, O.S., Akinkunmi, B.O., Alo, T.S.: A data mining model for predicting computer programming proficiency of computer science undergraduate students. Afr. J. Comput. ICT **5**(1), 43–52 (2012)
11. Cortez, P., Silva, A.: Using data mining to predict secondary school student performance. In: Proceedings of 5th Future Business Technology Conference, Oporto, Portugal (2008)
12. Ramaswami, M., Bhaskaran, R.: A CHAID based performance prediction model in educational data mining. Int. J. Comput. Sci. Iss. (IJCSI) **7**(1), 10–18 (2010)
13. Adelman, C.: Answers in the Tool Box: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment. U.S. Dept. of Education Office of Educational Research and Improvement, Washington, DC (1999)
14. Berkner, L., He, S., Forrest, E.: Descriptive Summary of 1995–96 Beginning Post Secondary Students: Six Years Later, NCES 2003-151. U.S. Department of Education, National Center for Education Statistics, Washington, DC (2002)
15. Astin, A.W.: How "good" is your institution's retention rate? Res. High. Educ. **38**(6), 647–658 (1997)
16. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence, p. 3754 (1997)
17. Baker, R.S.J.D.: Data mining for education. In: McGaw, B., Peterson, P., Baker, L. (ed.) International Encyclopedia of Education, 3rd ed. Elsevier, Oxford (2009)
18. Romero, C., Ventura, S.: Educational data mining: a survey from 1995 to 2005. Expert. Syst. Appl. **33**(1), 135–146 (2007)

The Effectiveness of Renewable Energies Projects in Kuwait—PAAET Solar Energy Project

Majed Alhajeri and Wathek Ramez

Abstract In Kuwait, 70% of electricity is generated from oil power station where the total irradiation on a plane equals 1891 kWh/m^2/year and the average monthly temperature rises about 42.3 °C, so the solar energy can be considered as a main renewable source of energy in Kuwait. 70% of produced energy is used in air conditioning and other services for households, while the commercial sector consumes only 30% of the produced electricity (42.58 billion kWh in 2010). The Public Authority for Applied Sciences and Training (PAAET) started to design Kuwait's first solar energy station, and this project is supported by Pan Arab Consulting Engineers (PACE) and was proposed to cover 17,542 m^2 with a total proposed production of 2046 MWh/Year. This research report discusses the proposed design in order to find the main problems causing low performance, especially the dust and the long period to get rate of return. A set of recommendation was listed.

Keywords Renewable energies · PAAET solar energy project · Kuwait

1 Introduction

Burning of fossil fuels produces non-renewable energy, which apart from the apparent threat of running out poses considerable risks to the environment [1]. Extracting as well as consuming them invariably also affects the economy, owing to the associated significant inflation rates throughout the globe [2]. This has led many countries to turn toward using and developing cleaner alternative energy sources, [3] solar energy being one of them. Increasingly backed by numerous researches, solar energy has been utilized to generate electricity—mainly via solar cells, such as photovoltaic (PV) cells—water heating purposes and a number of other useful applications [3]. Most developed countries in the recent times have successfully equipped themselves

M. Alhajeri (✉)
The George Washington University, Washington, USA
e-mail: allamh3@hotmail.com

W. Ramez
Ahlia University, Manama, Bahrain

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_14

with innovative technologies that can enable them to power entire cities sourcing just solar energy [4]. For the oil-rich gulf countries, where hydrocarbon fuels provide an abundant source of energy, studies and projects are being introduced proposing sufficient tapping of the also abundant solar energy. The conversion of solar energy into useful electrical and thermal energy has been given due consideration as the Public Authority of Applied Sciences and Training (PAAET) of Kuwait has commenced the design of the country's first solar energy station. The Pan Arab Consulting Engineers (PACE) has been funding this £10 million project, and the completion of which was scheduled to be by 2013.

The renewable energy projects should be clearly reviewed before implementation in order to find the benefit of these projects. PAAET project in Kuwait is one of those projects which will be reviewed in this dissertation because this project has high cost which takes into consideration Kuwait has a large oil production field. The main aim of this project is to study the suitability of applying PAAET project in Kuwait, so this study will conclude the validity of this project and how the output will be maximized.

The following questions will be answered in this research: What is the suitability of PAAET project depending on the cost and the revenue? What is the benefit of using renewable energies instead of oil fuel? Is solar energy suitable to be used in Kuwait? What are the main recommendations to enhance the performance of this project?

2 Theoretical Data

PV principle: Basically, the system of PV collects the solar energy from the sun and converts it into electrical energy. When the coming sunlight strikes the PV cell, then the light photons are absorbed by the semiconducting material. As a result of this absorption, the negative layer frees electrons which then move toward the positive layer. Electrons flow creates an electric current, which can be captured in a circuit that connects the two layers. The produced electricity can be directly used, sold back to a utility company or stored in batteries (Fig. 1).

Design steps of PV cell: The followed steps when a design for PV cell is required involve: Appropriate system selection. Provide a suitable area for the PV system. Determine the resistant materials for sunlight and climate. Locate the array to minimalize shading from foliage. Making a model for the wanted building and design the electrical codes. Check that the design encounters the local utility.

PV system components: The basic components for the PV cell are presented in Fig. 2.

PV module (PV Array) The PV module is the part that involves a group of PV cells, and it is the main responsible about converting the sunlight into electrical energy. Commonly, it is made from silicon (Si) or a thin film of silicon, polycrystalline and cadmium telluride. After reviewing some PV cells available in the market, it was found that most of these cells have

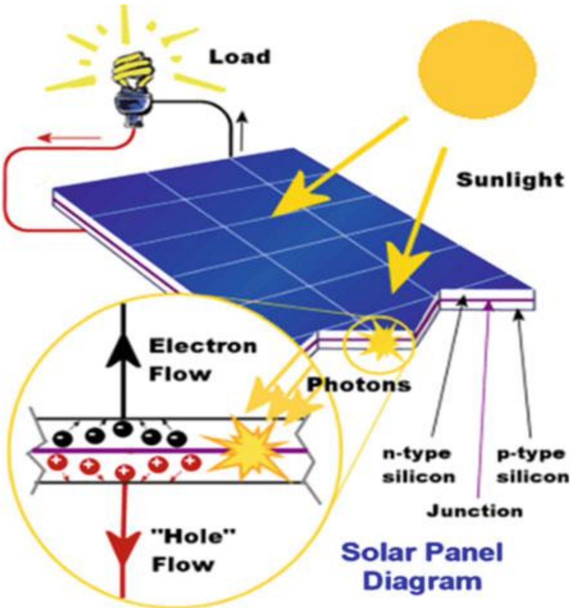


Fig. 1 Principle of work for PV cell [4]

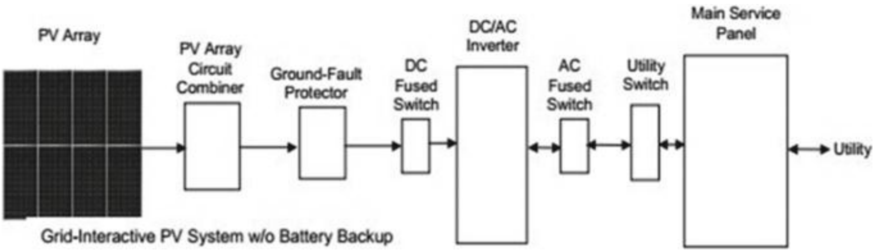


Fig. 2 Photovoltaic system without batteries

8 m^2 (25 ft^2) to 1.6 m^2 (5 ft^2) area. Surface density of 20 kg/m^2 (4 lb/ft^2).

Wiring and mounting systems: The second components are the wiring and mounting systems which usually denote to "balance of system equipment" (BOS). The main function of these components is to accumulate the solar models into the electrical and structural systems.

DC/AC inverter: The inverter is the power control unit that is used for alternating-current loads.

Gauges and meters: Gauges and meters (metering) are utilized for PV system evaluation and monitoring.

PV System Performance: The amount of the produced electricity by the PV system based on the sunlight intensity which is not constant in all days. For this reason, the produced current from the PV system could be also not constant (variable). The overall performance of the PV system is affected by the some factors which are

Standard Conditions of Test: This term indicates to the main conditions which are rated and created by the manufacturing process of the system during testing process where the DC current that generated by solar system is fixed and calibrated by the manufacturers under general conditions of test. The significant of the previously mentioned conditions is to allow the user to assess and compare the system simply. The summary of these conditions as the following (California Energy Commission, 2003):

- Temperature of the solar cell is equal to 25 °C
- The percent of error (tolerance) $<= \pm 5\%$ of the rating,
- ASTM Standard Spectrum; solar spectrum as filtered by passing through 1.5 thickness of atmosphere.
- The intensity of solar (irradiance) is equal to 1000 W/m^2
- The output of solar system is equal to 100 Ws of power under STC.

Generally, the power that generated from this system is proportional inversely with the system temperature increasing. Also, in the normal case, the array temperature reaches to approximately (50–75) °C. The design efficiency as well as performance of typical system rises to approximately (89%) at the previously mentioned conditions where this system which is expected to generate 100 W (California Energy Commission, 2003).

Dirt and Dust Accumulation: As previously mentioned, the weather nature in Kuwait region encourages and enhances accumulation process of dirt and dust on the solar photovoltaic surface which will lead to reject some sunrays as well as decrease the generated power. Also, the possibility raining in Kuwait region is low; therefore, the solar array must be cleaned periodically using both automatic and manual cleaner. The researches in the field of solar energy illustrate that the reduction percent that occurs because the accumulation of dust on surface of the solar is equal to 7%. In addition, the system that must produce 100 W will produce approximately (93 W).

Mismatch and wiring losses: The electrical devices such as the wires which are used in the solar model have the extreme power production always with value lower than the total of definite extreme production power might be produced by the system. This difference between actual output and the design can effect on the solar system performance. On the other hand, the electrical devices and wires resistance are comparatively high by (3% losses in the power). A factor of the reasonable reduction for these losses is approximately reached to 95% or 0.95 (California Energy Commission, 2003).

Conversion losses from Dc to Ac: The DC power that generated by the solar system must be transformed to AC power using different types of diverters. During this process, some of energy is lost (California Energy Commission, 2003).

3 Mathematical Data

Design steps: In order to size the PV system, the first essential step is to consider the required energy from this PV which can be determined by illustrating all of the daily loads. These loads comprise anything that utilizes electricity from the power source like batteries, radios, televisions and lights. Some of these loads require electricity all the time like refrigerators, and others use electricity less like power saws. In order to find the total energy consumption, the wattage of the appliance must be multiplied by the number of hours it is used in a day, and then, the total required power is the summation of all resultant amount of power. The next step is to calculate the losses in energy by the system or what called the equivalent power from PV. The second step is the PV modules sizing. This step is needed because the amount of the produced power relies on the PV modules size. In this step, the main focus is on calculating the total produced peak watt which is associated with the climate and solar intensity for the wanted region, and it involves two steps which are

Calculating the total needed Watt-peak gradient for PV modules through dividing the total daily watt-hours per required by 3.43 to find the total Watt-peak rating required for the PV panels in order to achieve correct peak with taken into account the Watt-peak rates can be estimated from weather agencies.

Evaluating the required number of photovoltaic panels for the solar system through using the amount of output Watt-peak of the system where it should be divided by available PV modules in the project. This step gives the minimum number of photovoltaic panels.

The third step is inverter sizing. This means that: The inverter rating cannot be less than the total amount of watt uses. It must have a nominal voltage similar to that of the used battery. Related to the rule of thumb, the size of the inverter must be 25–30% bigger than the appliance total watt.

The fourth step is the battery sizing. After reviewing some of the available batteries for solar photovoltaic, it was found the mostly used type for PV cells is the deep cycle battery. This type can be discharged to low energy level and cycle (rapid) recharged for long time reaches to years and days. In order to size this battery, it must be considered that it should be large to save adequate power to run the system at the days that no sunlight appears and in the night. Sizing the battery or achieving the suitable size can be achieved through implementing the following equation:

$$\begin{aligned}\text{Battery Capacity (Ah)} = {} & \text{Total Watt} - \text{hours per day used by appliances} \\ & \times \text{Days of Autonomy } (0.85 \times 0.6 \times \text{nominal battery voltage})\end{aligned}$$

The above equation involves the following steps: Calculating the total Watt-hours per day is needed for the applications. The above value must then divide by 85% which refers to the losses in battery. Then, the depth of discharge must be considered through dividing the given value in the previous step by 0.6. The obtained value from the previous step is then divided by the nominal voltage of the battery. Finally, the number of autonomy days is multiplied by the final answer obtained from the previous step.

The final step involves solar charge controller sizing. In the selection process of the best solar charge controller, it must be considered that this controller can encounter the voltage value of the PV array, the batteries voltage. The selected solar charge controller must have a large capacity adequate to overcome the current generated from the PV array.

After reviewing some of the previous studies, it was found that for huge projects, the best solar charge controller type is the "set type." Also, the sizes of these controllers based on the overall photovoltaic feed current which is transferred into charge controller. Another factor that must be considered when sizing these controllers is the photovoltaic panel configuration wither it series or parallel configuration. The equation that expresses the solar charge controller sizing is presented below. In this equation, the total short circuit current of the PV array is multiplied by a constant value which is 1.3.

$$\text{Solar charge controller rating} = \text{total short circuit current of VP array} \times 1.3$$

Solar Radiation theory

The direct solar radiation refers to the directly coming radiations from the sun and received at the surface of earth. This direct radiations spread on the earth in a semi-circular orbit as presented in Fig. 3.

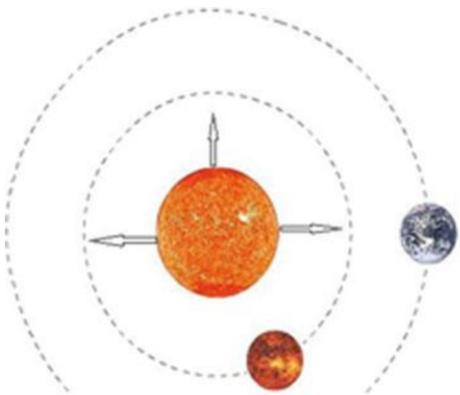
Useful equations: In this part, some of the useful equations in calculating the solar energy are illustrated.

Radiation energy

Firstly, the equation that expresses the radiation energy from the sun is called the Stephen–Boltzmann equation, and it is given as followed:

$$q = \sigma \varepsilon A(T_{\text{Source}}^4 - T_b^4)$$

Fig. 3 Direct sun radiations toward the earth surface



where q rate of energy transfer (power) (Watt). σ Steven-Boltzmann constant (5.67 * 10^{-8} W/m^2 K^4). ε emissivity of space. A radiation area (m^2). T_{Source} the temperature of the surrounding (K). T_b the temperature of the body that accepts the radiation (K).

Sunlight intensity: The following equation expresses the overall intensity from the sun:

$$L_{\text{total}} = 4\pi R^2 \cdot F_{\text{sun}} = 4\pi R^2 \cdot \sigma \cdot T^4$$

R: the radius of the sun (m). F is the flux.

4 Results and Analysis

Given Kuwait's typical desert climate that is accompanied by the coastal humidity, the overall nature of the climate can be said to be comparatively extreme, as marked by the significant variation in temperatures throughout the year. Table 1 presents the climatological information of Kuwait's climate based on a 15 year period database of monthly averages.

The wide range of temperature is apparent from the table above. Also illustrated is the monthly average temperature recorded between June of 2011 and April of 2012 in Fig. 4 that follows [5]. Based on this information, a high potential is seen for solar energy in Kuwait.

This project is aimed at provided an effective alternative to Kuwait energy needs; the latter is currently sourced as much as 70% from fossil fuels as the thermal power stations continuously produce electricity from the petroleum resources. Since solar

Table 1 Climatological information of Kuwait

Month	Mean temperature °C		Mean total rainfall (mm)	Mean number of rain days
	Daily minimum	Daily maximum		
Jan	8.5	19.5	30.2	5
Feb	10.0	21.8	10.5	3
Mar	14.0	26.9	18.2	3
Apr	19.5	33.9	11.5	3
May	25.4	40.9	0.4	1
Jun	28.9	45.5	0.0	0
Jul	30.7	46.7	0.0	0
Aug	29.5	46.9	0.0	0
Sep	26.2	43.7	0.0	0
Oct	21.5	36.6	1.4	1
Nov	14.5	27.8	18.5	3
Dec	9.9	21.9	25.5	3

Fig. 4 Monthly average temperature in Kuwait from 6/2011 to 4/2012

power proves to be much more economical as compared to oil, immense favorable outcomes are expected for Kuwait, financially as well as environmentally. Solar energy is suggested to be accounting for 10% of Kuwait's total electricity generation which can be grid connected without much of compensation [6].

Case study: PAAET

The utilization of renewable energy resources is becoming an important global issue worldwide. In order to cope with this global trend, the development of solar energy in Kuwait as a sustainable energy resource is one of the visions of the PAAET (Public Authority for Applied Education and Training). It is known that the Arab Gulf region and thus the Kuwait have high solar radiation during the summer months. Therefore, the peak load demands during this summer period can be easily compensated by the solar energy. In addition to this vision, the PAAET aims to ensure an energy supply that is characterized by being sustainable and practical for Kuwait region.

For Ardiyah School Campus's (for primary education) engineering, the ministry of PAAET has contracted PACE. A PV system with the capacity in the range of 1–2 MW is to be used to cover a portion of the Campus's electricity demand. This PV system will be one of the first large PV systems in Kuwait. A clear vision for both PAAET and PACE is set. This vision is that the PV system must set an example for Kuwait. For such systems and beside the requirements of having high performance and high quality, these systems must serve and assist educational purposes. Thus, the optimal conditions of the campus must be taken into consideration when designing this system. Other essential parameters for the designing phase include risk analysis, capacity building and monitoring. The PACE and in order to complete their mission, they need some help. They can get this help from Ecofys, which has good experience (25 years of experience) in the PV installation field in various regions in the world. The below figure shows a prototype of Ardiyah School Campus.

School campus Design

The design process of this school as any other construction project needs a preliminary stage before going further with the complete design. The main considerations to take into account during this stage include

The solar system must be installed on the campus building's roofs.

The solar arrays are to be connected to the buildings' electrical points. These arrays are connected in the form of smaller batches.

The least capacity of the system must be 1 MW, but if 1.5 MW can be obtained, it is preferable. It is expected from operating this system to cover up to 20% of the campus annual electricity consumption.

The expected obtained results from this preliminary design process are listed below:

- Understanding the annual energy generation (per building) as well as the peak installed capacity definitions.
- Understanding the connections between the array and the building's electrical points.
- Setting the peak capacity of the PV films as well as the annual produced energy from buildings 13 A & B and 22 A & B.
- Estimating the total system's cost, Ardiyah School Campus total capacity, total installed capacity and total roof area.

Referring to the made assumptions for the ground cover's ratio, tilting angle and the distance from shading objects, the roof spaces available to exploit the solar incidence from the campus buildings are calculated. The ground cover ratio was taken as 57%, and the system was designed using the Tenesol, TE240-60P+ modules.

Three different configuration types were used to design the invertor; these three configurations are in fact corresponding to three different sizes. These three base configurations were upscaled according to building number as well as the total system size. The three invertor's configuration types are SMA, SB 1200 for 1–3 kW systems SMA, SB 3800 for 3–16 kW systems. SMA, SMC 10000TL for 10–100 kW systems.

In order to minimize the cost associated with the maintenance, installation and operation processes of the PV system, it is recommended to utilize a maximum of three different configuration inverter types. The total feasibility of the PV system at this site is equal to MW.

Total electricity production

Based on the simulated results obtained from the concept design, the roof-based system's total electricity production is calculated. With the given performance ratio equals to 77.7%, the annual production is equal to 1422 kWh/KW. Thus, the system will produce 2002 MWh per year.

Array connection to the network

In order to connect any renewable system to the local grid, it is of vital importance to follow certain regulations and standards. Since Kuwait has no such regulations framework for the renewable systems grid connection, this system and thus its connection's protocol are to be designed following the "European best-practise standards." The design of the interconnection of the system may vary as a result of the need for defining the regulatory issues. For instance, this may include regulating for extra switches for the breaker installed between the distribution boards and the invertors.

Thin film foil

Both the performance and yield ratios for the metal roofs, presented in the buildings 6, 13, 15 and 22, were simulated using the PVL144 laminate (having unisolar type). In PV solar applications, any roofs not facing the optimal south direction will necessarily result in losses in the solar irradiation. Here in this case, some of the sport facilities roofs' account for this problem. These buildings are building number 13 and building number 22. On the other hand, these roofs have no nearby shading objects, and thus, the only form of solar irradiation loss is associated of the tilting angle and the orientation. When designing the C-Si module, a maximum acceptable loss is taken to be 6%, this in fact consists of two types; shading and non-ideal oriented irradiation, the first contribute by 3.3% and the later by 2.7%. Three tilting angles, 3°, 8° and 13°, were used in simulating the corresponding irradiation losses. This was made for the four orientations, east, south, west and north. Table 2 represents the resulted irradiation losses. It is important to note that the north-oriented thin film laminates tilted with an angle more than 3° do not satisfy the pre-defined 6% total irradiation loss.

Potential for maximum capacity per m^2 for this section is equal to (65 Wp/m^2). Also, the maximum power feasible for this installation is equal to (0.65 MWp) and can produce approximately 971 MWh/year. Table 3 represents specified ground area, calculated system power as well as the annual production for the thin film roof area.

The previous chart represents the annual production for female and male building where building number 1 indicates to communal building for male and female, and also, building number 2 indicates to Physical Education Complex Sport Building. From the chart, it was concluded that the annual production for male building is larger than the annual production for female building.

Array connection to the network

Cost estimation

Table 2 Irradiation losses with different tilt and orientation

Tilting angle °	Orientation (azimuth angle °)	Irradiation loss
3	0	2.1
8	0	−0.3*
13	0	−2.1*
3	180	5.7
8	180	9.3
13	180	13.4
3	90	3.9
8	90	4.4
13	90	5.3
3	−90	3.9
8	−90	4.4
13	−90	5.4

* Indicates significant at 5%

Table 3 Yield values with different tilt and orientation

B # no	Building name	Usable roof area (m^2)	Peak capacity (kW)	Annual production (MWh/year)
B# 06	Female—communal building	504	31	45
B# 13	Female—physical education complex sport building	4224	261	390
Total for rest of buildings		**29,325**	**292**	**435**
B# 15	Male—communal building	384	24	35
B# 22	Male—physical education complex sport building	5431	335	500
Total for rest of buildings (boys campus)		**57,406**	**359**	**535**
Total for all film roofs		**86,731**	**651**	**971**

Within three years, the price of solar module has decreased more than approximately (60%). The figure below illustrates the 30 years learning curve of PV modules that has a (20%) decrease with every doubling of the production volume. Also, the shortage of silicon during (2005–2007) years resulted in a slight increase in prices in these years. Last year's overcapacity production of around 2 GW resulted in a module "dumping" on the world market. Module prices can currently be bought around 0.8–0.9 \$/Wp.

Economic of system

Table 4 presents the economics of the (1.40 MWp PV) system on Ardiyah Campus. Also, the total investment in this field is calculated to be approximately (\$3,620,000) taking into account that the yearly maintenance and operation cost are estimated to

Table 4 Estimated price for the system

System		
Total power installed	KW	1407
Annual production	MWh/year	2002
Estimated number of modules	#	6500
Investment		
Current investment CAPEX	M$	2.32
O&M Cost first year	k$	47
Economics		
LCOE, 25 years, 0% inflation*	$/kWh	0.072
Average production cost/kWh	$/kWh	0.1256

* Indicates significant at 5%

Table 5 Minimum system price range

System		
Total power installed	KW	1407
Annual production	MWh/year	2002
Estimated number of modules	#	6500
Investment		
Current investment CAPEX	M$	3.62
O&M cost first year	k$	54
Economics		
LCOE, 25 years, 0% inflation*	$/kWh	0.095
Average production cost/kWh	$/kWh	0.1256

* Indicates significant at 5%

be ($54,000) which include cleaning cost, insurance cost as well as general O&M costs (Table 5).

4.1 Building 6 Module Specifications

The PV systems on building 6 are designed with the Tenesol TE240-60P+ modules.

Inverter specifications

The PV systems on building 6 are larger than 30 KW and are designed with inverter configuration 3 as defined in the general design section (SMA's SMC 10000TL inverter).

Design and available roof area

Optimization 15o:
The shading losses from row to row are determined by the tilt of the each row and the path of the sun during the year. The optimal tilt of 15o, orientation to south and a diffuse irradiation of 30% in Kuwait are used to calculate the shading losses. In the view of PACE, the project has an experimental character and will be used as example for educational purposes. Therefore, PACE requested to design a system with optimal performance, better than international standards. We therefore have set an upper limit to the irradiation losses of 3.3%. This limit results in a distance *y* of 3.0 times the height of the module row. The pitch, which is defined as the PV arrays base to base distance, is 1.72 m with a standard PV module width of ±0.99 m (Fig. 5). The final 77 cm space between two module rows is just enough for walking space, closer distances are not advised, although German and EU standards would allow some closer distances.

Distances to shading object
The optimal distance from module to module is 3.0 times the height of the module row. We use a similar factor for the distances from shading objects positioned from the east to south to west direction of the PV array, in order to limit the shading losses

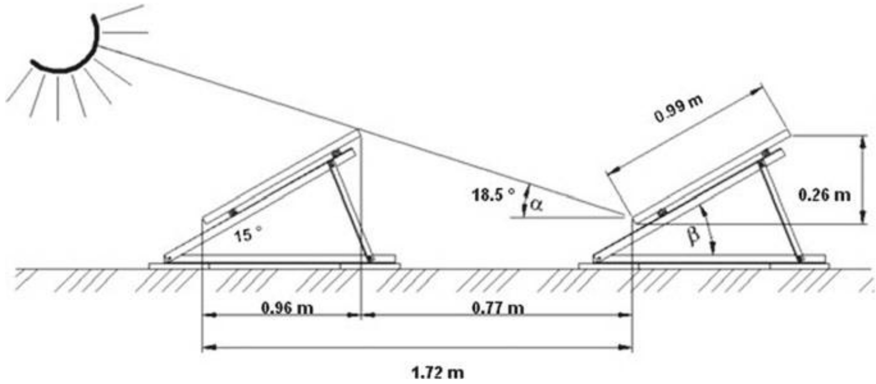


Fig. 5 Solar array configuration for calculation of shading losses

Table 6 Building 6

B # No	Building name	Usable roof area 2	Peak capacity kW	Annual production MWh/year
B# 06	Female—communal building	796	61.3	87

during the winter time, sunrise and sunset. This is better than the German standards. Distances from shading objects as close as 1.3 times the height of the shading object can be allowed according to the German standards. In the current calculations, we have used the most strict assumptions with maximal losses of 3.3%, as we can still reach up to 1.40 MWp. Therefore, the proposed optimal output design can be used as a minimum value to reach with a system designed for optimal output. We estimate an optional 10–15% increase possible in space and power if more losses are allowed according to the existing international standards.

Available roof area
The minimal roof area available on building 6 is shown in Table 6, and the following figure gives an indication of the position of the PV systems on building 6 designed with irradiation losses better than the maximum of 3.3%.

5 Conclusion and Recommendations

5.1 Conclusion

According to literature review, Kuwait is one of the richest countries in solar energy where the temperature in summer is around 50 °C, and also, the solar intensity and the wave length for the UV are suitable to generate a power in form of electrical or thermal. In Kuwait hot climates, the thin film technology has not been widely tested and practical, and hence, the present solutions are limited. The PAAET project is a new solar energy project in Kuwait to produce the electricity from the solar energy. The PAAET project would have an operation on land of area approximately equal to 17,542 m^2 to estimate the produced electricity about 2046 MWh/year, and this project will reduce the pollution and gas emissions but the foils bond to the roof or support structure, which can be facilitated by the hot climate. The rate of return for this project is more than 20 years depending on the price of the electricity locally and internationally. The rate of return of the project is very long, so the project should be optimized using a supplementary design like wind energy or using focus mirrors to increase the solar intensity.

5.2 Recommendations

Study the feasibility to use the solar energy in water distillation/desalination instead of PV technologies. Optimize the performance of the PVs using focus mirrors. Design an integrated system using wind energy to be operated during dust conditions to work as supplementary system. Use mathematical modules to get the effective cost of the solar cells. Use different software as solid work to represent the performance of solar cells technology.

References

1. Abdullah, A.H., Ghoneima, A., Al-Hasanb, Y.: Assessment of grid-connected photovoltaic systems in the Kuwaiti climate. Renew. Energy **26**(2), 189–199 (2002)
2. Al Mahdi, N., Al Baharna, N., Zaki, F.: Assessment of solar radiation models for the Gulf Arabian countries. Renew. Energy **2**(1), 65–71 (1992)
3. Alaruri, S., Amer, M.: Empirical regression models for weather data measured in Kuwait during the years 1985, 1986, and 1987. Sol. Energy **50**(3), 229–233 (1993)
4. Al-Hasan, Y.: Electricity generation cost between proposed photovoltaic station and conventional units in Kuwait. Renew. Energy **12**(3), 291–301 (1997)
5. PAAET ECOSYS, 2011, project data
6. Ghoneim, A.: Environmental impacts of grid connected photovoltaic system adapted for peak load reduction in Kuwait. Renew. Sustain. Energy Rev. **11**(2), 345–351 (2011)

Skin Lesion Analyser: An Efficient Seven-Way Multi-class Skin Cancer Classification Using MobileNet

Saket S. Chaturvedi, Kajol Gupta, and Prakash S. Prasad

Abstract Skin cancer is an emerging global health problem with 123,000 melanoma and 3,000,000 non-melanoma cases worldwide each year. The recent studies have reported excessive exposure to ultraviolet rays as a major factor in developing skin cancer. The most effective solution to control the death rate for skin cancer is a timely diagnosis of skin lesions as the five-year survival rate for melanoma patients is 99% when diagnosed and screened at the early stage. Considering an inability of dermatologists for accurate diagnosis of skin cancer, there is a need to develop an automated efficient system for the diagnosis of skin cancer. This study explores an efficient automated method for skin cancer classification with better evaluation metrics as compared to previous studies or expert dermatologists. We utilized a MobileNet model pretrained on approximately 1,280,000 images from 2014 ImageNet Challenge and finetuned on 10,015 dermoscopy images of HAM10000 dataset employing transfer learning. The model used in this study achieved an overall accuracy of 83.1% for seven classes in the dataset, whereas top2 and top3 accuracies of 91.36% and 95.34%, respectively. Also, the weighted average of precision, weighted average of recall, and weighted average of f1-score were found to be 89%, 83%, and 83%, respectively. This method has the potential to assist dermatology specialists in decision making at critical stages. We have deployed our deep learning model at https://saketchaturvedi.github.io as Web application.

Keywords Skin cancer · Dermoscopy · Classification · Convolutional neural network

S. S. Chaturvedi (✉) · K. Gupta · P. S. Prasad
Department of Computer Science & Engineering, Priyadarshini Institute of Engineering & Technology, Nagpur 440019, India
e-mail: saketschaturvedi@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_15

1 Introduction

Skin cancer is an emerging global health problem considering the increasing prevalence of harmful ultraviolet rays in the earth's environment. The researchers had discovered a further 10% depletion of the ozone layer will intensify the problem of skin cancer with an additional 300,000 non-melanoma and 4,500 melanoma cases each year [1]. Currently, every year, 123,000 melanomas and 3,000,000 non-melanoma cases are recorded worldwide [1–5]. The recent study on the prevention of skin cancer reports 90% of non-melanoma and 86% of melanoma cases induced by excessive exposure of ultraviolet rays [6, 7]. The UV radiation detriments the DNA present at the inner layers of skin, triggering the uncontrolled growth of skin cells, which may even emerge as a skin cancer [8].

The most straightforward and effective solution to control the mortality rate for skin cancer is the timely diagnosis of skin cancer as the survival rate for melanoma patients in a five-year timespan is 99% when diagnosed and screened at the early stage [9, 10]. Moreover, the most mundane skin cancer types BCC and SCC are highly treatable when early diagnosed and treated adequately [9, 11]. Dermatologist primarily utilizes visual inspection to diagnose skin cancer, which is a challenging task considering the visual similarity among skin cancers. However, dermoscopy has been popular for the diagnosis of skin cancer recently considering the ability of dermoscopy to accurately visualize the skin lesions not discernible with the naked eye. Reports on the diagnostic accuracy of clinical dermatologists have claimed 80% diagnostic accuracy for a dermatologist with experience greater than ten years, whereas the dermatologists with experience of 3–5 years were able to achieve diagnostic accuracy of only 62%, the accuracy further dropped for less-experienced dermatologists [12]. The studies on Dermoscopy imply a need to develop an automated efficient and robust system for the diagnosis of skin cancer since the fledgling dermatologists may deteriorate the diagnostic accuracy of skin lesions [13–15].

Although the method is complicated, deep learning algorithms have shown exceptional performance in visual tasks and even outperformed humans in gaming, e.g., Go [16], Atari [17] and object recognition [18], which has lead to conduct the research on automated screening of skin cancers [9]. Several studies have been done to compare the dermatologist level, and Deep learning-based automated classification of skin cancer [19, 20]. Esteva et al. reported a benchmark study comparing the performance of dermatologists and a CNN model over 129,450 clinical images, showing the CNN model performs at par or better than dermatologists [20]. In recent years, the trend has shifted to deep neural networks (DNNs) [21], which were proposed to overcome the drawbacks of previous models [9, 22–29]. Although DNNs require huge data for the training, they have an appealing impact on medical image classification [29]. The current literature mostly employs transfer learning to solve large dataset problem. Transfer learning is a method where a model trained over another similar task is finetuned for the given task. Mostly, the melanoma screening works employing DNNs have trained a network from scratch [25, 30], or employs transfer knowledge [23–27] from ImageNet. The main difference between the DNN architecture and implementation framework—Caffe [23, 26] is the most common framework, and ResNet [27], AlexNet [24], VGG-16 [28] are most common architectures.

Previous work in dermoscopic-automated skin cancer classification has lacked generality capability [29, 31, 32], and have not achieved pleasing results for multi-class skin cancer classification [20, 33, 35]. This study explores an efficient automated method for the classification of dermoscopy skin cancer images. We utilized a MobileNet convolutional neural network [36] pretrained on approximately 1,280,000 images from 2014 ImageNet Challenge [30] and finetuned on HAM10000 dataset [37] which contain 10,015 dermoscopy images employing transfer learning [38]. The MobileNet model classified skin lesion image with performance better or comparable to expert dermatologists for seven classes. We also conducted data analysis on the dermoscopy images of skin cancer from HAM10000 dataset to uncover the relation of skin cancer with several parameters to strengthen the understanding of skin cancer.

2 Method

2.1 Dataset

We have utilized HAM10000 dataset [37] for the training and validation in this study. HAM10000 dataset is a benchmark dataset with over 50% of lesions confirmed by pathology. The dataset consists of a total of 10,015 dermoscopy images, which includes 6705 Melanocytic nevi images, 1113 Melanoma images, 1099 Benign keratosis images, 514 Basal cell carcinoma images, 327 Actinic keratosis images, 142 Vascular images, and 115 Dermatofibroma images with 600×450 pixels resolution. Sample images of skin cancer types from HAM10000 are represented in Fig. 1.

2.2 Data Preprocessing

The preprocessing of skin lesion images was done by using Keras ImageDataGenerator [39]. The 57 null Age entries in the dataset were filled using the mean filling method [40]. The Dermoscopy images in the dataset were downscaled to 224×224 pixel resolution from 600×450 pixel resolution to make images compatible with the MobileNet model [36]. The 10,015 images in the dataset were split into the training set (9077 images) and validation set (938 images). The dataset images with no duplication in training data were selected for the validation set so that the authenticity in the validation process can be maintained.

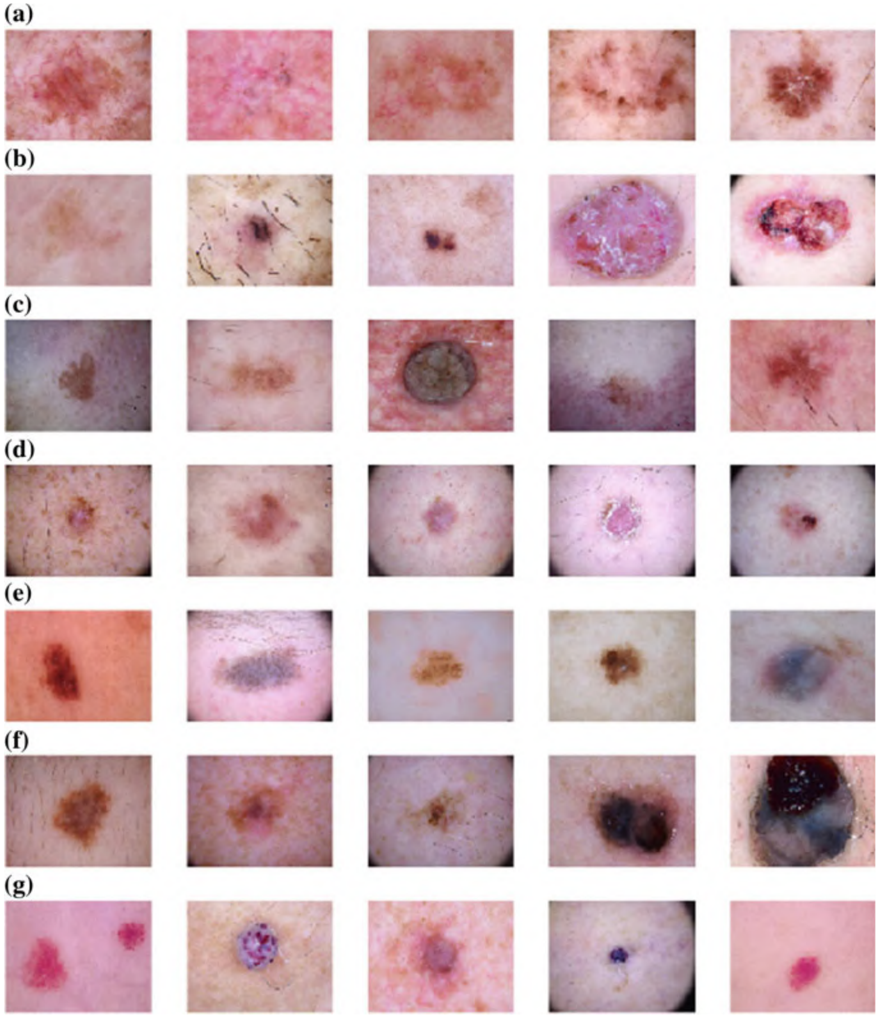


Fig. 1 Sample images from HAM10000 dataset for cancer types **a** Actinic keratosis **b** Basal cell carcinoma **c** Benign keratosis **d** Dermatofibroma **e** Melanocytic nevi **f** Melanoma **g** Vascular lesions

2.3 Data Augmentation

HAM10000 dataset has an unbalance distribution of images among the seven classes. Data augmentation [41] brings an opportunity to rebalance the classes in the dataset, alleviating other minority classes. Data augmentation is an effective means to expand the size of training data by randomly modifying several parameters of training data images like rotation range, zoom range, horizontal and vertical flip, fill_mode, etc. [41]. We conducted data augmentation of minority classes in the dataset: Melanoma,

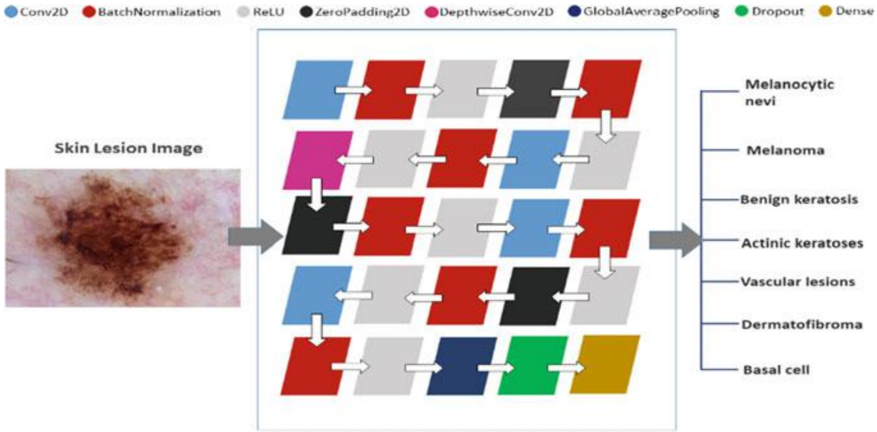


Fig. 2 MobileNet architecture used in the current study for the classification of skin lesion image among seven skin cancer types

Benign keratosis, Basal cell carcinoma, Actinic keratosis, vascular lesion, and dermatofibroma to generate approximately 6000 images in each class giving a total of 38,569 images in the training set.

2.4 Training Algorithm

The MobileNet model is ideal for mobile and embedded vision applications as they have lightweight DNN architecture [36]. We used MobileNet convolutional neural network [36] pretrained on 1,280,000 images containing 1,000 object classes from the 2014 ImageNet Challenge [30]. The 25 layered MobileNet architecture was constructed for the current study, which employs four Conv2D layers, seven BatchNormalization layers, seven ReLU layers, three ZeroPadding2D layers, and single DepthwiseConv2D, GlobalAveragePooling, Dropout, and Dense layers as shown in Fig. 2. The training of the model was done on a training set of 38,569 images using transfer learning [38] with batch size and epochs as 10 and 50, respectively. The categorical crossentropy loss function, Adam optimizer, and metric function accuracy, top2 accuracy, and top3 accuracy were used to evaluate MobileNet model performance.

2.5 Evaluation Metrics

The overall performance of the model was evaluated with several evaluation metrics: Accuracy, Micro Average of Precision (MAP), Micro Average of Recall (MAR), and

Micro Average of F1-score (MAF). The weighted average for recall, precision, and f1-score was evaluated by using the following mathematical expressions.

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

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

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{3}$$

$$\text{F1-score} = 2\left(\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}\right) \tag{4}$$

3 Results

The calculations were performed on Kaggle kernel having 4 CPU cores with 17 GB RAM and 2 CPU cores with 14 GB RAM [42]. Model evaluation was performed by calculating categorical accuracy, top2 accuracy, top3 accuracy, classification report, and confusion matrix. Further, the loss and accuracy curves were plotted to validate the model's performance for the optimization and prediction phase.

3.1 Dataset Analysis

The important observations recorded during the data analysis of the HAM10000 dataset are shown in Fig. 3 (i) Actinic keratosis, Basal cell carcinoma, Dermatofibroma, and Melanocytic nevi are not much prevalent below the age of 20 years. Whereas Melanoma and Vascular lesions can occur at any stage of life. (ii) The peak age for skin cancer is found at 45 years, while they are more common between the age of 30–70. (iii) Back, Lower Extremity, Trunk, Upper Extremity and Abdomen are heavily compromised regions of skin cancer.

3.2 Model Validation

The validation of the model was conducted on 938 unknown sample images from the validation set. We evaluated micro and weighted average for precision, recall, and f1-score to evaluate the MobileNet model performance on unknown images of the validation set. The Weighted Average of 89, 83, 83%, and Micro Average

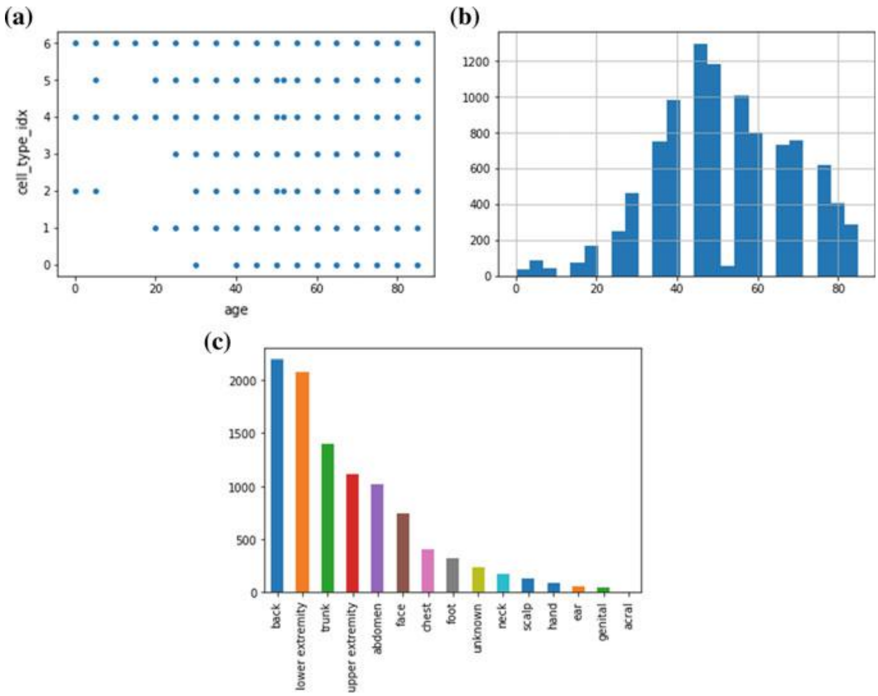


Fig. 3 Exploratory data analysis performed on the HAM10000 dataset **a** comparative study of skin cancer type on the *y*-axis with respect to age on the *x*-axis: the seven classes of the study represent 0, 1, 2, 3, 4, 5 and 6, respectively, on the *y*-axis. **b** Comparison of a number of cases of skin cancer on the *y*-axis with respect to age on the *x*-axis. **c** The number of skin cancer cases on the *y*-axis with respect to the location of skin cancer on the human body on the *x*-axis

of 83, 83, 83%, was recorded for precision, recall, and f1-score. The MobileNet model shows best precision, recall, and f1-score value for Melanocytic Nevi. The multi-class classification report showing micro average and weighted average for precision, recall, and f1-score are represented in Table 1.

The comparison of the current study with other related previous work is represented in Table 2. The majority of previous work is done on two or three classes, and their accuracies and recall vary between approximately 66–81% and 60–76%, respectively. In the study [20], they reported 48.9 and 55.4% classification accuracy evaluated for nine classes using CNN models. In the study [33], classification accuracy for ten classes using multi-track CNN was reported to be 75.1%. Also, in the study [34] they reported accuracy as 70%, 76%, 74%, and 67% for seven classes using InceptionResnetV2, PNASNet-5-Large, SENet154, and InceptionV4, respectively. In this study, we achieved categorical accuracy of 83.15%, top2 accuracy of 91.36%, top3 accuracy of 95.3%, and recall of 83% using MobileNet. Our seven-way skin cancer classification method has performed better than previously proposed computer-aided diagnosis systems in terms of both accuracies and recall.

Table 1 Multi-class classification report showing micro average and weighted average for precision, recall, and F1-score

Classes	Precision	Recall	F1-score
Actinic keratosis	0.36	0.38	0.37
Basal cell carcinoma	0.55	0.87	0.68
Benign keratosis	1.00	0.13	0.24
Dermatofibroma	0.21	0.50	0.30
Melanoma	0.28	0.69	0.40
Melanocytic nevi	0.95	0.93	0.94
Vascular Lesions	0.73	0.73	0.73
Micro average	**0.83**	**0.83**	**0.83**
Weighted average	**0.89**	**0.83**	**0.83**

Table 2 Comparison results of the current study with previous related work

Source	Year	Classifier	Number of classes	Accuracy %
[33]	2016	Multi-tract CNN	Ten	[a]75.1
[20]	2017	CNN	Three	69.4
		CNN-PA		72.1
		CNN	Nine	48.9
		CNN-PA		55.4
[34]	2019	InceptionResnetV2	Seven	70.0
		PNASNet-5-Large		76.0
		SENet154		74.0
		InceptionV4		67.0
	2019	Current study	Seven	**83.15** (cat)
				91.36 (top2)
				95.84 (top3)

[a]We have converted the recall and accuracy in percentage to compare them with the current study

Additionally, the proposed method is more efficient considering the faster processing capability and lightweight architecture of MobileNet.

3.3 Confusion Matrix

The confusion matrix for our model was evaluated for seven classes. Each element of confusion matrix shows the comparison between the true label and predicted label for each image in the validation set. Our model showed the best result for Melanocytic nevi by making a correct prediction for 696 images out of 751. Basal cell carcinoma and Melanoma were correctly determined for 26 images out of 30 and for 27 images out of 39, respectively. The diagnosis of Benign keratosis was most challenging due to their similar appearance with Melanoma and Melanocytic nevi. Only ten correct predictions were recorded for Benign keratosis.

3.4 Loss and Accuracy Curves

In order to examine learning, generalizing, and performance of the model, we computed training-validation loss curve (Fig. 4a) and training-validation accuracy curves for categorical (Fig. 4b), top2 (Fig. 4c) and top3 (Fig. 4d) accuracies. The model shows a good learning rate as the training accuracy increase with the number of iterations along with symmetric downward sloping of the training loss curve. The small gap between training and validation curves represents a good-fit, showing model can generalize well on unknown images.

We have developed a Web application to provide an effective automated online tool for the multi-class classification of dermoscopy skin lesion images. This Web application is available for social use at https://saketchaturvedi.github.io.

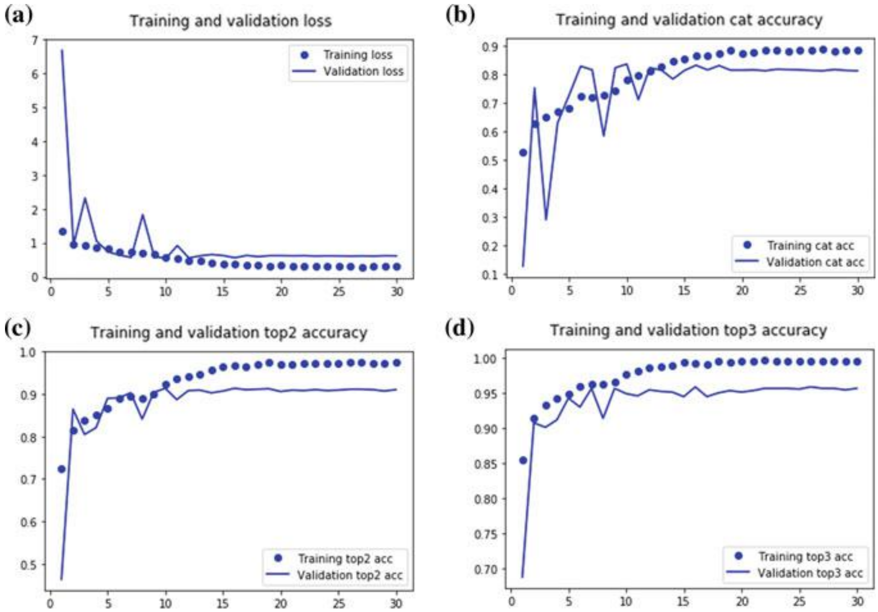


Fig. 4 Skin cancer classification performance curves of MobileNet model **a** training and validation loss **b** training and validation categorical accuracy **c** training and validation top2 accuracy **d** training and validation top3 accuracy

4 Conclusion

The skin cancer incidences are intensifying over the past decades; the need of an hour is to move toward an efficient and robust automated skin cancer classification system, which can provide highly accurate and speedy predictions. In this study, we demonstrated the effectiveness of deep learning in automated dermoscopic multi-class skin cancer classification with the MobileNet model trained on a total of 38,569 dermoscopy images from HAM10000 dataset. We matched the performance of expert dermatologists across seven diagnostic tasks with an overall accuracy of 83.1% for seven classes in the dataset, whereas top2 and top3 accuracy of 91.36% and 95.34%, respectively. Also, the weighted average of precision, the weighted average of recall, and the weighted average of f1-score were found to be 89%, 83%, and 83%, respectively. We conclude that MobileNet model can be used to develop an efficient real-time computer-aided system for automated medical diagnosis systems. As compared to previously proposed models, the MobileNet model has shown accurate and robust performance in addition to its faster and lightweight architecture.

The future work may deal with the utilization of patient's personalized data such as genes, age, and color in addition to the current study for skin cancer diagnosis. This additional feature can be advantageous to develop personalized computer-aided systems for the diagnosis of skin cancers.

References

1. Stewart, B.W., Wild, C.: International Agency for Research on Cancer, and World Health Organization. World cancer report (2014)
2. Cakir, B.O., Adamson, P., Cingi, C.: Epidemiology and economic burden of nonmelanoma skin cancer. Facial Plast. Surg. Clin. North Am. **20**(4), 419–422 (2012). https://doi.org/10.1016/j.fsc.2012.07.004
3. Rogers, H.W., Weinstock, M.A., Feldman, S.R., Coldiron, B.M.: Incidence estimate of nonmelanoma skin cancer (Keratinocyte carcinomas) in the U.S. population, 2012. JAMA Dermatol. **151**(10), 1081–1086 (2015). https://doi.org/10.1001/jamadermatol.2015.1187
4. Stern, R.S.: Prevalence of a history of skin cancer in 2007: results of an incidence-based model. Arch. Dermatol. **146**(3), 279–282 (2010). https://doi.org/10.1001/archdermatol.2010.4
5. WHO: Skin cancers WHO (2017)
6. Koh, H.K., Geller, A.C., Miller, D.R., Grossbart, T.A., Lew, R.A.: Prevention and early detection strategies for melanoma and skin cancer current status. Arch. Dermatol. **132**(4), 436–443 (1996). https://doi.org/10.1001/archderm.1996.03890280098014
7. Parkin, D.M., Mesher, D., Sasieni, P.: Cancers attributable to solar (ultraviolet) radiation exposure in the UK in 2010. Br. J. Cancer **105**(2), S66–S69 (2011). https://doi.org/10.1038/bjc.2011.486
8. Canadian Cancer Society. Risk factors for melanoma skin cancer (2018). https://www.cancer.org/cancer/melanoma-skin-cancer/causes-risks-prevention/risk-factors.html. Accessed 31 Mar 2019
9. Cancer facts & figures 2016. Atlanta, American Cancer Society (2016). https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2016.html. Accessed 31 Mar 2019

10. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2017. CA Cancer J. Clin. **67**(1), 7–30 (2017). https://doi.org/10.3322/caac.21387
11. Neville, J.A., Welch, E., Leffell, D.J.: Management of nonmelanoma skin cancer in 2007. Nat. Clin. Pract. Oncol. **4**(8), 462–469 (2007). https://doi.org/10.1038/ncponc0883
12. Morton, C.A., Mackie, R.M.: Clinical accuracy of the diagnosis of cutaneous malignant melanoma. Br. J. Dermatol. **138**(2), 283–287 (1998). https://doi.org/10.1046/j.1365-2133.1998.02075.x
13. Binder, M., Schwarz, M., Winkler, A., et al.: Epiluminescence microscopy. A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. Arch. Dermatol. **131**(3), 286–291 (1995). https://doi.org/10.1001/archderm.1995.01690150050011
14. Piccolo, D., Ferrari, A., Peris, K., Daidone, R., Ruggeri, B., Chimenti, S.: Dermoscopic diagnosis by a trained clinician vs. a clinician with minimal dermoscopy training vs. computer-aided diagnosis of 341 pigmented skin lesions: a comparative study. Br. J. Dermatol. **147**(3), 481–486 (2002). https://doi.org/10.1046/j.1365-2133.2002.04978.x
15. Argenziano, G., et al.: Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. J. Am. Acad. Dermatol. **48**(5), 679–693 (2003). https://doi.org/10.1067/mjd.2003.281
16. Kittler, H., Pehamberger, H., Wolff, K., Binder, M.: Diagnostic accuracy of dermoscopy. Lancet. Oncol. **3**(3), 159–165 (2002). https://doi.org/10.1016/S1470-2045(02)00679-4
17. Silver, D., et al.: Mastering the game of go with deep neural networks and tree search. Nature **529**, 484–489 (2016). https://doi.org/10.1038/nature16961
18. Mnih, V., et al.: Human-level control through deep reinforcement learning. Nature **518**, 529–533 (2015). https://doi.org/10.1038/nature14236
19. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. **115**(3), 211–252 (2015). https://doi.org/10.1007/s11263-015-0816-y
20. Vestergaard, M.E., Macaskill, P., Holt, P.E., Menzies, S.W.: Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. Br. J. Dermatol. **159**(3), 669–676 (2008). https://doi.org/10.1111/j.1365-2133.2008.08713.x
21. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017). https://doi.org/10.1038/nature21056
22. Rogers, H.W., Weinstock, M.A., Feldman, S.R., Coldiron, B.M.: Incidence estimate of non-melanoma skin cancer (Keratinocyte carcinomas) in the US population. JAMA Dermatol. **151**(10), 1081 (2015). https://doi.org/10.1001/jamadermatol.2015.1187
23. Stern, R.S.: Prevalence of a history of skin cancer in 2007. Arch. Dermatol. **146**(3), 279–282 (2010). https://doi.org/10.1001/archdermatol.2010.4
24. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015). https://doi.org/10.1038/nature1539
25. Arbib, M.A.: The Handbook of Brain Theory and Neural Networks. MIT Press (1998)
26. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (2017). https://doi.org/10.1145/3065386
27. Liu, L., Yan, R.J., Maruvanchery, V., Kayacan, E., Chen, I.M., Tiong, L.K.: Transfer learning on convolutional activation feature as applied to a building quality assessment robot. Int. J. Adv. Robot. Syst. **14**(3), 172988141771262 (2017). https://doi.org/10.1177/1729881417712620
28. Szegedy, C., et al.: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9 (2015). https://doi.org/10.1109/CVPR.2015.7298594
29. Ding, P., Zhang, Y., Deng, W.J., Jia, P., Kuijper, A.: A light and faster regional convolutional neural network for object detection in optical remote sensing images. ISPRS J. Photogramm. Remote Sens. **141**, 208–218 (2018). https://doi.org/10.1016/j.isprsjprs.2018.05.005
30. Masood, A., Al-Jumaily, A.A.: Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms. Int. J. Biomed. Imaging 323268 (2013). http://dx.doi.org/10.1155/2013/323268

31. Khosla, A., et al.: ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. **115**(3), 211–252 (2015). https://doi.org/10.1007/s11263-015-0816-y
32. Rosado, B., et al.: Accuracy of computer diagnosis of melanoma: a quantitative meta-analysis. Arch. Dermatol. **139**(3), 361–367 (2003). https://doi.org/10.1001/archderm.139.3.361
33. Burroni, M., et al.: Melanoma computer-aided diagnosis: reliability and feasibility study. Clin. Cancer Res. **10**(6), 1881–1886 (2004). https://doi.org/10.1158/1078-0432.CCR-03-0039
34. Kawahara, J., Hamarneh, G.: Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers. In: Wang, L., Adeli, E., Wang, Q., Shi, Y., Suk, H.I. (eds.) Machine Learning in Medical Imaging. MLMI 2016. Lecture Notes in Computer Science, p. 10019 (2016). https://doi.org/10.1007/978-3-319-47157-0_20
35. Milton, M.A.A.: Automated Skin Lesion Classification Using Ensemble of Deep Neural Networks in ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection Challenge (2019)
36. Hardie, R.C., Ali, R., De Silva, M.S., Kebede, T.M.: Skin Lesion Segmentation and Classification for ISIC 2018 Using Traditional Classifiers with Hand-Crafted Features. https://arxiv.org/abs/1807.07001
37. Howard, A.G., et al.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (2017). https://arxiv.org/abs/1704.04861
38. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-sources dermatoscopic images of common pigmented skin lesions. Sci. Data **5**, 180161 (2018). https://doi.org/10.1038/sdata.2018.161
39. Weiss, K., Khoshgoftaar, T.M., Wang, D.: A survey of transfer learning. J. Big Data **3**(1), 1345–1359 (2016). https://doi.org/10.1109/TKDE.2009.191
40. Image Preprocessing—Keras Documentation. Keras (2019). https://keras.io/preprocessing/image/. Accessed 31 Mar 2019
41. Pandas: Working with missing data—pandas 0.22.0 documentation (2019). https://pandas.pydata.org/pandas-docs/stable/user_guide/missing_data.html. Accessed 31 Mar 2019
42. Mikolajczyk, A., Grochowski, M.: Data augmentation for improving deep learning in image classification problem. In: International Interdisciplinary Ph.D. Workshop (IIPhDW), pp. 117–122 (2018). https://doi.org/10.1109/IIPHDW.2018.8388338
43. Kaggle: Your Home for Data Science. https://www.kaggle.com/. Accessed 31 Mar 2019

Real-Time Object Detection in Remote Sensing Images Using Deep Learning

Vijender Busi Reddy, K. Pramod Kumar, S. Venkataraman, and V. Raghu Venkataraman

Abstract Deep learning technology has grown vastly in the present era and penetrating to all research fields. Deep learning mechanism for objection detection is gaining prominence in remote sensing data analysis. The main challenge of object detection in remote sensing data is that the objects appear small and different objects look similar in the images. Another challenge in remote sensing data especially for real-time detection is the volume of the data handling which is substantially large. This paper presents a method based on Faster-RCNN network with required modifications to suit real-time object detection in remote sensing images.

Keywords Deep learning · Satellite data · Remote sensing · Neural networks · IRS satellite

1 Introduction

Deep learning (DL) technique is an efficient machine learning methodology to solve many complex tasks in image processing, signal processing, and natural language processing. Scene classification and object detection are key elements in remote sensing data analysis. Significant research has been done on scene classification in remote sensing data. Convolution neural networks (CNNs) [7] are popular and

V. B. Reddy (✉) · K. Pramod Kumar · S. Venkataraman · V. Raghu Venkataraman
Department of Space, Advanced Data Processing Research Institute (ADRIN), Government of India, Secunderabad, India
e-mail: vijender@adrin.res.in

K. Pramod Kumar
e-mail: pramod@adrin.res.in

S. Venkataraman
e-mail: kalka@adrin.res.in

V. Raghu Venkataraman
e-mail: director@adrin.res.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_16

widely used for image classification. They use pixel patterns from the image for classification.

Nowadays, image classification methods in remote sensing have shown good results with reasonable accuracy [14]. Supervised learning is a common method used in image classification, where first images are subjected to training with labeled scenes and followed by inferencing an unknown scene. Detection of objects is a key component in satellite imagery analytics. Therefore, a good object detection method is very critical for remote sensing image analytics.

After 2013, there was a wide impetus on object detection which was due to Imagenet ILSVRC challenge [10]. In 2013, Imagenet ILSVRC challenge average object detection rate was only 22.581%. In ILSVRC2014, the average object detection rate is greatly improved to 43.933%. Subsequently, the detection rate is increased every year. Currently in 2017, the average object detection rate increased to 73.13%. Above works included RCNN [17], Overfeat [15], GoogLeNet [8], DeepID-Net [5], etc., models.

CNNs perform the operation similar to the human eye. Eye's visual cortex contains cells that are arranged in complex manner [9]. These cells are very sensitive to small subregions in the visual field. The cells in the subregions act as local filters and extract the local correlations present in the image [9]. Similarly, CNNs use local filters to extract strong local correlation and replicate the weights across the entire visual field. The wights are obtained to form feature map.

Compared to image classification, object detection is a more challenging task and complex. Real-time object detection, target identification/recognition, tracking are very crucial better decision making. Several applications require to detect objects in real-time such as autonomous vehicles, object tracking, face detection and recognition, activity recognition, robotics, defect identification in manufacturing industry, and disaster management.

In object detection, first the data is classified and then localization of object is done. Searching for the object in an entire image consumes more time. Hence, most of the models select the probable locations and then search for objects in the selected locations. Region-based convolutional neural networks (RCNN) [17] first select the regions then searches for objects in the selected region. Traditional RCNN [17] and Faster-RCNN [17] are found to be highly time-consuming and also has lesser accuracy when used in large volume of remote sensing data.

Following adaptations to Faster-RCNN [17] are made to overcome the issues of time and accuracy as for as RS images are concern.

– We propose a tile-based method for training and inference.
– Preprocessing of data is done to improve accuracy.
– Incorporated a new "Misc" label for objects that are similar to objects of interest.

The rest of the paper is organized as follows: Sect. 2 briefly explains about previous research works in deep learning on RS images. Section 3 shows the proposed method for object detection in real time. Section 4 shows the inference results and analysis. Finally, Sect. 5 concludes the paper.

2 Related Work

This section briefly describes some of the object detection mechanisms in deep learning.

Over the past several years, more and more studies have been using DL for processing of remote sensing (RS) imagery [20]. DL proved to be efficient for processing in both optical (hyper-spectral and multispectral imagery) and radar images, in extracting different land cover types such as road extraction, building extraction [3, 6, 14]. Deep learning architectures such as convolutional neural networks (CNNs), deep auto-encoders, deep belief networks, and recurrent neural networks [7] with long short-term memory (LSTM) model have already been explored for RS tasks [3, 4, 6, 12, 13]. It should be noted that most studies with DL for RS utilize the same area/location for the classification purposes, e.g., land cover or object detection [3]. However, multi-temporal images are usually required to reliably identify specific objects [3].

Many outstanding object detection algorithms use region proposals generation algorithms [21]. The goal of region proposal generation is to eliminate substantial irrelative areas which contain no objects; in a short time, it contributes to design more sophisticated classifiers without considering too many negative samples. A good region proposals generation algorithm should be able to suit for multi-scale objects and have a high recall performance and faster convergence.

Zhao et al. [21] adopt multi-scale sliding windows to generate region proposals. Qi et al. [16] provide a saliency-based method to generate object detection region proposals for deep learning and machine learning. The saliency method is computed in phase spectrum of Fourier transformation which can reduce the complexity significantly. Chen et al. [2] adopt sliding window technology on gradient images which can effectively reduce the number of region proposals. Cheng et al. [22] use the selective search method to generate region proposals and achieve good performance in variant objects. These methods have been proved to have good performance in some specific scenarios.

Ren et al. [17] improved the region-based technique by using CNN (Faster-RCNN). It is tested on MS COCO, PASCAL VOC 2007, and 2012 data set and shown good accuracy.

3 Object Detection in RS Images

Multiple challenges, viz. volume of data (a typical very high-resolution scene of C2S [1] with 16,000 × 60,000 resolution is 1.5 GB), objects of interest such as buildings, airplanes, and vehicles are smaller in size with respect to scene size. These challenges lead to higher false alarms. To substantiate the argument, the following scenarios are

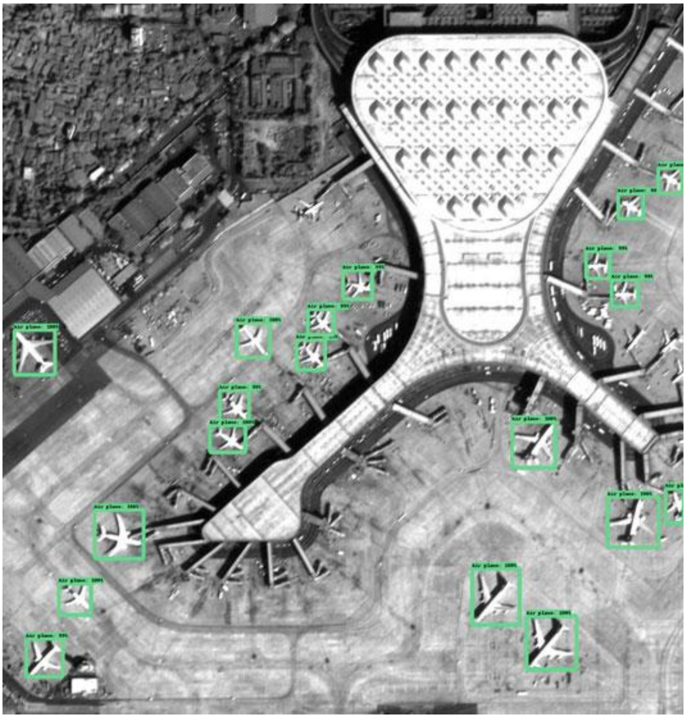


Fig. 1 Airplane detection with multiple size images

detailed below. Faster-RCNN model is trained using airplanes in various images similar to shown in Fig. 1. This trained model is used for inference in the following two scenarios.

Scenario 1 This scenario explains the issues with look-alike objects. When Fig. 2 is subjected to inference, two objects (*A* and *B*) are detected. Here, object *A* is identified correctly. However, object B is a false alarm (building) looks similar to object in consideration, i.e., airplane.

Scenario 2 The other issue with the object detection is image size. Images with different sizes are taken for training. When Figs. 1 and 3 are given for inference to airplane object detection model. However, trained model does not give intended accuracy on combined image of Figs. 1 and 3 which is shown in Fig. 4. The main reason for this is found to be difference in image sizes used for training and inference. This process is believed to suppress features of interest during the search reduction or squash operation. Hence, the model is unable to detect the objects accurately.

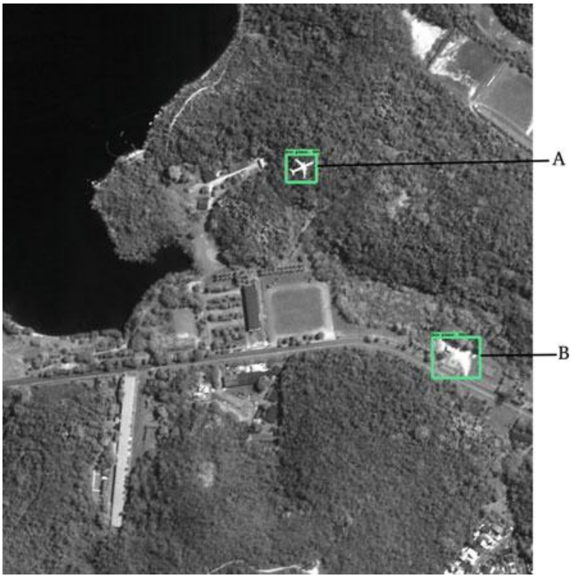


Fig. 2 Issues with airplane like similar object

Fig. 3 Airplane detection with multiple size images

3.1 Real-Time Object Detection in RS Images

The main aim of this paper is to provide an effective method to detect objects in remote sensing images. The proposed method works for bigger images and infer the images without squashing the input image.

Data set preparation: Data set preparation is one of the crucial tasks in deep learning because accuracy of the model depends on the data set given for training. Since images are quite large in size, it is not possible to subject the complete image for training due to current memory and processing constraints. We propose tile-based method to prepare the data for training the model. Figure 5a shows the data set preparation

Fig. 4 Airplane detection with combination of Figs. 2 and 3

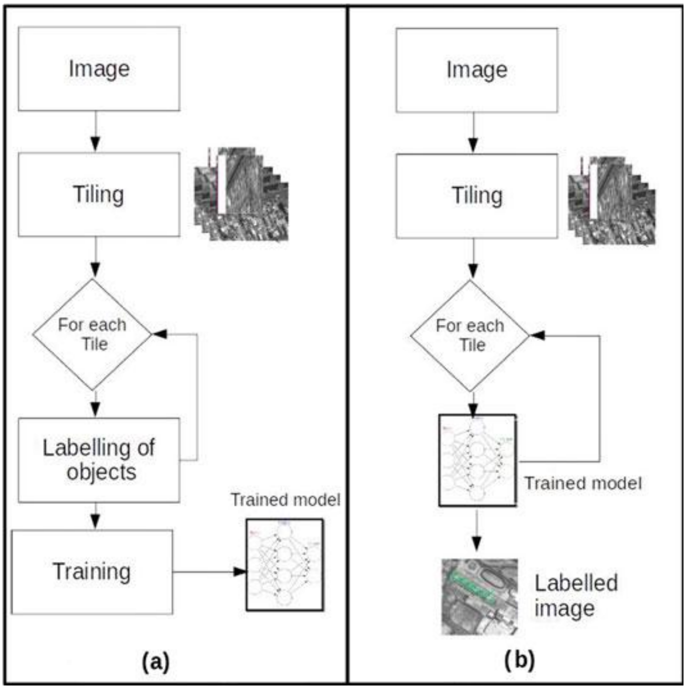


Fig. 5 **a** Data set preparation and training operations **b** inference

operations. RS images which have objects of interest are taken for preparing the data set. Each image is tiled and labeled by identifying the feature in each of the tile image. The labeled tiles are used to train the object detection model to detect objects of interest such as buildings, airplanes, and vehicles.

Training: Training of Faster-RCNN model [17] is done by using the prepared data and labels set. Faster-RCNN model contains convolution and max-pool layers. Convolution layer is configured with stride 16, and max-pool layer is configured with stride 2. Gradient decent backpropagation method with a learning rate 0.0002 is used for training.

Inference: In real time, the satellite data is processed by noise filters. Figure 5b depicts the inference method in object detection. The processed image is tiled and subjected to inference. After inferring all the tiles, the result is written back on the original untiled image.

4 Results and Analysis

The proposed method is applied on C2S [1] image data. Tiling is done for the size of 1024 × 1024 with an overlap of 256 pixels and labeled the objects in the tile by using LabelImg tool [11]. Nvidia K40 GPU is used for training and inferencing. Tile size is chosen based on the memory available on the GPU. Faster-RCNN model [17] is used in out methodology. 200 images with 2,00,000 iterations are performed to train the network.

The input image is divided into tiles and given for inference as per the method shown in Fig. 5b. The output of the inference is written on original untiled image with bounding boxes and confidence percentage.

Initially, the training is done on Faster-RCNN model with the data set prepared based on Fig. 5a. The inference is done on test images with the trained model. The average of precision, recall, and accuracy parameters on ten test images is 0.74, 0.76, and 0.71, respectively. Even though precision is greater than 0.7, real-time object detection should be more precise and required to be equal to 1.

So, a preprocessing method is applied before subject into training and inference. Each image is processed using image processing filters such as noise filters and contrast enhancement before subjecting into training. Faster-RCNN model is trained by using the new preprocessed data set. During the inference operation, the image is preprocessed in the same way as in training and then given for inference to trained model. The average of precision, recall, and accuracy parameters on ten test images is 0.76, 0.43, and 0.39, respectively. Even though precision is increased, the recall is decreased significantly because of false alarms. During the preprocessing operations, some objects become brighter and become look a like objects. Hence, these are detected as airplane objects. It is essential to reduce the false alarms in object detection.

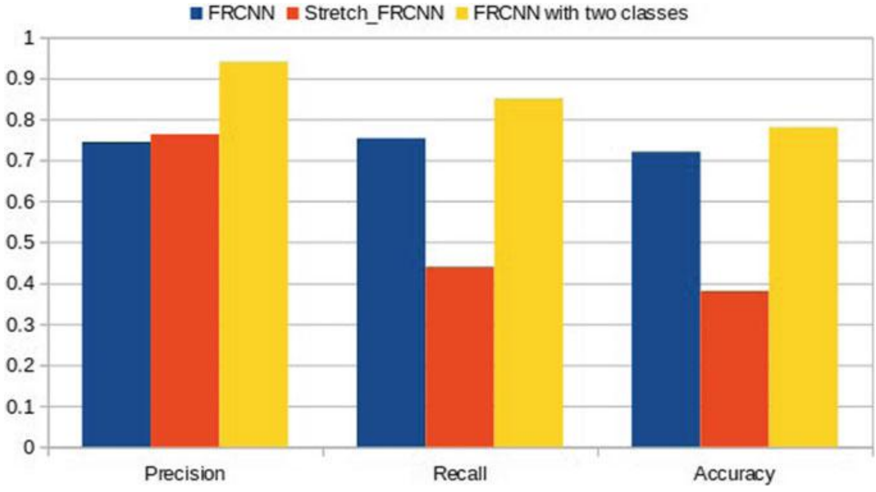


Fig. 6 Performance results of FRCNN, modified_FRCNN, FRCNN_2classes

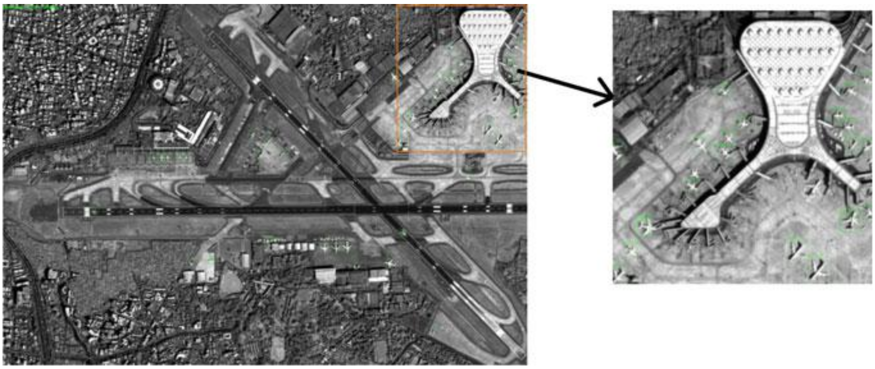


Fig. 7 Inference with FRCNN_2classes model on Fig. 4

In the third approach, we have introduced one more object "Misc" in the data set to reduce the false alarms. We have identified all the false alarms and labeled them "Misc" objects. Then the training is performed with two objects. Figure 7 shows the result of the inference on one test image. The average of precision, recall, and accuracy parameters on ten test images is 0.93, 0.85, and 0.79, respectively. It is clearly shown that the proposed method with multiple objects has performed better (Fig. 6).

Figure 6 shows the performance parameters (precision, accuracy, and recall) [19] of the above three trained models. These values are taken as average of ten images' inference. It clearly shows that the trained model with multiple objects accurately identifying the objects.

To test the robustness and reliability of the proposed model on other RS images and targets, ship detection on Resourcesat-2A [18] satellite's Advanced Wide Field

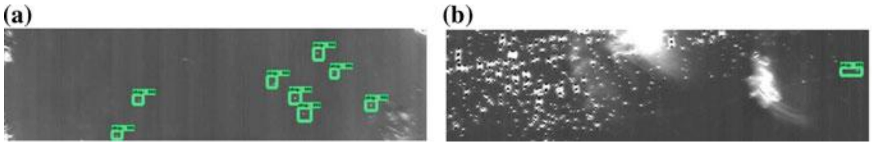


Fig. 8 Ship detection in Resourcesat-2A images

Sensor (AWiFS) data was carried out. AWiFS data has 56 m spatial resolution, and ships are very small with tale feature. AWiFS data is preprocessed and used for data set preparation as per the method given in Fig. 5a. Ships with wakes are labeled and trained with our proposed methodology. Faster-RCNN [17] model is trained using data sets and ship labels. Trained model is used to infer the AWiFS data as per the method given in Fig. 5b. Figure 8a, b shows the inference results. Figure 8b shows good accuracy even though many similar objects are present in the image.

5 Conclusion and Future Work

In this paper, we have proposed a method to detect objects in large remote sensing images. The proposed method is tested on C2S data and Resourcesat-2A satellite's AWiFS data. Experiments are done on multiple images to show the robustness and reliability. The proposed method not only worked well for large remote sensing images but also precision is more than 90%. In the future, we will train the model to detect multiple object sizes.

References

1. Cartosat 2 series satellite. https://www.isro.gov.in/Spacecraft/cartosat-2-series-satellite-2
2. Chen, X., Xiang, S., Liu, C., Pan, C.H.: Vehicle detection in satellite images by hybrid deep convolutional neural networks. IEEE Geosci. Remote Sens. Lett. **11**(10), 1797–1801 (2014)
3. Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y.: Deep learning-based classification of hyperspectral data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **7**(6), 2094–2107 (2014)
4. Chen, Y., Zhao, X., Jia, X.: Spectralspatial classification of hyperspectral data based on deep belief network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **8**(6), 2381–2392 (2015)
5. DeepID-Net. https://arxiv.org/abs/1412.5661
6. Geng, J., Fan, J., Wang, H., Ma, X., Li, B., Chen, F.: High-resolution SAR image classification via deep convolutional autoencoders. IEEE Trans. Geosci. Remote Sens. **12**(11), 2351–2355 (2015)
7. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2017)
8. GoogleNet. https://www.ai.google/research/pubs/pub43022
9. Hubel, D., Wiesel, T.: Receptive fields and functional architecture of monkey striate cortex. J. Physiol. **195**, 215–243 (1968)
10. Imagenet LSVRC competition. www.image-net.org/challeges/LSVRC/
11. Label image tool. https://github.com/tzutalin/labelImg

12. Liang, H., Li, Q.: Hyperspectral imagery classification using sparse representations of convolutional neural network features. Remote Sens. **8**(2), 99 (2016)
13. Lyu, H., Lu, H., Mou, L.: Learning a transferable change rule from a recurrent neural network for land cover change detection. Remote Sens. **8**(6), 506 (2016)
14. Mnih, V., Hinton, G.E.: Learning to detect roads in highresolution aerial images. In: Proceedings of the Europe Conference on Computer Vision, pp. 210–223 (2010)
15. Overfeat network. https://arxiv.org/abs/1312.6229
16. Qi, S., Ma, J., Lin, J., Li, Y., Tian, J.: Unsupervised ship detection based on saliency and S-HOG descriptor from optical satellite images. IEEE Geosci. Remote Sens. Lett. **12**(7), 1451–1455 (2015)
17. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. arXiv:1506.01497v3 [cs.CV] (2016)
18. ResourceSat-2 and ResourceSat-2A satellites of ISRO. https://directory.eoportal.org/web/eoportal/satellite-missions/r/resourcesat2
19. Wu, H., Zhang, H., Zhang, J., Xu, F.: Fast aircarft detection in satellite images based on convolutional neural networks. In: 2015 IEEE International Conference on Image Processing (ICIP), Quebac City, QC, pp. 4210–4214 (2015)
20. Zhang, F., Du, B., Zhang, L.: Scene classification via a gradient boosting random convolutional network framework. IEEE Trans. Geosci. Remote Sens. **54**(3), 1793–1802 (2016)
21. Zhao, A., Fu, K., Sun, H., Sun, X., Li, F., Zhang, D., Wang, H.: An effective method based on ACF for aircraft detection in remote sensing images. IEEE Geosci. Remote Sens. Lett. **14**(5), 744–748 (2017)
22. Zhou, P., Cheng, G., Liu, Z., Bu, S., Hu, X.: Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping. Multidimens. Syst. Signal Process. **27**(4), 925–944 (2016)

Malaria Detection Using Convolutional Neural Network

Pankaj Pundir, Shivam Aggarwal, and Maroti Deshmukh

Abstract In the tropical and subtropical countries, malaria has been a challenge, which really needs a quick and precise diagnosis to stop or control the disease. It is a serious and sometimes fatal disease caused by a parasite that commonly infects a certain type of mosquito which feeds on humans. The traditional microscopy technique has a few weaknesses which incorporate time utilization and reproducibility of the results. This paper deals with the automatic identification of malaria-infected cells using deep learning methods. Deep learning methods have the advantage of being able to automatically learn the features from the input data, thereby requiring minimal inputs from human experts for automated malaria diagnosis. An automated diagnostic system can significantly improve the efficiency of the pathologists and also reduce the need for dedicated pathologists in rural villages.

Keywords Malaria · Binary classification · Neural network · Convolutions · Image preprocessing

1 Introduction

Malaria is a life-threatening disease caused by parasites that are transmitted to people through the bites of infected female Anopheles mosquitoes. One-celled parasite—Plasmodium—is transmitted by a mosquito which is a vector or carrier. According to the report released by the World Health Organization (WHO), there were 214 million cases of malaria in 2015 and 438,000 deaths [1]. In most cases, malaria can only be diagnosed by a manual examination of the microscopic slide. Whole

P. Pundir · S. Aggarwal · M. Deshmukh (✉)
Department of Computer Science and Engineering, National Institute of Technology Uttarakhand, Srinagar, India
e-mail: marotideshmukh@nituk.ac.in

P. Pundir
e-mail: pankaj369.cse16@nituk.ac.in

S. Aggarwal
e-mail: ashivam093.cse16@nituk.ac.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_17

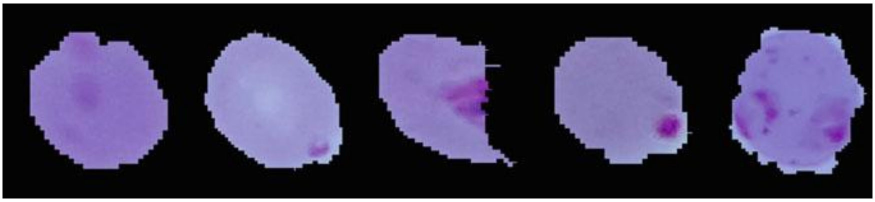


Fig. 1 Infected cells

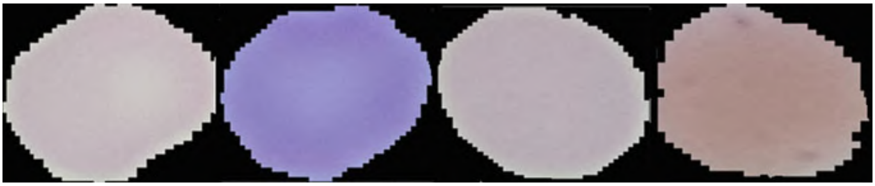


Fig. 2 Uninfected cells

slide imaging, which scans the conventional glass slides in order to produce digital slides, is the most recent imaging modality being employed by pathology departments worldwide. Figures 1 and 2 show human red blood cell samples obtained using whole slide images.

So as to give a solid finding, essential preparing and concentrated human assets are required. Tragically, these assets are a long way from being satisfactory and much of the time regularly inaccessible in immature zones where jungle fever has a stamped prevalence. Along these lines, an automated framework can give an effective answer to this issue. Due to this drawback, the primary objective of this study is to develop a computerized malaria detection framework that replicates the conventional gold standard diagnosis of malaria to identify malaria parasites in blood smear images. Such a diagnostic system can be of immense help to malaria specialists in the diagnostic process. An automated diagnostic system can significantly improve the efficiency of the pathologists and also reduce the need for dedicated pathologists in rural villages. In microscopy diagnosis, peripheral blood smear images are stained using a popular stain called Giemsa to perform efficient visual identification of malaria parasites. The staining procedure slightly colors the healthy RBCs but highlights parasites, platelets, artifacts, and white blood cells (WBCs) with a deep purple color as shown in Fig. 1. To identify malaria infection, it is necessary to divide the purple-colored blood cell components into two categories using visual cues: parasite-infected blood cells as shown in Fig. 1 and non-parasitic blood cell shown in Fig. 2 components such as platelets, artifacts, and WBCs. Therefore, this research addresses the area of a computer-aided malaria detection system using a deep learning architecture. Recently, deep learning techniques have gained increasing attention from researchers for its great capability in learning complex patterns within images; thus, it is used for building an automated diagnostic system for malaria

[2, 5]. In [2], machine learning algorithms like SVM [15] and naive Bayes classifiers [4] were utilized to achieve accuracies of 84% and 83.5%, respectively.

Most of the work done in this field is based on image processing methods and machine learning. Both of these methods require human effort in feature engineering and do not provide a reliable and stable solution [14] which includes an image processing method to detect malaria parasite in thin blood smear and classify them [15]. In this paper, a fast and precise system was developed using stained blood smear images. Watershed segmentation technique is used to acquire plasmodium infected and non-infected erythrocytes, and the relevant feature was extracted. Many different machine learning classifiers like SVM [16] and KNN are used. As features can be very complex, deep learning methods are used for automatic feature extraction and provide state-of-the-art results.

2 Methodology

The fundamental aim of this paper is to build an automatic system for the detection of malaria within the smeared blood sample image. The promising performance of CNNs is accompanied by the availability of a huge amount of annotated data. The steps for the proposed method contain input image, preprocessing, convolutional neural network, and classification. With scarcity for annotated medical imagery, transfer learning (TL) [10] methods are used where pre-trained deep learning models are either fine-tuned on the underlying data or used as feature extractors to aid in visual recognition tasks [11]. These models transfer their knowledge gained while learning the generic features from large-scale datasets like ImageNet [12] to the underlying task. The transfer of previously learned skills to a new situation is generalized, rather than unique to the situation. The existing pre-trained models are trained on a dataset which contains more complex images and shows less correlation with the malaria dataset, and thus, training from scratch is preferred.

2.1 Image Preprocessing

The aim of preprocessing is to remove unwanted objects and noise from the image to facilitate image segmentation into meaningful regions. We have performed data preprocessing before feeding it into a neural network. We have applied some basic image processing.

1. Grayscale conversion
2. Resize image to 50×50
3. Morphological operations
4. Image augmentation.

As thresholding will degrade the information within the image, thus morphology is preferred. As an image is of a single red blood cell, it lacks complex features. But, the infected cells may show less correlation with another cell of the same category and require recognizing minute features.

2.2 Convolutional Model

The neural net is implemented using a deep learning framework—Keras [13]. Various layers including dropout, MaxPooling2D, and batch normalization are used to prevent model from overfitting. The model is compiled with Adam optimizer with learning rate = 0.001, decay = 0.001/64, and batch size = 64.

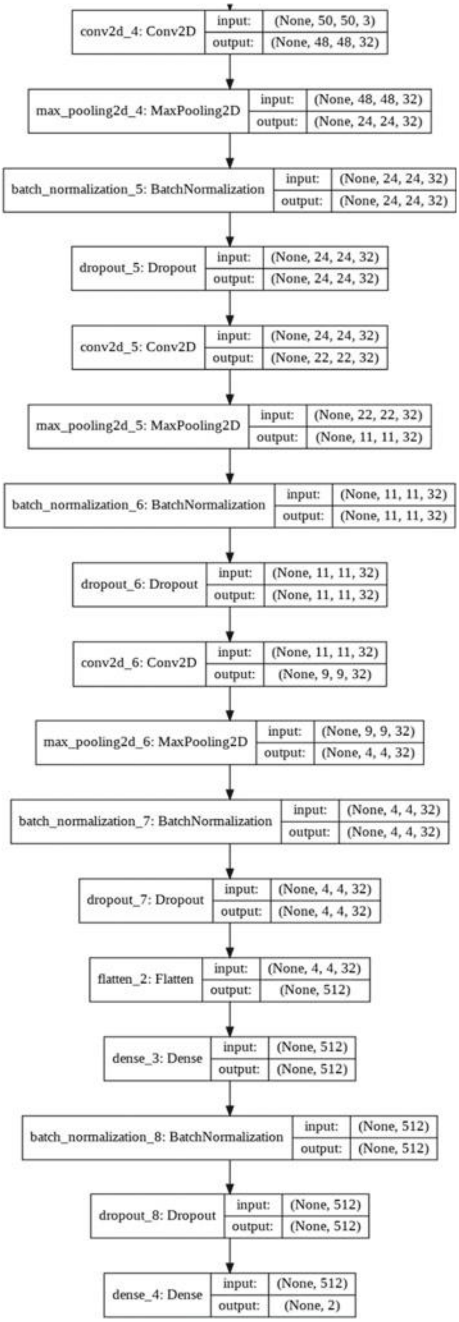
1. Conv2D (32, batch normalization, MaxPooling, dropout)
2. Conv2D (32, batch normalization, MaxPooling, dropout)
3. Conv2D (32, batch normalization, MaxPooling, dropout)
4. Flatten
5. Dense (512, batch normalization, dropout)
6. Dense (512)
7. Dense (2).

Each Conv2D layer is accompanied with MaxPooling2D, batch normalization, and dropout. Initial layers extract prominent features, and higher layers extract more complex features. Model is compiled with Adam optimizer, and the evaluation metric is set to 'accuracy.' A detailed architecture is provided in Fig. 3.

3 Experimental Results

The dataset is a standard dataset taken from the official NIH Web site [7]. Researchers at the Lister Hill National Center for Biomedical Communications (LHNCBC), part of the National Library of Medicine (NLM) [6], have developed a mobile application that runs on a standard Android smartphone attached to a conventional light microscope [8]. Giemsa-stained thin blood smear slides from 150 P. falciparum-infected and 50 healthy patients were collected and photographed at Chittagong Medical College Hospital, Bangladesh. The smartphone's built-in camera acquired images of slides for each microscopic field of view. The images were manually annotated by an expert slide reader at the Mahidol Oxford Tropical Medicine Research Unit in Bangkok, Thailand. The de-identified images and annotations are archived at NLM. A level-set-based algorithm to detect and segment the red blood cells has been tested [9]. Each image consists of a single red blood cell. Each cell is categorized under 'infected' and 'uninfected.' Infected cells show a blobby structure when passes through stain to show different colors. Data comprises equal numbers of infected and non-infected cells. There are total of 27,560 images out of which

Fig. 3 Model architecture



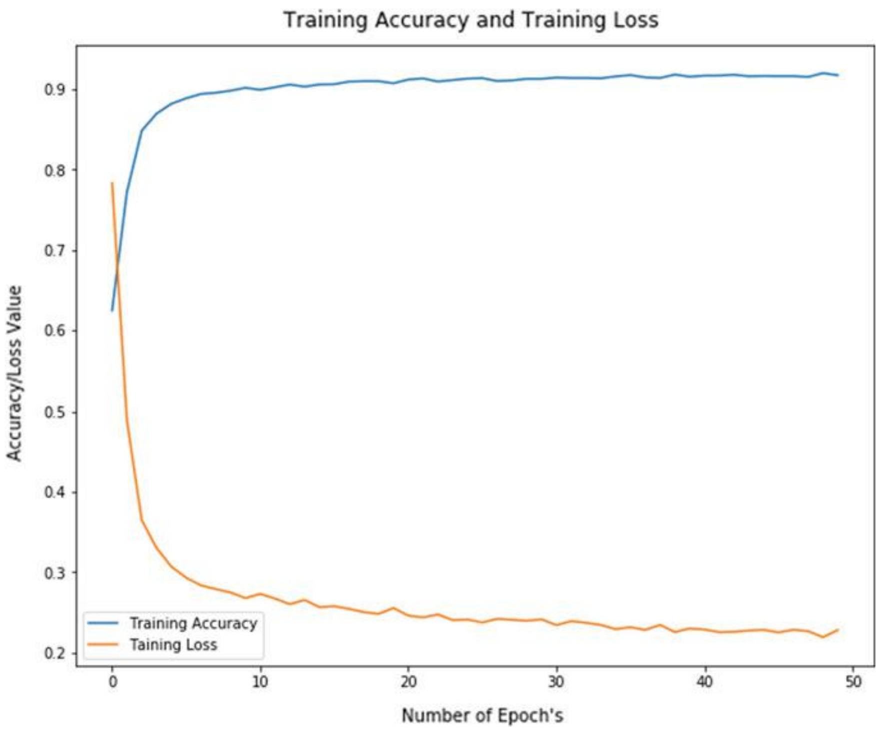


Fig. 4 Training accuracy and loss curve

Table 1 Accuracy and loss of model

	Loss	Accuracy
Training	0.147	95.3
Test	0.127	96.02

13,780 are of infected cells and other 13,780 are of non-infected cells. Further, we have split this dataset into training and test set. Validation split of 10% is performed during training in Keras [13]. The training accuracies and loss curves are shown in Fig. 4, and corresponding loss and accuracies are provided in Table 1.

3.1 Platform Used

The model is accelerated with Nvidia Titan X GPUs that were provided online in Kaggle [17]. This benchmark shows that enabling a GPU to your Kernel results in a 12.5× speedup during training of a deep learning model. The model is trained for 50 epochs and learned minute features from the training set.

Table 2 Confusion matrix

Predicted/Actual	Yes	No
Yes	TP = 2630	FP = 120
No	FN = 99	TN = 2663

3.2 Metric

A graph is plotted between number of epochs vs accuracy loss. We generated a confusion metric corresponding to our results as shown in Table 2, by which we can calculate some important parameters like precision, recall. Basic terminology used in the confusion metric can be defined as follows:

1. True positives (TP): These are cases in which the model predicted yes, and actual result was also yes.
2. True negatives (TN): The model predicted no, and actual result was also no.
3. False positives (FP): The model predicted yes, but actual value was no.
4. False negatives (FN): The model predicted no, but actual was yes.

Some important metrics calculated from the confusion metric are defined as follows:

1. **Accuracy**: Classification rate or accuracy is given by relation Eq. (1)

However, there are problems with accuracy. It assumes equal costs for both kinds of errors.

2. **Precision**: Precision or confidence denotes the proportion of predicted positive cases that are correctly real positives [16]. It is also known as true positive rate (TPR). Precision is defined by Eq. (2)

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

3. **Recall or sensitivity**: Recall or sensitivity is the proportion of real positive cases that are correctly predicted positive. This measures the coverage of the real positive cases [16]. Recall is calculated by Eq. (3).

$$\text{Recall} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{3}$$

4. **F-measure**: Since we have two measures (precision and recall), it helps to have a measurement that represents both of them. We calculate an F-measure which uses harmonic mean in place of arithmetic mean as it punishes the extreme values more. The F-measure will always be nearer to the smaller value of precision or recall. calculated by Eq. (4).

Table 3 Results

Metric	Value (%)
Accuracy	96.02
Precision	95.63
Recall	95.68
F-measure	95.65

$$\text{F-measure} = \frac{2PR}{P + R} \tag{4}$$

As the project deals with medical application where the life of a person has utmost importance, precision cannot be taken as a crucial factor as with larger and biased data precision values biases, and thus, recall is a better measure than precision. This measure ensures the cases detected as positives are the real ones. F-measure is also adopted to provide an overview of model's accuracy as it considers both precision and recall. Values of these parameters for our model are shown in Table 3.

4 Comparison

A research [18] made the classification of malaria cells using deep learning approach. [18] used pre-trained VGG that facilitates the role of expert learning model and SVM as domain-specific learning model. Initial 'k' layers of pre-trained VGG are retained, and (n-k) layers are replaced with SVM. VGG19-SVM resulting in the classification accuracy of 93.1% and in the identification of infected falciparum malaria. Our proposed method follows a convolution network trained from scratch with image augmentation, and our results exceed [18] with 20% testing split. The obtained result shows the potential of deep learning in the field of medical image analysis, especially malaria diagnosis.

5 Conclusion

This paper presents a deep learning technique to classify human red blood cells that are infected with malaria. The proposed malaria detection system is based on morphological processing for the detection of the state. Features like irregular shape and size are extracted within the initial convolution layers. This system diagnoses the RBCs in the image as infected or non-infected cells. Compared to manual results, system results showed excellent efficiency of about **96.02**% accuracy for the detection of malaria within a single blood cell image. Further, the blood cells should be stained to show malaria cell properties.

References

1. World Health Organization: Disease burden of malaria. http://www.who.int/mediacentre/factsheets/fs094/en/
2. Das, D.K., et al.: Machine learning approach for automated screening of malaria parasite using light microscopic images. J. Micron **45**, 97–106 (2013)
3. Shawe-Taylor, J., Cristianini, N.: Support vector machines. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, pp. 93–112 (2000)
4. Rish, I.: An empirical study of the naive Bayes classifier. In: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, vol. 3, no. 22 (2001)
5. Tek, F.B., Dempster, A.G., Kale, I.: Parasite detection and identification for automated thin blood film malaria diagnosis. J. Comput. Vis. Image Underst. **114**(1), 21–32 (2010)
6. Malaria Detection Database: https://peerj.com/articles/4568/
7. Dataset Link: https://ceb.nlm.nih.gov/proj/malaria/cellimages.zip
8. Poostchi, M., Silamut, K., Maude, R.J., Jaeger, S., Thoma, G.: Image analysis and machine learning for detecting malaria. Transl. Res. **194**, 36–55 (2018)
9. Kose, M., Incel, O.D., Ersoy, C.: Online human activity recognition on smart phones. In: Workshop on Mobile Sensing: From Smartphones and Wearables to Big Data, vol. 16, no. 2012, pp. 11–15 (2012)
10. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)
11. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 806–813 (2014)
12. Deng, J., et al.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE (2009)
13. Gulli, A., Pal, S.: Deep Learning with Keras. Packt Publishing Ltd. (2017). https://keras.io/
14. Mohammed, H.A., Abdelrahman, I.A.M.: Detection and classification of malaria in thin blood slide images. In: 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE), pp. 1–5. IEEE
15. Wang, L. (ed).: Support Vector Machines: Theory and Applications, vol. 177. Springer Science & Business Media (2005)
16. Powers, D.M.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation (2011)
17. Online Datascience Platform: www.kaggle.com
18. Vijayalakshmi, A.: Deep learning approach to detect malaria from microscopic images. Multimed. Tools Appl. 1–21 (2019)

Drone-Based Face Recognition Using Deep Learning

Adam Deeb, Kaushik Roy, and Kossi D. Edoh

Abstract The next phase of facial recognition research is continuing to finding ways of improving the accuracy rates of models when input images are taken from less than ideal angles and distances, from lower quality images, and from images that do not show much facial information of the person. In this paper, we attempted to use convolutional neural network (CNN) models to accomplish these tasks and attain an improved top accuracy. In this study, we compared three different deep learning models: VGG16, VGG19, and InceptionResNetV2; when testing them in several different facial recognition tasks, using the DroneFace dataset. We used three of the most accurate CNNs, when tested by using the ImageNet database, in an attempt to show that they can be used to achieve high drone face recognition accuracy. After applying these three CNNs to the image dataset used in the study, we compared the accuracy achieved by using each deep learning model in order to see which model was best able to handle and interpret images presented it, when the images provided are taken from a drone. Specifically, we tested how the heights at which the images were taken, by the drone, affected the accuracy of the model at detecting who the photographs were taken of. We attempted to achieve this by training the model at large heights and testing at low heights, by training the model at large heights and testing at low heights, and by training on a random set of 80% of photographs of all subjects and testing on the remaining 20% of photographs of all subjects.

A. Deeb (✉)
Department of Computer Science, North Carolina State University, Raleigh, NC 27695, USA
e-mail: adeeb@ncsu.edu

K. Roy
Department of Computer Science, North Carolina A&T University, Greensboro, NC 27401, USA
e-mail: kroy@ncat.edu

K. D. Edoh
Department of Mathematics, North Carolina A&T State University, Greensboro, NC 27401, USA
e-mail: kdedoh@ncat.edu

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_18

1 Introduction

Facial recognition is a very broad area of study with many different applications; while there has been a lot of research done on deep learning and facial recognition, there is still much more to learn. Facial recognition is an application of deep learning, which is a small subset of machine learning, also a subset of artificial intelligence. Recently, deep learning has been used extensively in computer vision research, and today, we find ourselves with several usable approaches. AlexNet, VGGnet, ResNet, and Google Inception, developed in that order and with each being increasingly deep, have been the leading models in image recognition using the ImageNet dataset. Deep learning has been used extensively in computer vision research, and today, we find ourselves with several usable approaches. In the study, we used a type of neural network called convolutional neural networks (CNNs). CNNs are one of the most commonly used and most effective of these approaches, as it is usually composed of many trainable layers and can be used for a number of computer vision applications [1]. The three previously mentioned CNNs have achieved top marks on the ImageNet Database Competition, attaining some of the top accuracy, top-1 and top-5 accuracies, between 2012 and 2016. The ImageNet project contains around 14,197,122 images from 21,841 different categories, and the organization holds an annual competition where participants report their accuracies on the 1.2 million images, belonging to 1000 different classes, that are provided to them [2]. Nowadays, facial recognition has already advanced immensely since to its beginning, but very little research has been done on how this technology would perform if it were to be implemented in drones. Drones are autonomously piloted aircrafts which can easily reach places which would be unreachable for people and easily see things that would be very difficult for the human eye to see [3]. Facial recognition is really accurate when cameras capture the image of the subject from perfectly direct angle, zero degrees of elevation, and at a very close distance, but the images that would be captured by a drone would usually be from drastically different conditions. Based on the study of Hsu and Chen [4], two of the most popular, readily available, online facial recognition softwares, Face++ and Rekognition, were put to the test using the dataset Labeled Faces in the Wild by Huang, Ramesh, Berg, and Learned-Miller in order to test their abilities in achieving desired functionality when presented with these images. In [4], the study of Hsu and Chen defined an accuracy of 0.75, 75 percent, as the "standard of distinguishability." Face++ and Rekognition achieved approximately this standard from distances of 3–12 m and at a height of up to 4 m. The results of the study showed, as predicted, that the programs were able to recognize images when taken in those ideal conditions at extremely high rates of success, while the conditions of the photographs started to become less ideal, the success rate of recognition drastically fell. This is a bit startling as when drones with facial recognition functionality will be used in society, they will be required to be able to function in much higher heights, than even this dataset provides, and in poor weather conditions, and with less facial information provided to it. In this paper, we are using the three best performing ImageNet deep learning models, and we will try to highlight any deficiencies any model might show when combating these varying conditions.

1.1 Related Works

Mini-drones are increasingly used in video surveillance. Their aerial mobility and ability to carry video cameras provide new perspectives in visual surveillance which can impact privacy in ways that have not been considered in a typical surveillance scenario [4]. One of the more popular utilizations of facial recognition technology using drones is with military UAVs, unmanned aerial vehicles. More specifically, the United States Army has been known to use UAVs for detecting and tracking potentially dangerous targets [5].

Knowing the potential future payoff using drones for public security and safety, the University of Kent tested its limitations by doing four experiments attempting to highlight these limitations shown by drones and the facial recognition implementation accompanied by drones [6]. The four experiments conducted by the researchers of Kent where first was to allow human observers to attempt to match very low-quality distance shots from an array of images to a very high-quality front face photograph of an unfamiliar person. The second test turned the matching test to a recognition test by applying the same test only this time and the subject who will be pictured is now a familiar face. In the third test, it is similar to test by also using a familiar subject but now a still image is replaced by a 10 s video of the moving subject. The final test was to see whether observers could determine the sex, race, age of unfamiliar subjects from the drone-captured footage. The results of the study presented information about how difficult these real-world situations, where drones would be at steep angles and from far distances and heights, would be for a human to analyze and produce consistently accurate results [6].

In terms of security, the identification and tracking of an attacker, virtual or real world, are of the upmost importance. In [7], an idea of using facial biometrics to protect technology, like PCs and smartphones, was the best way to defend against attacks and to also help determine the attacker if he decides to threaten your device. Not only that, but the authors also bring an example of a Japanese protection firm, who has set up a drone which will take off once the intruder alarm of one of its customers is set off. The job of this drone is to immediately identify the intruder and chase him, capturing real-time images of any wrong-doing.

The only other available research data has been provided by the creator of the dataset. In [3], its authors, Hwai-Jung Hsu and Kuan-Ta Chen, tested the accuracy of two popular online facial recognition services, Face++ and Amazon Rekognition, when asked to correctly identify images taken at each combination of heights and distances.

2 Dataset

The images used in this study were obtained from the DroneFace dataset. With this being a recently established dataset, there is currently very little research concerning it. DroneFace was established to try and further advance drone capabilities and to resolve the current limitations of facial recognition. If one were to utilize the functionalities of a drone trying to locate an individual, it would be very problematic if one would have to have the drone placed perpendicularly and in close proximity to the person.

In the dataset, there are 11 subjects, 7 males and 4 females, 2,057 pictures composed of 620 raw images, 1,364 frontal face images, and 73 portrait images. Raw images are taken in 3,680 × 2,760 resolution with ultra-wide field of view (170°) under daylights, and the resolutions of the facial images are between 23 × 31 and 384 × 384. The raw images are taken from 1.5, 3, 4, and 5 m high, and from 2 to 17 m away from the subjects with 0.5 m interval [5].

3 Methods

With the current improvements, some began to conduct research using facial and image detection for applications other than the classification. Research has been done using models to detect causes of distracted driving by studying changes in facial features and to detect hidden messages in images that are used as a cover [2]. With this still being a relatively new area of study, there is still some debate and uncertainty over which methods provided the best balance of accurate results while also using the simplest code, and there is still a lot of growth and development.

Facial detection, in the context of using image and machine learning, is the ability to determine whether there are any faces in a given image and, if any are found, return where the face is located. Facial detection from a single image can be a challenge, because of the extreme variability of the features of anyone face. Changes in expression, pose, scale, and many other factors can have a great impact on the appearance of a person's face [8]. Facial recognition can be used to solve two major issues, face verification, 1:1 matching, and face identification, 1:N matching. In this study, we attempt to explore both approaches to facial recognition.

First, the DroneFace dataset had to be organized so that the data could be set up correctly for each assessment, to train. In the first experiment, the images used were all the faces of each subject, cropped faces from the provided distance shots, and they were organized into subfolders containing only the images of one subject in each subfolder and then, using the split-folders package of python; splits folders with files into train, validation, and test files; to split all the images into only train and validation folders, still separated by subject. So, the train and validation sets were made up of 11 folders, each containing the images of its corresponding subject taken from all heights and distances. The train and validation sets contained different images that were randomly split, 80% for training, and 20% for validation from the same set of photographs. Then, we added the model. In this paper, we used VGG16, VGG19, and InceptionResNetV2 and used the same optimizer, stochastic gradient descent with a learning rate of 0.005, and the same loss function, sparse categorical cross-entropy. Then, by using the ImageDataGenerator function, we were able to import the train and validation images into python and then we could move on to train the model by using the model.fit_generator function. We used the matplotlib package, used for plotting in Python, to create the accuracy and loss graphs; for an example of the accuracy graphs; see Figs. 1, 2, 3, 4 and 5.

Then, the images had to be structured in such a way so that we can see how accuracy changes when height is accounted for. The two ways to accomplish this was to test the models' ability to recognize the same person from a set of photographs of only themselves at different heights and distance, and also, to test the models' ability to recognize a certain subject when presented with a photograph from a set of photographs containing images of all subjects at different heights and distances. In the first scenario, the train set consisted of only images of the individual subject taken from certain heights, or distances, and the validation set consisted of images of the same person from certain heights, or distances, either closer than or farther from the train set. In order to see how well each model is able to distinguish a certain subject from a group of subjects at varying heights and distance, the train set contained two folders; one containing images of the target subjects at the desired distances, or heights, and the other containing images of all other subjects at the same distances and heights.

4 Results

4.1 Subject-Wise

In the subject-wise classification experiment, see Figs. 1 and 3, InceptionResNetV2 reached 50.56% accuracy in two epochs, 90% in four epochs, and eventually reached a top accuracy of 99.36% in its 19th epoch, after hovering between 95 and 98% after the fifth epoch. VGG19 stayed between 9.5 and 13% in the first four epochs, reached

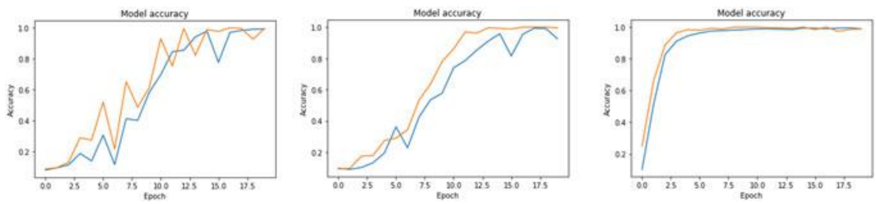


Fig. 1 Subject-wise, VGG16, and VGG19, respectively, orange—test, blue—train

57% in its 10th epoch, and moved to 91.17 in its 14th epoch. It reached its top accuracy of 99.36% in its 18th epoch. While it reached its top accuracy one epoch earlier, it took much more gradual progression to reach a noteworthy accuracy. VGG16 lingered between eight and 19% in its first five epochs, reached 59% accuracy after its 10th epoch, and obtained its top accuracy of 99.36%, the same as InceptionResNetV2 in its 20th epoch. VGG16 took the longest time to achieve significant, consistent improvements in results, about seven epochs. It also had the most erratic progression toward its top accuracy, going from 13.79 to 30.34% after its sixth epoch back down to 11.49% after its seventh epoch and from 97.89 to 78.03% after its 16th epoch and back up to 97.06%.

4.2 Train 0–4 m Test-5 m, Validation

When moving on to look at the effects of height on the recognition of an individual, using one versus one classification, there were four experiments done each person, independently. The models were trained from heights of 1.5–4 m and tested on a height of 5 m, and vice versa, and they were also trained on images taken from heights 1.5 and 3 m and tested on images taken from heights of 4 and 5 m. When trained from heights of 1.5–4 m, the IncpetionResNetV2 model reached a top training accuracy of 97.83%, the VGG19 had a top training accuracy of 72.79%, and VGG16 reached a top accuracy of 92.14%. The results do not seem to be very representative, as they are not consistently increasing values, rather they are very erratically going up and down, and they all had a testing accuracy of 0%, seldom jumping up to 100%. Upon observing that the validation accuracy was incredibly low, we decided to start training on 1.5 and 3 m and testing with heights of 4 and 5 m. Doing this increased the training accuracy slightly and allowed the VGG19 and inception models to show some consistency, increasing values after each epoch, reaching 100% accuracy in the eighteenth and seventh epoch, respectively, and the validation accuracy also showed improvement for the InceptionResNetV2 model with it reaching 82% accuracy in the fourth epoch and staying around 85% after that. VGG16 showed to significantly decrease in accuracy, as it stayed around 50%, without a notable increase, training and validation accuracy.

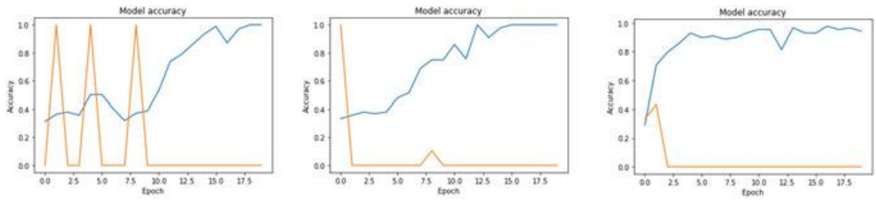


Fig. 2 Train 0–4, Test 5, VGG16, VGG19, respectively, and InceptionResNetV2, validation

4.3 Train 5 m Test 0–4 m, Validation

When training on 5 m heights and testing from heights of 1.5 m to 4 m, InceptionResNetV2 achieved a top training accuracy of 100% and validation accuracy 34.59%, VGG19 reached a top training accuracy of 100% and validation of 38.2%, and VGG16 had a top training accuracy of 100% and validation accuracy of 34.22%. When being trained on heights of 4 m and 5 m and test on heights of 1.5 m and 3 m, InceptionResNetV2 achieved the same accuracies, 100 and 34.59%, respectively, but VGG19 and VGG16 had much worse results, having very similar training and validation accuracies, both staying around 50%. Both cases showed fairly poor results with neither showing effective validation accuracies, but seemingly, training from farther away and testing using closer heights showed to be more accurate. In both tasks, training with close heights and training with farther away heights, InceptionResNetV2 seemed to perform better than both VGGnet models.

4.4 Train 0–4 m Test-5 m, Identification

When performing the identification tasks, we started with training on heights from 1.5 to 4 m and testing on a height of 5 m. InceptionResNetV2 achieved a top training accuracy of 99.8% and validation accuracy 100%, VGG19 reached a top training accuracy of 100% and validation of 100%, and VGG16 had a top training accuracy of 100% and validation accuracy of 91.54%. These results for InceptionResNetV2

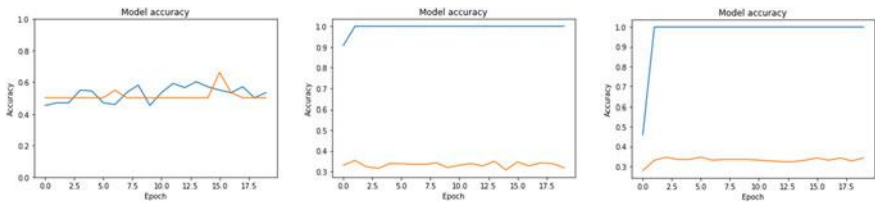


Fig. 3 Train 5, Test 0–4, VGG16, and VGG19, respectively, validation

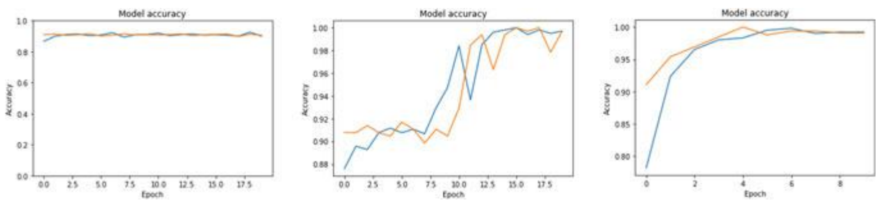


Fig. 4 Train 0–4, Test 5, VGG16, VGG19, and InceptionResNetV2, respectively, identification

and VGG19 seemed much more promising as the models actually seemed to progressively show improvement at distinguishing features, but VGG16 showed almost no improvement throughout all 20 epochs. When being trained on heights of 5 m and tested on heights of 1.5–4 m, InceptionResNetV2 achieved a top training accuracy of 100% and validation accuracy 95.24%, VGG19 reached a top training accuracy of 91.95% and validation of 91.56%, and VGG16 had a top training accuracy of 100% and validation accuracy of 94.35%. InceptionResNetV2, again, and VGG16 provided evidence that it had the capabilities to identify faces at a height higher than that of the height it was trained with, while VGG19, in this case, had much more stagnant results showing little improvement.

4.5 Train 5 m, Test 0–4 m, Identification

When being trained on heights of 5 m and tested on heights of 1.5–4 m, InceptionResNetV2 achieved a top training accuracy of 100% and validation accuracy 95.24%, VGG19 reached a top training accuracy of 91.95% and validation of 91.56%, and VGG16 had a top training accuracy of 100% and validation accuracy of 94.35%. InceptionResNetV2, again, and VGG16 provided evidence that it had the capabilities to identify faces at a height higher than that of the height it was trained with, while VGG19, in this case, had much more stagnant results showing little improvement.

5 Discussion

All three models very adequately categorize the subjects when they are presented with random test images. Generally, after being completely trained, 20 epochs, each model had approximately the same accuracy. With that said, InceptionResNetV2 increased in accuracy at a much faster rate and needed much less time and a lower number of training epochs to reach its top accuracy. The VGG16 and VGG19 models would take longer to reach an accuracy comparable to the InceptionResNetV2 model. Depending on the deep learning model or the task, identification or validation, the effect of heights varies from having a seemingly large effect to having a seemingly

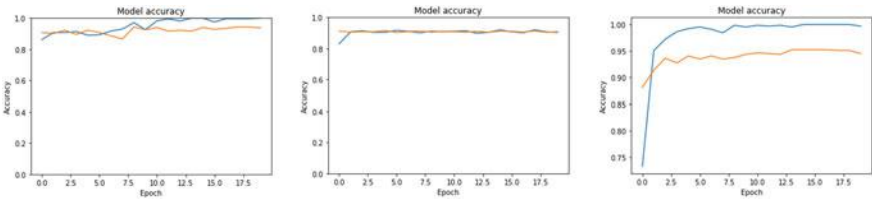


Fig. 5 Train 5 m, Test 0–4 m, VGG16, VGG19, and InceptionResNetV2, respectively, identification

insignificant effect on the accuracy of a model on a certain task. The identification task appeared to be affected much less prominently by the varying heights at which the images were taken, but in the validation task, the models looked as though they were heavily affected by the differences in heights. The model that had the most consistent, accurate results was InceptionResNetV2. It not only had the smoothest, the fastest rate of increase in the subject-wise examination, but it also had the highest accuracy when attempt to account for the differences in heights between train and test sets in the identification and validation tasks.

6 Conclusion

Depending on the deep learning model or the task, identification or validation, the effect of heights varies from having a seemingly large effect to having a seemingly insignificant effect on the accuracy of a model on a certain task. Each model had its own strengths and weaknesses, and each test brought out what each of those was. All three models were able to achieve above 99% accuracies, on test and train sets, in the subject-wise assessment and in the identification task and were able to reach a 100% training accuracy in the validation experiment but results for the test set would usually significantly drop, to about a 35% top testing accuracy. The reason for this is probably the lack of photographs available in the train set. While all models seemed to have high rate in correctly distinguishing between a finite group of individuals when trained on a diverse set of images of all individuals in the group, InceptionResNetV2 needing the least amount of training before producing accurate results, it is obvious that unless we provide these models with a large coherent set of images for training, accurate results will not be reached. This is shown by the differing train and test accuracies when the sample of images was significantly reduced and was significantly less diverse. Unfortunately, with real-world practices, such as searching for a lost individual or criminal, it would be hard to generate such a dataset to use to train the models, but CNN models do show that they could have the ability to perform these tasks in the correct circumstances.

Acknowledgements We would like to acknowledge the support from the National Security Agency (Grant #H98230-19-1-0012), the National Science Foundation (NSF) (Grant HRD #1719498), and Army Research Office (Contract No. W911NF-15-1-0524).

References

1. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M.: Deep learning for visual understanding: a review. Neurocomputing **187**, 27–48 (2016)
2. Koustubh: ResNet, AlexNet, VGGNet, inception: understanding various architectures of convolutional networks (2017). https://cv-tricks.com/cnn/understand-resnet-alexnet-vgg-inception/

3. Hsu, H., Chen, K.: Face recognition on drones: issues and limitations. In: Proceedings of the first workshop on micro aerial vehicle networks, systems, and applications for civilian use, pp. 39–44 (2017)
4. Bonetto, M., Korshunov, P., Ramponi, G., Ebrahimi, T.: Privacy in mini-drone based video surveillance. In: 2015 IEEE International Conference on Image Processing (ICIP) (2015)
5. Hsu, H., Chen, K.: DroneFace: an open dataset for drone research. In: Proceedings of the 8th ACM on multimedia systems conference, pp. 187–192 (2017)
6. Bindemann, M., et al.: Person identification from aerial footage by a remote-controlled drone. Sci. Rep. **7**(1) (2017)
7. Satayanaryana, S., Niharika, D., Sravya, P., BhanuPriya, G.: Advanced face detection to trend analysis in image processing based on drone primitives. Int. J. Pure Appl. Math. **115**(6), 31–36 (2017)
8. Tiddeman, B., Burt, M., Perrett, D.: Prototyping and transforming facial textures for perception research. IEEE Comput. Graph. Appl. **21**(4), 42–50 (2001)

Traffic Sign Recognition for Self-driving Cars with Deep Learning

Daniel Xie, Emmanuel Nuakoh, Prosenjit Chatterjee, Ashkan Ghattan, Kossi Edoh, and Kaushik Roy

Abstract The purpose of this research was to create a model for an autonomous car in traffic sign recognition. A high-accuracy model is needed to analyze the signs. Previous studies have mainly been centered on European countries, and the models created in Europe are not applicable to American autonomous cars. The contributions of this paper are twofold. First, this study generated a dataset that was collected and annotated in order to establish a suitable model for the USA. The dataset was custom made and acquired by using camera footage that was converted into individual frames. The dataset was named Cyber Identity and Biometrics Lab Traffic Sign Dataset Version 1 (CIB TS V1). Then, it was annotated into different classes and labels with LabelIMG. With a customized program, we used the annotations to crop out images and categorized them. Second, the data was run through a deep learning algorithm called modified AlexNet. A lighter version of the AlexNet was used for our experiments. Results showed that the model achieved above 99% accuracy on the validation set.

Keywords Autonomous car · Traffic sign recognition · Deep learning · Image data collection

D. Xie (✉)
North Carolina School of Science and Math, Durham, NC, USA
e-mail: xie21d@ncssm.edu

E. Nuakoh · P. Chatterjee · A. Ghattan · K. Roy
Department of Computer Science, NC A&T State University, Greensboro, NC 27411, USA
e-mail: ebnuakoh@aggies.ncat.edu

P. Chatterjee
e-mail: pchatterjee@aggies.ncat.edu

A. Ghattan
e-mail: aghattan@aggies.ncat.edu

K. Roy
e-mail: kroy@ncat.edu

K. Edoh
Department of Mathematics, NC A&T State University, Greensboro, NC 27411, USA
e-mail: kdedoh@ncat.edu

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_19

1 Introduction

Throughout the advancement of technology, one of the most recent accomplishments is the creation of self-driving cars by companies such as Volkswagen, Audi, and Toyota. These self-driving cars are designed using a model to view the traffic signs that they are approaching, so they can abide by the traffic standards. A lot of research in traffic sign detection has been done in Europe [1], including the DFG dataset in Slovenia [2] and the BelgiumTS in Belgium. However, there are great differences between European traffic signs and American ones, so the European model cannot be implemented in the USA. Also, even though there have been studies on American-based traffic sign datasets, such as the LISA dataset [3], few studies have explored the effects of customized datasets on the creation of the model. Recently, deep learning techniques have been used for traffic sign detection and recognition [4, 5] and produced promising results. This study will establish a deep learning model [6] for autonomous cars in traffic sign recognition using customized datasets.

There have been studies conducted by researchers in regard to traffic sign recognition in autonomous cars, but few of them were conducted using the American traffic signs with customized datasets [4].

Tabernik and Skocaj [4] utilized the DFG dataset, which is a Slovenian-based traffic sign dataset, using end-to-end learning with ResNet and AlexNet. The DFG dataset is a very unique dataset due to its large number of classes and its generally small dataset size. The strength of this dataset in their research is that all the images are already cropped and annotated, so it is unnecessary to do any data wrangling to run this data through an algorithm to produce a model. However, it contains 200 classes with 5254 training images and 1703 test images; and every class has minimum 20 images [2], which is generally insufficient data for 200 classes. The traffic signs were also created by a Slovenian company, which would not be applicable in the USA due to its foreign signs. One of the only few available datasets in the USA is the LISA dataset [3]; however, it has low resolution and has been used in many studies.

Another study used a unique method in creating a model for traffic detection in autonomous cars [7]. The researchers converted images of road signs into grayscale, which were filtered with simplified Gabor wavelets (SGW). Then, they located the objects in the image that resemble traffic signs and selected those images. Next, they used a neural network to classify traffic signs into subclasses. Using this model, the researchers achieved a speed of process at 6.3 frames per second with a 99.43% accuracy with the image sizes being 1360×800 pixels. However, it is noticed that the dataset was not highly mentioned in the research; therefore, it could be entirely possible that the dataset was not sufficient or representative in real-world applications. What the researchers recommended in the study was dimension reduction of feature vectors and the use of a GPU.

In another study, Forson used a dataset for the model that included images that were 32×32 pixels in RGB scale [5]. The training set included 34,799 images, the validation set had 4,410 images, and the test set included 12,630 images. Forson used the grayscale version of the image and convoluted it with three layers. The model resulted in high accuracy with 99% on the validation set and 98% on the test. The researcher suggested using batch normalization and other neural networks, such as ResNet and Xception.

2 Methodology Used

The overall end goal is to use deep learning technology to create a model for traffic sign recognition. The core of the deep learning algorithm is a modified version of AlexNet [6]. The AlexNet is a type of convolutional neural network that is useful for image classification. To feed the deep learning model, roadside actual video data has been collected and separated into individual frames. For each frame, an annotation procedure has been conducted to assign categories based on the image contents. Then, the annotated images feed into the modified AlexNet [6] that is comprised of max pooling, convolutional, flatten, and dense layers with parameters as shown in Fig. 1.

2.1 Creating Annotations with LabelIMG

In order to create a customized dataset, we prepared raw camera footage that was organized frame by frame. The data was collected mainly in Greensboro, North Carolina area. The data was categorized into different days based on when the camera footage was taken. The data we used was formatted into the six days it was collected, as shown in Fig. 2. The first folder labeled 100318 was taken on 10/03/18, and

Fig. 1 AlexNet [6] architecture used in this study

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 32, 124, 124)	832
max_pooling2d_1 (MaxPooling2	(None, 32, 41, 41)	0
conv2d_2 (Conv2D)	(None, 64, 37, 37)	51264
max_pooling2d_2 (MaxPooling2	(None, 64, 12, 12)	0
conv2d_3 (Conv2D)	(None, 128, 8, 8)	204928
max_pooling2d_3 (MaxPooling2	(None, 128, 2, 2)	0
flatten_1 (Flatten)	(None, 512)	0
dense_1 (Dense)	(None, 64)	32832
activation_1 (Activation)	(None, 64)	0
dropout_1 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 10)	650
activation_2 (Activation)	(None, 10)	0

Total params: 290,506
Trainable params: 290,506
Non-trainable params: 0

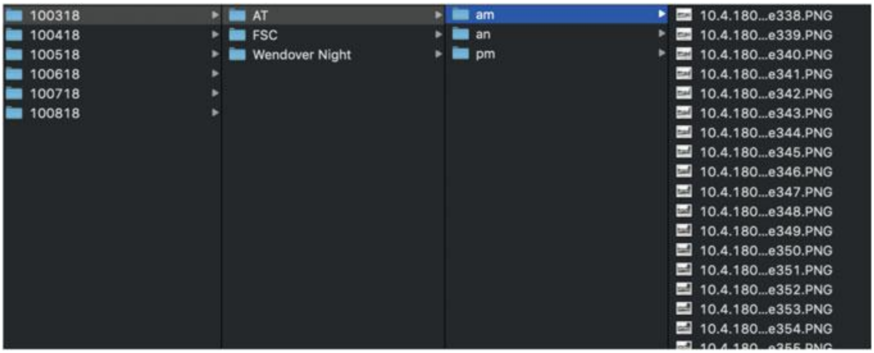


Fig. 2 Raw data structure folder

each following day was consistent with the format mm/dd/yy. Within the 100318 folder, there were different locations given, including an area named AT (NC A&T University), FSC (Friendly Shopping Center), and Wendover Night (Wendover is the street's name and the camera footage was taken during the night). The AT folder was then composed of am (before midday), an (afternoon), and pm (evening). Every other folder unmentioned of its contents was filled with images and did not include any other folders.

Before creating the annotations, a folder was made for the directory in which the annotations would be saved. After the directory for the annotations was assigned, Label IMG was installed from this link: https://github.com/tzutalin/labelImg. The installation varies from different operating systems, but all were generally similar and instructions were included in the link. Once the installation was complete, we located the directory in which the program was saved in the terminal and then launched it with the command "python3 labelIMG.py".

We annotated the data using LabelIMG [8], which opens images frame by frame in designated folder with the "Open Dir" option. Then, a save folder was designated where the annotations were saved with the "Change Save Dir" option (Fig. 3). The authors went through frames to find traffic signs whenever it appeared in the frame image (Fig. 4). Then, a rectangle box around the traffic sign was created and a category was assigned to it (Fig. 5).

For the 100318 folder, the researchers annotated 93 frames in the 100318/AT/am folder, 58 frames in the 100318/AT/an folder, and 244 frames in the 100318/AT/pm folder. For the 100418 folder, there were 150 frames that were annotated. Some of these frames had more than one traffic sign, resulting in a total of 736 sample images.

For the 100618 folder, the researchers added 730 annotations, some of which had more than one traffic sign annotated. After all of the annotations were converted into data samples, there were a total of 1743 images creating a total of 17 different categories. However, some of these categories had fewer than 50 images, and therefore were removed. The end product of the total dataset was 1628 data samples with 10 different classes and was split in an 80/20 ratio between the training and test set,

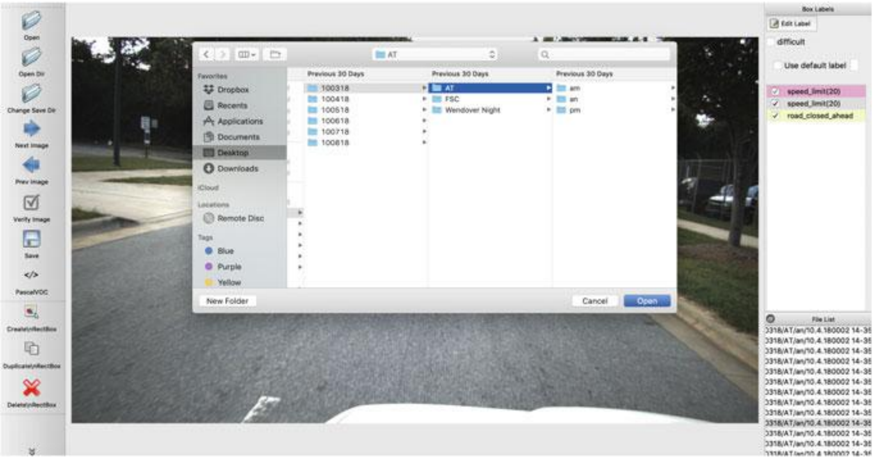


Fig. 3 Open directory in LabelIMG

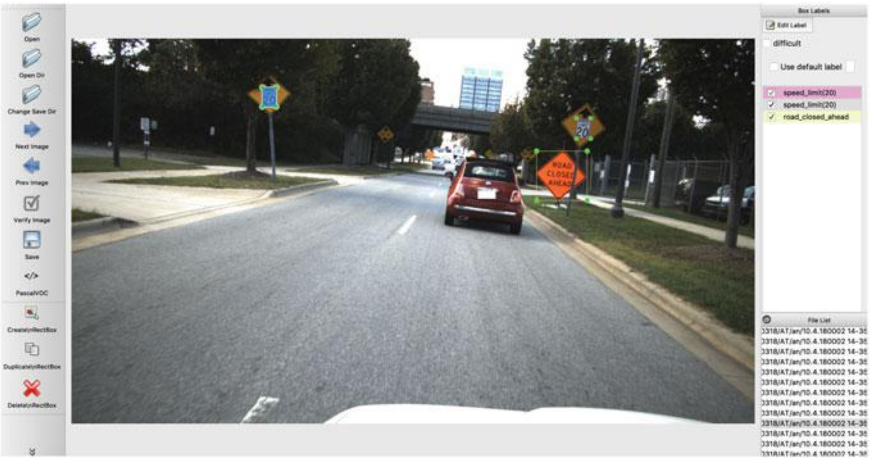


Fig. 4 Annotating traffic signs

which can be seen in Table 1. The dataset was named Cyber Identity and Biometrics Lab Traffic Sign Version 1 (CIB TS V1).

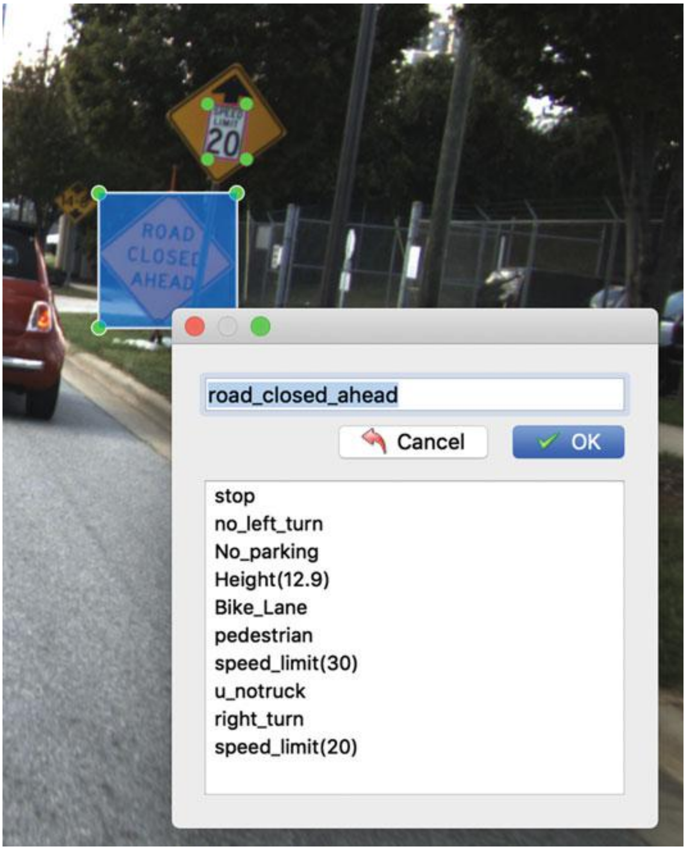


Fig. 5 Labeling a traffic sign

Table 1 Training and test data used for experiments

Class number and label	Training samples	Test samples	Total samples
0. stop	181	52	233
1. no_left_turn	112	29	141
2. No_parking	309	81	390
4. Bike_Lane	211	49	260
5. pedestrian	146	29	175
6. speed_limit (30)	43	10	53
7. u_notruck	73	19	92
12. speed_limit (35)	102	28	130
13. u_yield	87	16	103
15. u_railroad_intersection	38	13	51
Total	1302	326	1628

2.2 Converting Annotations into Cropped Images

A Python program was created in order to run through all the annotations to extract the traffic signs and associate them with their given labels. It ran through the annotations folder to examine each xml file, within which the program located a label and identified its respective coordinates. Next, the program used that information, found the annotation's respective frame file, and cropped out that image. At last, that particular image was placed in a folder by the program, which aligned with its given label.

3 Results

The authors created a modified AlexNet model in order to accommodate hardware limitations available during training for the computer. The CIB TS V1 dataset was fed into a multiclass AlexNet with 40 epochs to generate a model. The model utilized a categorical cross-entropy loss function, which was not only optimized for multiclass datasets but also useful for a 10 class dataset. The specifications on the multiclass AlexNet are given in Fig. 1.

3.1 Confusion Matrix

Confusion Matrix is the output of the deep learning model, which shows what the validation samples' true labels were and what the model predicted them to be. The elements in the diagonal of the matrix show the counts for the model predicting the correct classes. The validation data of the model was shown in Fig. 6. It can be seen that most images were predicted correctly. Table 2 shows the classes and their corresponding labels. The only error was in class 4, which was the height (12.9) traffic sign, and it was confused with class 7, which was the speed_limit (30) traffic sign.

3.2 Loss Graphs

The loss graphs seen in Fig. 7 were built off the loss values for each epoch in the outputs from the model. It was generally considered that having a loss value above 3.0 would be deemed unacceptable, because it meant that the model did not adequately and appropriately fit the dataset. The program was run with 40 epochs, and Fig. 7a represents the loss values of the first 20, while Fig. 7b represents the loss values second 20. The value of the loss function greatly decreased by the 5th

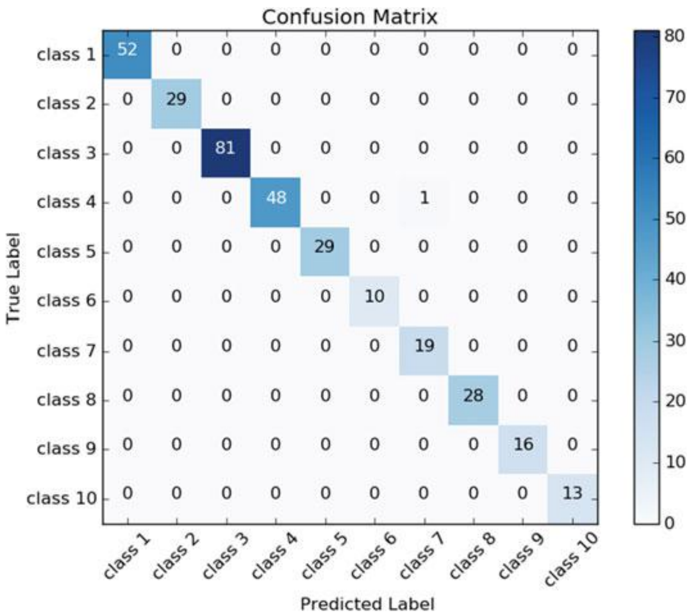


Fig. 6 Confusion matrix of validation

Table 2 Class names in confusion matrix

Class #	Class name
1	stop
2	no_left_turn
3	No_parking
4	Height (12.9)
5	Bike_Lane
6	pedestrian
7	speed_limit (30)
8	u_notruck
9	right_turn
10	speed_limit (20)

epoch and stayed very close to zero consistently afterward. In Fig. 7b, it shows a lot of staggering, but that is due to how enlarged the graph is, and what is seen as very unstable is mainly very small changes in loss in the different epochs.

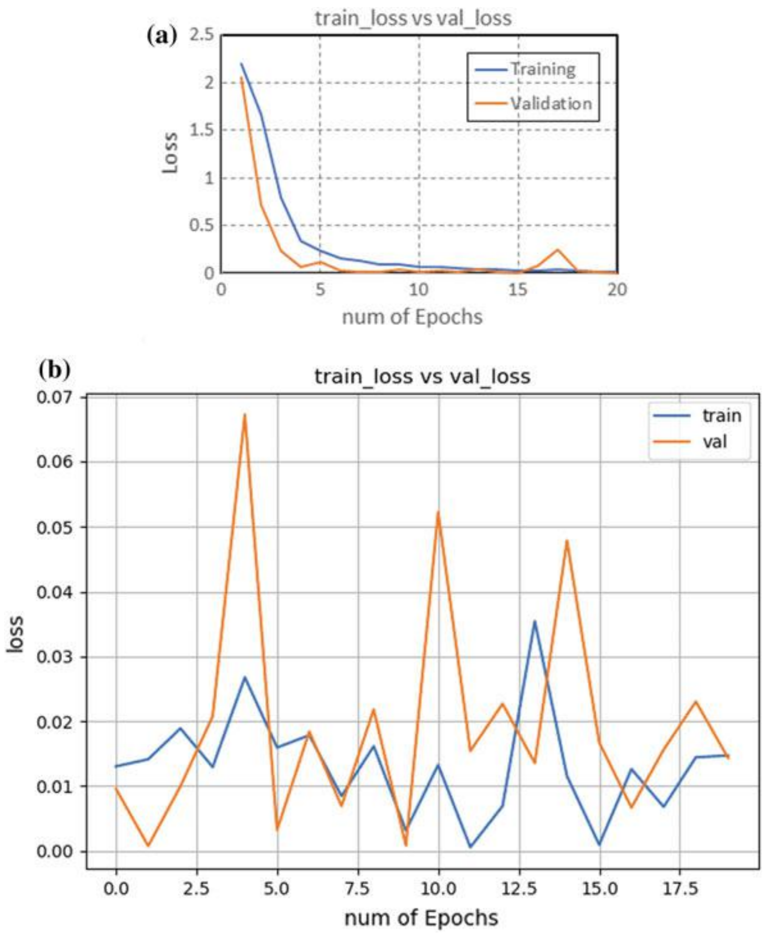


Fig. 7 Train loss and validation loss graphs, **a** the first 20 epochs, **b** the second 20 epochs

3.3 Accuracy Graphs

The accuracy looks at what percentage of the data model can correctly identify the labels for the images. In the two graphs in Fig. 8, the charts show that the accuracy quickly began to rise over the course of the 40 epochs. After about 5 epochs, the accuracy for validation goes above 97% and continues to increase. After about 10 epochs, training goes over 98% accuracy. Afterward, both accuracies stay rather consistent. In Fig. 8b, the graph always stays above 99% which is very high accuracy.

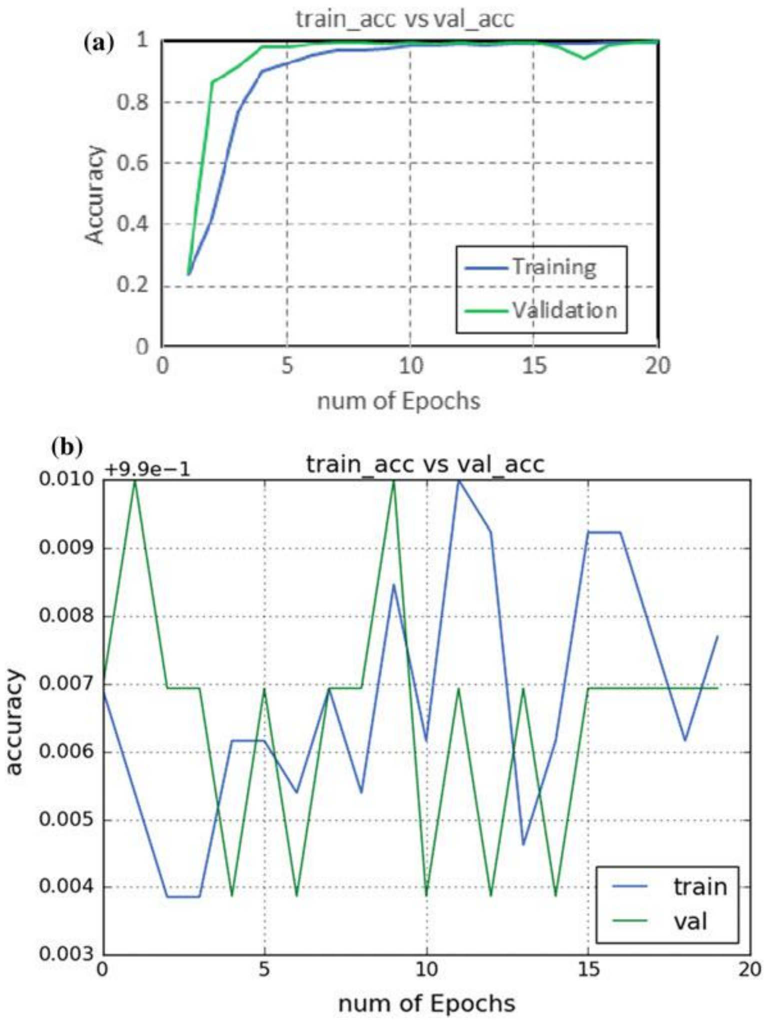


Fig. 8 Train accuracy and validation accuracy graphs, **a** the first 20 epochs, **b** the second 20 epochs

4 Conclusion

The AlexNet model utilized many different kinds of layers, including convolution, max pooling, dense layers, and a few more. It was very useful in image classification, and we used it to create the high-accuracy model that identified traffic signs. The accuracy of the model also highly depends on the dataset. Creating a customized dataset is beneficial in learning how datasets are created. Since most traffic sign research occurs in Europe, their models could not be implemented in the USA due to how varied the American and European signs may be. Creating annotations requires manual operation and analysis in the program, which may be time-consuming. However, it is necessary to create a good model for autonomous vehicles to accurately and appropriately identify traffic signs, so that a safer traffic environment could be ensured. The results of the model show an estimate of 99.69% accuracy from the validation set. This is a substantially high outcome; however, it has issues that it contains 10 different traffic signs, while there are a few dozen different ones. For example, a speed limit sign has many different versions which each require their own separate class. It is recommended that future work should include more annotations, depending on how many different classes are needed to create a suitable model. Also, there should be more traffic signs that could be included on areas such as highways. These datasets may also be processed with a more complex deep learning algorithm, such as ResNet. These results can be compared with the results of the data being run with the modified AlexNet in this work.

Acknowledgements We would like to acknowledge the support from the National Security Agency (Grant #H98230-19-1-0012), the National Science Foundation (NSF), and Grant HRD #1719498.

References

1. Møgelmose, A., Liu, D., Trivedi, M.: Detection of US traffic signs. IEEE Trans. Intell. Transp. Syst. **16**(I.6) (2015)
2. DFG Consulting d.o.o.: DFG traffic sign data set (n.d.). Accessed https://www.vicos.si/Downloads/DFGTSD
3. Møgelmose, A., Trivedi, M., Moeslund, T.: Vision based traffic sign detection and analysis for intelligent driver assistance systems: perspectives and survey. IEEE Trans. Intell. Transp. Syst. **13**(I.4) (2012)
4. Tabernik, D., Skocaj, D.: (2019). Accessed https://arxiv.org/pdf/1904.00649.pdf
5. Forson, E.: Recognising traffic signs with 98% accuracy using deep learning (2017). Accessed https://towardsdatascience.com/recognizing-traffic-signs-with-over-98-accuracy-using-deep-learning-86737aedc2ab
6. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems (2012)
7. Shao, F., Wang, X., Meng, F., Rui, T., Wang, D., Tang, J.: Real-time traffic sign detection and recognition method based on simplified Gabor wavelets and CNNs. Sensors (Basel) **10** (2018). https://doi.org/10.3390/s18103192
8. LabelIMG: Image labeling tool from https://github.com/tzutalin/labelImg

Identifying the Association Rule to Determine the Possibilities of Cardio Vascular Diseases (CVD)

Avijit Kumar Chaudhuri, Anirban Das, and Madhumita Addy

Abstract Depending on the insured's age, life insurance companies set the premiums. There are age-slabs for which premiums are set and after a proper medical examination, after a certain age, life insurance is issued. Major insurance company in India such as India Life Insurance Corporation Limited is seeking medical screening for any applicant above 45 years of age. Candidates whose health is not commensurate with age have been observed. This is particularly true of cardiovascular diseases (CVD). Therefore, the same can be tailored for individual candidates based on their medical test history instead of premiums based on age slabs. Checking for CVD, however, requires a number of medical tests, prompting both the applicant and insurance companies to use this method. This can be streamlined by conducting only main medical tests to determine the cardio-vascular system status of the applicant. The paper outlines the primary tests needed to be conducted to assess a person seeking health insurance's risk of cardiovascular disease. A series of association rules are developed to classify the risk of cardiovascular disease, using three well-proven methods. The three methods are Clustering of K Means, Decision Tree and Logistics Regression. This study suggests that premiums for health insurance should be based on the results of main assessments and their analysis of association rules. The combination of the three methods minimizes the Type 1 error.

Keywords Cardio vascular diseases (CVD) · CART · Logistic regression · K-means algorithm · Decision trees

A. K. Chaudhuri
Techno Engineering College Banipur, Banipur College Road, Habra 743233, India
e-mail: c.avijit@gmail.com

A. Das
University of Engineering & Management, Kolkata, India
e-mail: anirban-das@live.com

M. Addy (✉)
Imerit Technology Pvt. Ltd, Kolkata, India
e-mail: madhuaddy7@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_20

1 Introduction

The Data Mining Techniques (DMT) are very important for the medical management field to make the health and medical industry more powerful. Identification of inefficiencies and the best method that improve care and reducing costs is ascertained by using both data and the patterns and other meaningful information gathered from the analysis of data methodically. Thus Data Mining (DM) have a great future in healthcare industry. Other industries except the healthcare industry already implements and endorsed efficient DM and analytic strategies. Similarly, like data analytics and business intelligence the term DM can be defined as distinctive things to disparate people. In a straight forward way the DM can be defined as investigation of huge datasets to develop patterns and use those patterns to compute or anticipate the tendency of future events [1]. DM constitutes not only for the analysis or study of large volume of data sets but also for small or medium size data sets.

We can broadly classify analytics as follows:

- Descriptive analytics—Find the characteristics of manifestation
- Predictive analytics—Anticipating that can be established
- Prescriptive analytics—Making the conclusion for taking action for a specific consequence.

DM implements the intermediate level—predictive analytics level. Resolved patterns from different large and heterogeneous data sources are associated in DMTs. A new anticipated model can be designed by using these patterns. Thousands of industries profitably implement and use DMTs. Telecommunication industry exercises DM concepts for improvement of their promotional efforts. It helps determination of fraud in banking sectors. It can be used in various other fields like human life including business, education, medical, scientific and many more. Considering all the previous discussions we may found very few uses of DMTs in insurance industry. Only the academicians are used the most popular DMTs like regression analysis, choice modeling, rule induction, network or link analysis, clustering or ensembles, neural networks and decision tree analysis to prepare their research papers. Healthcare industry is always been very slow to adopt and comprehend the recent research findings into its regular practice.

Studies in the field of cardiovascular diseases, use of DMTs reveal that Decision tree, K-means and Logistic regression classifiers have obtained acceptable results. The authors have proposed association rules based on any of these single approaches that yield better results than others. The issues remain with any of these techniques and may affect the decisions if used in isolation. Use of any one of the techniques may result in type I error. The approach of this study evaluates three DM algorithms (K-Means Algorithm, Decision tree and Logistic regression) in their ability for identifying these key medical tests and creating association rules between the outcome of these tests and possibilities of CVDs.

2 Relevant Literature

Fixing premiums for health insurance in the health insurance market is a daunting problem in India. Many Indians also do not have healthcare coverage, and a lot of research work has been conducted, and much more needs to be done to identify the reasons to fix such coverage gaps. From the previous works we can identify that two key situations namely state and local level factors are beneficial to resolve the issue [2]. Firstly in local level individual person should take the responsibility to correctly decide and interpret the reasons which can reduce the uninsured rates. Secondly we should determine whether unchangeable characteristics of individuals are responsible for increase in uninsured rate or not. With the existing research work we can identify several factors which differ from insured and uninsured persons. Further, despite recognizing the differences between these factors, we cannot resolve the disparity in premiums [3]. Our study advances the current research by constructing association rules to fix premiums to those who do not have any coverage in health insurance. We shall ultimately use these association rules to help for reducing healthcare coverage disparity [4]. Recognition of these association rules without insurance coverage is very important as these association rules will lower access to medical care [5] and it should have more avoidable medications [6, 7]. There is a great relation of poor health condition leading to early death with the inadequacy of insurance coverage. The growing tendency of persons without health insurance in the past years has clearly intensified the issue [2, 8]. We noticed an increase of 18.4 million in the number of people without insurance in the period 1989–1999. In 2003, the uninsured population level soared to 16% of the total population [9]. The figure indicates that at some stage about 60 million people over a given year are uninsured. We study the factors that lead to an increasing number of non-insured persons at both state and local level [2] and many socio-analytical factors [10]. There is a link to this topic of life-style variables.

We observed that gender is also under consideration as an important socio-demographic variables to determine healthcare coverage [5, 11–13]. We can find from several analysis that generally men are more health insurance protected than women [5, 10], but in several cases the opposite observations are found in practical field [11]. We also consider race or ethnicity as an important factor for health insurance coverage disparity [5, 12] but the people belong to minority community have less healthcare coverage [3, 5, 11, 12]. A number on analysis shows that the low income group people have less healthcare coverage [11, 14]. With the help of information collecting from different sources we may say that the status of occupation has a very close association with health insurance. Not only employment status but also employment type can affect healthcare coverage. We can find the dissimilarity in percentage of healthcare coverage in different states, regions of the same state and different districts of a state [11, 12, 14]. Considering several perspectives we can suggest that senior adults have more insurance than junior adults [11, 12, 14]. When we study [9, 11, 15] about the difference between veterans and non-veterans, someone can find that a minimum number of veterans enjoy insurance at a minimum level with respect to the remaining population. Education and knowledge of people also have significant effect on insurance divergence [13].

Along with the previously discussed factors lots of other factors are present which will act as important factors in existing disparity. We shall also propose other various probable beneficent factors, both in socio-demographic and in society, and their augmentation to the growing divergence in healthcare coverage. Further, in this analysis we shall exercise the machine learning approaches to build classification models. Previously, health insurance divergence was studied using primary statistical techniques such as logistic regression [3] and basic allegorical statistics [11, 13, 15]. We have used linear regression as the primary technique for many years to assimilate and reproduce pragmatic relationships between dependent and independent variables, primarily taking into account their well-known statistically justifiable optimization strategies. Nonetheless, the comprehensibility of the design suffers in many problem structures as the false linear approximation of a non-valid feature. Using existing technology, as is discussed in this paper, machine learning techniques can easily model ideas such as healthcare coverage. Gauss–Markov's opinion (such as multicollinearity and normality) does not compel these systems, which is a significant involvement in more traditional models [16]. Such methods have already been used to research other health issues such as factors affecting impatient mortality [17] and prenatal care [18].

3 Methods

The source of data used in this study is the Hungarian Institute of Cardiology. CVD data set available at http://archive.ics.uci.edu/ml/datasets/Heart+Disease. The subjects were randomly selected form a dataset of 155 patients who went for medical examinations. Biometric data of the following factors, collected during the physical examination are shown in Table 1.

Decision tree classifiers, K-Means and Logistic Regression have achieved acceptable results relative to other literature techniques. This type of analysis can play a vital role in developing patient outcomes, minimizing medicine costs, and further advanced clinical studies. The above mentioned three methods were applied as revealed from the literature review to extract the entire set of association rule preventing type 1 error.

4 Results and Discussions

The simulation results (using Python programming language and SPSS) show the potential of these models with good accuracy and subsequent convergence for CVD prediction.

Table 1 Description of dataset

	Attributes	Description	Values
1	AGE	Age in years	Continuous
2	SEX	Male or female	1 = Male, 0 = Female
3	CP	Chest pain type	1 = typical type, 2 = typical type angina, 3 = non-angina pain, 4 asymptomatic
4	TRESTBPS	Rest blood pressure	Continuous value in mm hg
5	CHOL	Serum cholesterol	Continuous value in mm/dl
6	FBS	Fasting blood sugar >120 mg/dl	1 = true, 0 = false
7	RESTECG	Resting electrocardiographic results	0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy
8	THALACH	Maximum heart rate achieved	Continuous
9	EXANG	Execersise induced Angina	1 = yes, 0 = false
10	OLDPEAK	S T Depression induced by exercise relative to rest	Continuous
1 1	SLOPE	Slope of the peak exercise ST segment	1 = unsloping, 2 = flat, 3 = downsloping
12	CA	Number of major vessels colored	0–3
13	THAL	Defect type	3 = normal. 6 = fixed defect. 7 = reversible defect
14	NUM	Diagnosis of heart disease	0 = no heart disease (<50% narrowing)
			1 = has heart disease (>50% narrowing)

4.1 Logistic Regression

The variables in the equation figure (Fig. 1) include coefficients for the line (fitted) and other relative coefficients data. The line formula derived from the output is

$$\log(p/1-p) = -5.5683 + 0.6219 * \text{cp} + 0.4466 * \text{restecg} + 0.9547 * \text{ca} + 0.4649 * \text{thal} \tag{1}$$

This figure (Fig. 1) indicates that the regression model predicts the dependent variable significantly well. This indicates the statistical significance of the regression model that was run. Here, the significant values of cp, restecg, ca and thal is less than 0.05, and indicates that, overall, the regression model statistically significantly predicts the outcome variable (i.e., it is a good fit for the data).

```
                           Logit Regression Results
==============================================================================
Dep. Variable:                Disease   No. Observations:                  153
Model:                          Logit   Df Residuals:                      148
Method:                           MLE   Df Model:                            4
Date:                Wed, 23 Oct 2019   Pseudo R-squ.:                  0.3459
Time:                        22:53:08   Log-Likelihood:                -68.423
converged:                       True   LL-Null:                       -104.61
                                        LLR p-value:                 7.187e-15
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -5.5683      0.973     -5.721      0.000      -7.476      -3.661
cp             0.6219      0.233      2.668      0.008       0.165       1.079
restecg        0.4466      0.219      2.035      0.042       0.017       0.877
ca             0.9547      0.269      3.547      0.000       0.427       1.482
thal           0.4649      0.112      4.142      0.000       0.245       0.685
==============================================================================
```

Fig. 1 The output window provides a way of evaluating which variables may have predictive value

Model Validation

The output result from logistic regression summarizes the accuracy of prediction and other related parameters for group classification are shown below.

The accuracy of the model = TP + TN/(TP + TN + FP + FN) = 0.74
The Misclassification = 1 − Accuracy = 0.26
Sensitivity or True Positive Rate = TP/(TP + FN) = 0.63
Specificity or True Negative Rate = TN/(TN + FP) = 0.87
Positive Predictive value = TP/(TP + FP) = 0.83
Negative predictive Value = TN/(TN + FN) = 0.68
Positive Likelihood Ratio = Sensitivity/(1 − Specificity) = 4.691
Negative likelihood Ratio = (1 − Sensitivity)/Specificity = 0.43

4.2 Decision Trees

Evaluating the Model

This tree diagram (Fig. 2) shows that:

- thal factor is the best CVD indicator using the CHAID test.
- If value of thal is 6 and 7 the next best predictor for thal is ca and the next best predictor for thal value 3 is age.
- For ca values 1, 2 or 3 thal and ca itself are the primary factor of CVDs.
- This is known as a terminal node because there are no child nodes below it.
- For ca value 0 or missing the next best predictor is age. If the value of age group is either less than or equal to 50 along with thal value of 7 or 6, 81.8% of patients can be treated as CVD patients. if the age group of patients are greater than 50, then 21.1% are attacked in CVDs.

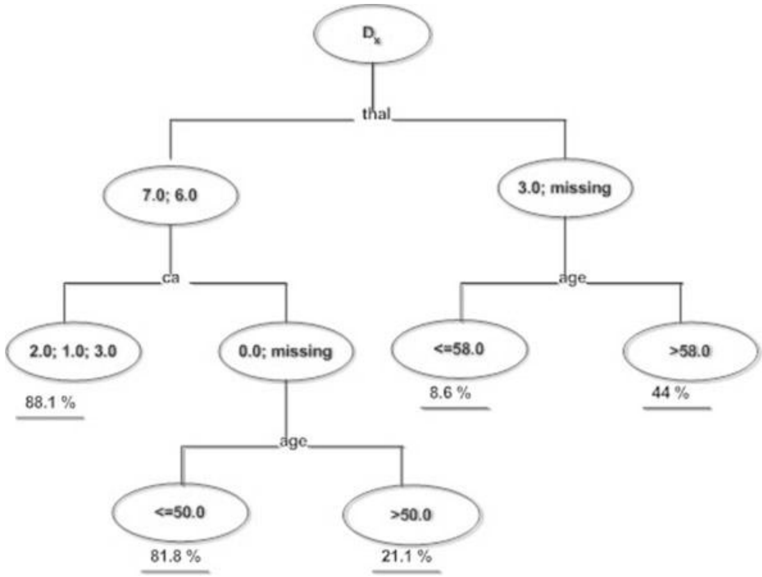


Fig. 2 Decision tree by CHAID algorithm

Fig. 3 Classification table of decision tree method

Observed	Predicted		
	0	1	Percent Correct
0	82	7	92.10%
1	20	46	69.70%
Overall Percentage	65.80%	34.20%	82.60%
Growing Method: CHAID			
Dependent Variable: heartdisease::category 0 1			

- On the other hand, if thal's minimum value is 3 or missing then the next best predictor will be age. For an age group of less than or equal to 58, about 8.6% of patients are attacked with CVD and for an age level of more than 58, about 44% are treated as heart patient.

The performance window (as shown in Fig. 3) offers a correlation between the outcome observed and the outcome expected. Method of growth: CHAID

4.3 K-Means Algorithm

Analysis with age, cp, ca and thal factor.

	Cluster				
	1	2	3	4	5
age	66	37	43	59	52
cp	3	3	3	3	3
ca	1	0	0	1	1
thal	5	4	4	5	5
heartdisease::category |0|1	0	0	0	1	0

Fig. 4 Final cluster centres

The analysis with age, cp, ca and thal factors shows (Fig. 4) that in the age group of 59 years may suffer from CVD due to cp value 3, and ca is 1 and thal is 5.

The association rule finally concluded is listed below.

4.4 Association Rules

4.4.1 Decision Trees

The set of association rule derived from the Decision Tree analysis is compiled below

If (6.0 < thal < 7.0) and (1.0 < ca < 3.0), Then the risk of CVD is 88%.
If (6.0 < thal < 7.0) and (ca is missing) and age <=50, Then there is an 81.8% chance of CVD.
If (6.0 < thal < 7.0) and (ca is missing) and age > 50, Then there is a 21.1% chance of CVD.
If (thal = 3.0) and age <=58, Then there is an 8.6% chance of CVD.
If (thal = 3.0) and age > 58, Then there is a 44% chance of CVD. 44%.

4.4.2 Logistic Regression

The set of association rule derived from the Logistic Regression analysis is compiled below

$$\log(p/1-p) = -5.5683 + 0.6219 * \text{cp} + 0.4466 * \text{restecg} + 0.9547 * \text{ca} + 0.4649 * \text{thal}$$

The Variables in the Equation table has several important elements. The statistics and associated probabilities provide an index of the significance of each predictor in the equation. The simplest way to assess statistic from the output is to take the significant values and if less than 0.05 reject the null hypothesis as the variable does

Comparison of accuracy levels and identification of significant variables for CVD dataset															
Approaches	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	Accuracy of predicting 1	Accuracy of predicting 0
Decision Tree	1											1	1	65.80%	34.20%
Logistic Regression			1				1					1	1	74.19%	25.81%
K Means	1		1									1	1		
Total	2		2				1					3	3		

Fig. 5 Comparison of accuracy level and identification of significant variables (CVD dataset)

make a significant contribution. In this case, we note that cp ($p = 0.008$), restecg ($p = 0.042$), ca ($p = 0.000$) and thal ($p = 0.000$) contributed significantly to the forecast of CVD but other variables did not (as p > = 0.05). So we will drop independents from the model when their effect is not significant by the statistic.

4.4.3 K-Means Algorithm

The set of association rule derived from the K-Means Algorithm analysis is compiled below

If (Age = 59 and ca = 1 and thal = 5 and cp = 3)
Then Probability of CVD is High
else
Probability of CVD is Low

The summary of the results and discussions are presented in the Fig. 5.

5 Conclusions

Objective of our work is to find the right age of persons suffering from CVDs. Literature review show that several DMTs of varied accuracy level exists. The above study shows that four variables (namely age, cp, ca and thal) typically affect CVD. The insurance companies fix the premiums based on these four variables. The slabs of the premium will vary according to the probability of CVD based on different values of these four variables. The techniques namely Decision tree classifiers, K-Means and Logistic Regression has been proposed to derive the complete set of association rule leading to identification of CVDs. Using any one of those methods is likely to contribute to the Type 1 error and none have 100% accuracy level.

Decision tree analysis show that factors namely thal, ca and age are significant to predict the probability of CVD. K-Means classifiers further shows that the age group for CVD can be lower if a partial variable set is considered. The logistic regression model indicates the probability of CVD as the sum of the product of these variables namely cp, restecg, ca and thal with their coefficients 0.6219, 0.4466, 0.9547 and .4649 respectively. That is, logistic regression establishes the weightages of the significant factors. However, a cluster of variables with specific values define the possibility of CVD. Therefore the significant factors need to be identified, their weightages determined and clustering with age is determined for customizations of insurance premiums. Therefore, the association rule can be established accordingly. Such findings can be the basis for customizing insurance premiums for both the categories of life and health instead of the traditional system of calculating premium on age-wise slabs.

The use of K-means Cluster approach suggests that the consideration of partial variable set reveal persons with age less than 40 can also suffer from CVD. This is true when the interactive effect of three variables namely sex, ca and trestbps are considered in lieu of all parameters. The decision tree further lays down the association rule strengthening the findings of K-Means analysis. The results of decision tree show that the existence of all three modifiable risk factors significantly increases the probability of CVD.

However, since the observations from K-Means analysis are more precise when age is focused, its results are included in the association rule. The Logistic Regression analysis show that the variables such as cp, restecg, ca and thal with definite weightages can lead to CVD. Thus it can be concluded that instead of using any one technique, the combination of most reliable techniques as listed in this study can be used to define the association rule.

References

1. Crockett, D., Eliason, B.: What is DM in Healthcare. Health Catalyst (2017)
2. Cunningham, P.J., Ginsburg, P.B.: What accounts for differences in uninsurance rates across communities. J. Health Care Org. Provision Fin. **38**(1), 6–21 (2001)
3. Glover, S., Moore, C.G., Probst, J.C., Samuels, M.E.: Disparities in access to care among rural working-age adults. J. Rural Health **20**(3), 193–205 (2004)
4. Wu, C.H., Kao, S.C., Su, Y.Y., Wu, C.C.: Targeting customers via discovery knowledge for the insurance industry. Exp. Syst. Appl. **29**(2), 291–299 (2005)
5. Monheit, A.C., Vistnes, J.P.: Race/ethnicity and health insurance status: 1987 and 1996. Med. Care Res. Rev. **57**, 11–35 (2000)
6. US Government Printing Office.: Healthy People 2010 Understanding and Improving Health. In: SDoHaHS (ed.) Office of Disease Prevention and Health Promotion, 2nd edn. (2000)
7. US Government Printing Office.: Healthy People 2010 Understanding and Improving Health. USDoHaHS (ed.) Office of Disease Prevention and Health Promotion, 2nd edn. (2000)
8. Herring, B.: The effect of the availability of charity care to the uninsured on the demand for private health insurance. J. Health Econ. **24**(2), 225–252 (2005)
9. Jonk, Y.C., Call, K.T., Cutting, A.H., O'Connor, H., Bansiya, V., Harrison, K.: Health care coverage and access to care—the status of Minnesota's Veterans. Med. Care **43**(8), 769–774 (2005)
10. Hendryx, M.S., Ahern, M.M., Lovrich, N.P., McCurdy, A.H.: Access to health care and community social capital. Health Serv. Res. **37**(1), 87–103 (2002)
11. Carrasquillo, O., Himmelstein, D.U., Woolhandler, S., Bor, D.H.: Going bare: trends in health insurance coverage, 1989 through 1996. Am. J. Pub. Health **89**(1), 36–42 (1999)

12. Carrasquillo, O., Carrasquillo, A.I., Shea, S.: Health insurance coverage of immigrants living in the United States: differences by citizenship status and country of origin. Am. J. Pub. Health **90**(6), 917–923 (2000)
13. Shi, L.Y.: Vulnerable populations and health insurance. Med. Care Res. Rev. **57**(1), 110–134 (2000)
14. Cardon, J.H., Hendel, I.: Asymmetric information in health insurance: evidence from the National Medical Expenditure Survey. Rand J. Econ. **32**(3), 408–427 (2001)
15. Woolhandler, S., Himmelstein, D.U., Distajo, R., Lasser, K.E., McCormick, D., Bor, D.H., et al.: America's neglected veterans: 1.7 million who served have no health coverage. Int. J. Health Serv. **35**(2), 313–323 (2005)
16. Uysal, M., Roubi, M.: Artificial neural networks versus multiple regressions in tourism demand analysis. J. Travel Res. **38**(2), 111–119 (1999)
17. Chae, Y.M., Kim, H.S., Tark, K.C., Park, H.J., Hoa, S.H.: Analysis of healthcare quality indicator using data mining and decision support system. Exp. Syst. Appl. **24**, 167–172 (2003)
18. Prather, J.C., Lobach, D.F., Goodwin, L.K., Hales, J.W., Hage, M.L., Hammond, W.E.: Medical DM: knowledge discovery in clinical data warehouse. In: Paper Presented at Annual Conference of AMIA 1997 (1997)

Prediction of Service-Level Agreement Violation in Cloud Computing Using Bayesian Regularisation

Archana Pandita, Prabhat Kumar Upadhyay, and Nisheeth Joshi

Abstract Service-level agreement (SLA) is a contract between the cloud service provider and consumer which includes terms and conditions of service parameters. The cloud service provider has to commit to service-level agreements, which ensures a specific quality of performance. A certain level of penalty is set if the provider performs SLA violations. Managing and applying penalties has become a critical issue for cloud computing. It is found to be of paramount importance that the violations are predicted well in advance so that the necessary measures can be taken. In this research work, various proactive SLA prediction models were designed, utilising the power of machine learning. We have used real-world data sets to highlight the accurate models for violation prediction in a cloud environment. Seeing violation prediction as a classification problem where incoming requests need to be predicted for the violation, we have used Bayesian regularised artificial neural network (BRANN) on different samples of real-world data set. Both the models show remarkable performance for predicting SLA violation. BRANN shows a significantly good average accuracy of 97.6%.

Keywords Bayesian regularised artificial neural network · Cloud computing · Neural network · Prediction · SLA violation

A. Pandita
Dept. of CSE, Birla Institute of Technology Offshore Campus, Ras al Khaimah, UAE
e-mail: archanapandita86@gmail.com

P. K. Upadhyay (✉)
Department of EEE, Birla Institute of Technology, Mesra, Ranchi, India
e-mail: uprabhatbit@gmail.com

N. Joshi
Department of CSE, Banasthali University, Vanasthali, Rajasthan, India
e-mail: jnisheeth@banasthali.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_21

1 Introduction

Cloud computing is one of the most cost effectual and flexible IT services [1]. With inventive concepts for new Internet services, developers do not need the huge investment expends in hardware for the deployment of their service or the human overhead to drive it. Cloud computing's motive is to make available a more significant amount of flexibility, scalability and availability of computing resources. To understand the importance of cloud computing, it is essential to understand its essential characteristics [2]. The main features are resource pooling, on-demand service and pay-per-use. Resource pooling allows multi-tenancy. There are different resources in terms of the virtual machine, storage, bandwidth, memory etc. which are shared by different users. These resources are dynamically allocated and deallocated to the users as per the demand. On-demand service offers massive scalability. It can scale from one to hundreds of systems and resources like storage and servers can be provisioned for various computations as per requirement. Pay-per-use reduces costs by letting the user pay according to the facilities and resources they use. The services consumed are recorded, controlled and monitored, which allows the facilitator to be entirely transparent with the customer. Cloud computing supports on-demand access to a set of resources shared over a network of remote servers [3].

Cloud computing delivers different services over the Internet such as servers, networking, databases, storage, analytics and various others. The services [4] can be categorised broadly as SaaS, PaaS and IaaS. Software as a service (SaaS) refers to the applications hosted by a service provider. A customer can access these applications via Internet. With the help of this service, we do not need to install or run the software. Rather than purchasing the software, customers subscribe to a SaaS service. This service provides high scalability which helps in providing better access or services for the customers. Platform as a service (PaaS) assists in providing an interface for the customers for developing, running and managing applications. Instead of creating and building software, this service delivers hardware and software tools required for building of the infrastructure. This service helps in providing an environment where the users can install their applications. They do not require maintenance for the applications. Infrastructure as a service (IaaS) provides functions such as monitoring the login and logout, security, load, data backup, recovery, etc. These services are policy-driven which are assumed to be much easier, faster and cost-efficient for operating and maintaining the infrastructure. Cloud computing provides the whole information technology as a service, which inherits the capability to process a massive amount of data by using the scheduling mechanism. Scheduling mechanism plays a vital role in cloud computing.

Microsoft Azure, Google Cloud Platform and Amazon EC2 are few cloud service providers (CSP) which provide services to cloud users—individuals and businesses. The CSP and cloud users are bound to each other with a contract known as service-level agreement (SLA) [5]. To ensure smooth and trustworthy services, it is imperative to define the SLA formally. The SLA serves as a guiding and monitoring tool for availability, performance and various other quality of service (QoS) parameters.

2 SLA Violation

Service-level agreement is a connection between a service provider and customers. It is a contract in which all the terms and conditions related to the quality of services are agreed upon by both the parties [6]. Any breach in the contract is known as service-level violation (SLV) and is subject to penalty [7]. SLA violation being our primary focus is emphasised in this work. Following are the three leading case of violation.

Case 1: When the requested and assigned resources do not match, resources like RAM, CPU and network are provisioned across the virtual machines, based on the policies and services in virtualisation architecture. If the requested amount of resource is more than assigned, then the system will not be able to fulfil the QoS and hence results in violation [8].

Case 2: When a degrade is caused in performance as a result of any failure, malfunction or multicomponent failure in the system [9, 10], the degrades can be monitored by the various tools, for example, real-time monitoring report.

Case 3: When the upgrade or the downgrade service request is not fulfilled within the agreed time frame [11], tools like IT service management (ITSM) are used for monitoring and control of scaling of services at different points of time.

Authors in [12] have used several allocated processors for violation detection. SLA is detected based on the processors allocated against the number of processors requested. At the application layer, SLA violation detection is proposed by [13] using cloud application SLA violation detection (CASViD). This architecture facilitates deploying, scheduling and allocation of cloud service. However, it does not have prediction features. There are many approaches in literature which use machine learning for SLA violation prediction. SLA violation is predicted during runtime in [14] using the regression technique on captured historical data. Authors in [15] put forward prediction of workload using the extension of regression in parallel applications. LoM2HiS, proposed in [16], uses heuristics to predict future violations based on predefined thresholds. Naïve Bayes model is used by authors in [17], which gives excellent performance. However, the data set used is generated by simulation and hence does not represent a real-world environment. In another similar research, authors [14] implemented regression technique in WEKA framework which is also not scaled to the real environment.

Based on the related work done by researchers, it is noticeable that research work on SLA violation prediction lacked real-world data sets and also, a variety of algorithms have not been tested to find the best classifiers. We think that there should be more practical methods that profiles and predicts SLA violations in the cloud by studying, comparing and testing various classification approaches to obtain the most accurate model for predicting future violations in real-world data sets. We plan on exploring new algorithms on the real-world data set to provide the best results for fault prediction in cloud computing.

3 Methods Used

Machine learning is the development of algorithms which learn from the historical data and can make a prediction when new data is exposed to it [18]. Artificial neural networks (ANN) are the computation models employing supervised, unsupervised and reinforcement learning inspired by biological neural networks which comprise several nodes. These nodes are connected which communicate through signals modelling the human brain. The output of each node is a nonlinear function of the sum of inputs. The connections have weights which increase or decrease as the learning proceeds. The training of a multilayer network becomes efficient when the backpropagation learning scheme is applied. Both regression and classification can be solved by the backpropagation algorithm. It is a well-ordered and productive model applicable for the computation of gradient. Since it requires a known outcome concerning its input to calculate the loss function gradient, it is a supervised learning approach. In this paper, we have used the methodology explained in further subsections.

3.1 Bayesian Regularised Artificial Neural Systems

Bayesian regularisation is a mathematical process that converts a nonlinear regression into a 'well-posed' statistical problem in the manner of a ridge regression [19]. As compared to the standard backpropagation method, BRANN is more robust and also eliminates the need for cross-validation which is a lengthy process. The fully connected network is divided into layers, and we have designed the model using Bayesian regularisation training function trainbr. Trainbr follows Bayesian regularisation backpropagation algorithm. The architecture of the network observed in the study has four independent input variables 'p_i' $= (pi_1, pi_2, pi_3, pi_4)$.

Weights (w_{kj}, $k = 1, 2, \ldots, N$) are used to associate input variables with each of N neurons in a hidden layer. The input in each neuron is equipped with an activation function $f(.)$ The combination activation functions used in this study are as follows

$$\begin{aligned}
&fhidden\,layer(.) = linear(.)\,and\,foutput\,layer(.) = linear(.)\\
&fhidden\,layer(.) = tangentsigmoid(.)\,and\,foutput\,layer(.) = linear(.)
\end{aligned}$$

Bayesian Regularisation artificial neural network learning algorithm updates the weight and bias values according to Levenberg–Marquartd optimisation. It minimises a combination of squared errors and weights and then determines the correct mixture so as to produce a network that generalises well, because of its good convergence and fast learning capability. As per Bayes' theorem, the probability of occurrence of two independent events $E1$ and $E2$ is obtained by Eq. (1)

$$P(E1\,and\,E2) = P(E1)P(E2) \tag{1}$$

also,

$$P(E2|E1) = P(E1\, and\; E2)/P(E1) \tag{2}$$

hence,

$$P(E1\, and\; E2) = P(E2|E1) * P(E1) \tag{3}$$

similarly,

$$P(E1|E2) = P(E1\, and\; E2)/P(E2) \tag{4}$$

combining (2) and (4) gives,

$$(E1|E2) = P(E2|E1)P(E1)P(E2) \tag{5}$$

where P $(E2|E1)$ represents a conditional probability of $E2$ given $E1$.

Providing solutions to several problems, they are difficult to overtrain since the training can be stopped based on the Bayesian criteria. It turns off the irrelevant parameters effectively and hence is tricky to overfit.

3.2 Normalisation

Normalisation is scaling the values from a minimum value to maximum value, which means adjusting the values to a common scale. Normalisation is significant in machine learning for data preparation so that the difference in the ranges of values does not change, which guarantees an unwavering convergence of weight and biases. In this paper, we have used min-max normalisation for both pre-processing and post-processing. Min-max is a normalisation technique which transforms X to Y linearly. For example, all features with the minimum value are transformed to 0 and maximum value to 1, and all other values get transformed between 0 and 1. The following equation is used

$$Y^{'} = \frac{(Y - \min(Y))}{(\max(Y) - \min(y))} \tag{6}$$

where Y is the original value and Y' is the scaled value. The minimum and maximum values are set as per the nature of data. To rescale the range between any arbitrary values say m and n, the following equation is used

$$Y' = m + \frac{(Y - \min(Y))(n - m)}{(\max(Y) - \min(y))} \tag{7}$$

3.3 Performance Measure

Measuring the performance is very significant to evaluate the model and validate if the expected output is achieved. In this paper, we have used two techniques for a performance measure, mean square error (MSE) and classification accuracy. MSE is the average of the square of the difference between the original values and the predicted values and is given by the following formula

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x - x_i')^2 \tag{8}$$

where $i = 1\ to\ n$ is an array of points x, and x' are the coordinates for each point. Classification accuracy is the ratio of correct predictions to the total number of data points and can be calculated using the following formula

$$A = \frac{P_c}{P_T} \tag{9}$$

where A represents the accuracy, P_c represents the correctly predicted values and P_T represents total prediction. The accuracy is mostly calculated in percentage and should be higher while MSE should be lesser.

In the proposed network, the input layer consists of three nodes, hidden layer with six nodes and output layer with single node as shown in Fig. 1.

Where X_1, X_2 and X_3 are three input variables of bandwidth, disk and memory. Activation function used for hidden layer and output layer are logistic sigmoidal and

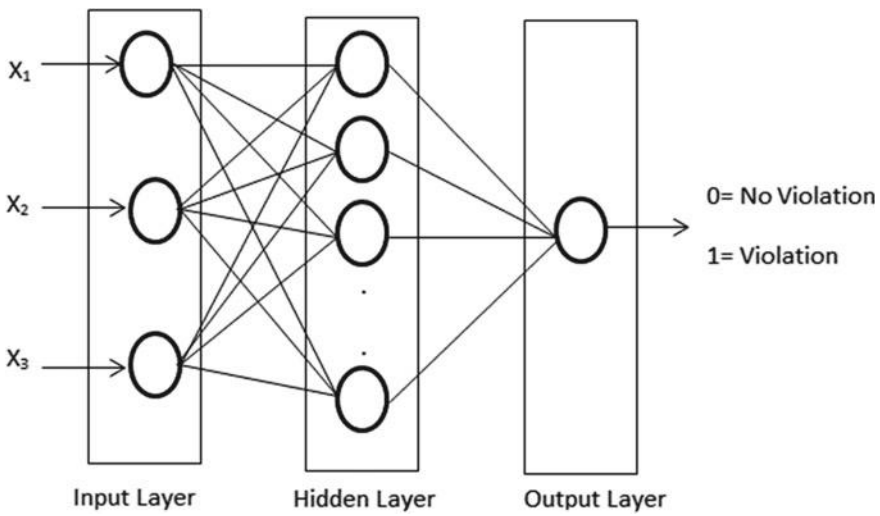


Fig. 1 Architecture

Table 1 Data set specifications

Dataset	Samples	Violation	Percentage (%)
DS1	870	77	9
DS2	2000	316	16

tangent hyperbolic function, respectively, as shown in Eqs. (10) and (11)

$$y = \frac{1}{1 + e^{-ax}} \tag{10}$$

$$y = \frac{1 - e^{-ax}}{1 + e^{-ax}} \tag{11}$$

where 'a' is slope parameter, x is input vector and y is the obtained response.

Features extracted from cloud Google trace data are presented at the input layer of the network, and net weighted input is computed from each node. Linear combiner output from these nodes pass through the activation function as mentioned in Eqs. (10) and (11) for different nodes of hidden and the output layer.

Errors are calculated from the node located in the output layer and are further back-propagated following Bayesian regularisation scheme. Weights are adjusted epoch by epoch which eventually results in a final weight matrix of size 6×4 for the hidden layer and 1×7 for output layer, where each layer nodes also have bias terms.

4 Data Set Description

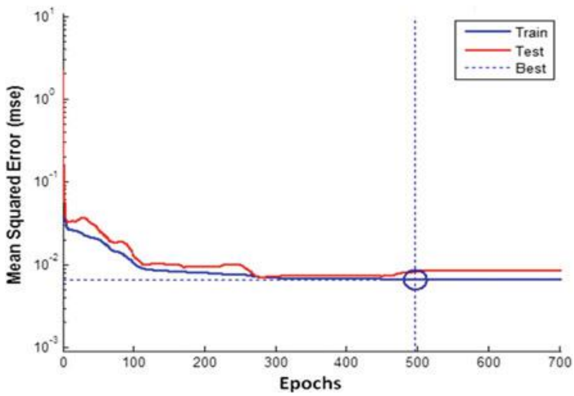
The experiment is carried out on the data set, which contains 29 days records of Google's Cloud [20]. The parameters bandwidth, disk and memory are scaled by dividing each with their respective highest value. We have used an inner join using MySQL database management system on events and usage tables in the Google cluster-usage traces data set to obtain any violation. Violation is found based on case 1 discussed in Sect. 3. Since Google cluster traces dataset consists of millions of rows, we have formulated two samples (DS1 and DS2) to test our prediction model. The reason to choose two samples is to avoid our model from overfitting problem. We have selected tuples at random for DS1, and to make sure the system will not face overfitting issue, we have chosen the second sample with twice of the fault rate. The specifications of the data sets are matched as follows (Table 1).

5 Results and Discussion

The model design and analysis have been carried out in MATLAB environment where codes have been written and executed. The data sets DS1 and DS2 were imported into the workspace, and the results obtained were saved and compared with original targets. We have applied BRANN algorithm to both the data set in which the size of hidden layers is set to be 6. To make sure that the result is not biased, the data is split into, training and test dataset randomly. 3-fold cross-validation is used to split the data into three partitions. The training of the model is done three times, wherein two-third and one-third data is used for training and testing, respectively. The final result is given by aggregating the results of all the runs. Prediction efficiency of the model is then obtained by finding the mean square error and the percentage accuracy. The data sets DS1 and DS2 have been presented to the proposed model, and prediction analysis is carried out. However, the results obtained for data set DS2 have only been discussed below.

While analysing training performance in SLA violation prediction using ANN models based on BRANN, training and testing curves thus obtained have been plotted. Figure 2 shows the best training performance for BRANN at epoch 628 with its respective mean square error values of 0.022977. It is also evident from these figures that predictive correlation is very high which reflects the fact that the model has acquired sufficient predictive ability. However, to remove plausible effects because of random values taken initially, quite a few independent BRANNs must be trained, and the reported value should be the average value of various runs. Figure 3 shows error histogram in which errors recorded in several instances have been depicted. In this figure, it can be noticed that the errors are falling in a narrow range of −0.1749 and 0.549 signifying that errors are symmetrically distributed about the zero mean for all the instances. Figure 4 shows the progress of other training variables when training is imparted to data set DS2. The training variables also include the gradient magnitude and the number of validation checks during offline training process. During training, sum square error (SSX) is constantly monitored while different values of momentum constant (mu) are assigned. The values of momentum constant, mu, was iteratively updated so as to avoid local minima problem in which sometimes network is likely to get stuck.

Fig. 2 BRANN training MSE performance (DS2)



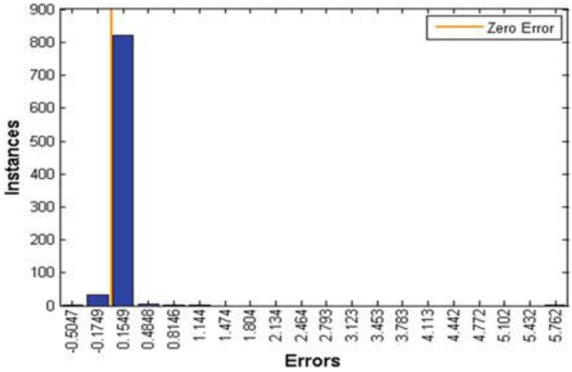


Fig. 3 BRANN error histogram (DS2)

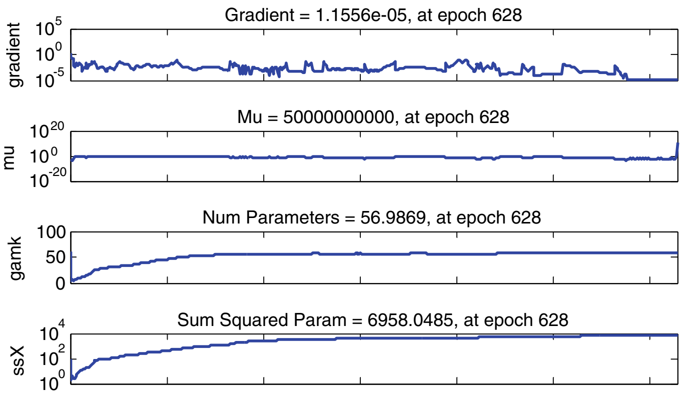


Fig. 4 BRANN training state (DS2)

A similar experiment was carried for data set DS1 on the same network, and the MSE for both the data sets is compared in Table 2. The average MSE is less than 0.038. The training statistics for BRANN, when applied to DS1 and DS2, are compared in Table 3.

Network is found to be fully trained at the iteration 716 for DS1 and 533 for DS2, where mean squared error is observed to be minimum and convergence of the algorithm is high. The curve shown in the Fig. 2 tends towards minimum gradient level at this iteration.

Table 2 MSE comparison for DS1 and DS2

Training function	Dataset	MSE	Average MSE
Trainbr	DS1	0.018376	0.0206765
	DS2	0.022977	

Table 3 Training statistics comparison for BRANN

Statistics	DS1	DS2
Epochs	716 iterations	533 iterations
Time	0:00:17	0:00:24
Gradient	3.21	1.55

Table 4 Accuracy

Algorithm	Sample	Accuracy (%)
BRANN	DS1	96.4
	DS2	97.6

5.1 Test Result

After the training is over, test data sets were presented at the input layer of BRANN, and performance accuracy has been recorded. The average accuracy of the models for data sets DS1 and DS2 are shown in Table 2. For data set DS1, BRANN has achieved 96.4% accuracy, and however, in the case of data set DS2, BRANN has 97.6% accuracy.

It is vital to provide reliable cloud service which does not violate the service-level agreement (SLA) and avoids penalties. After a study of various classification algorithms, we have applied BRANN algorithm on two different data sets and compared the percentage accuracy. The results obtained from this model appear to be quite encouraging, and hence, it can also be applied to the problems pertaining to different cloud scenario (Table 4).

6 Conclusion

This research work successfully compares the performance of BRANN fault predicting model on real-world data sets. Despite the data being skewed in terms of less number of violation in the data sets used, we obtained a high level of accuracy in the models we have designed. From the study, it has been observed that the Bayesian regularised artificial neural networks, which is an extension of backpropagation, provides the highest accuracy rates for given targets regardless of the dataset's size or structure. It avoids overfitting issues and builds the most generalised model which can be used with reliability. BRANN showed a very good performance accuracy of 96.4% and 97.6% for DS1 and DS2, respectively. This work can further be extended by using these prediction models in fault avoidance mechanisms and systems. This work can also be extended by testing more data sets on the above models to test the applicability of the mentioned results on a broader spectrum of real work usage in this domain.

References

1. Stanoevska-Slabeva, K., Wozniak, T.: Cloud basics–an introduction to cloud computing. In: Grid and Cloud Computing. Springer, Berlin, Heidelberg, pp. 47–61 (2010)
2. Regan, D.: Human brain electrophysiology: evoked potentials and evoked magnetic fields. Sci. Med. Part 3 (1989)
3. Ahmed, M., Chowdhury, A.S.M.R., Ahme, M., Rafee, M.M.H.: An advanced survey on cloud computing and state-of-the-art research issues. Int. J. Comput. Sci. Iss. (IJCSI) **9**(1), 201 (2012)
4. Jadeja, Y., Modi, K.: Cloud computing-concepts, architecture and challenges. In: 2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET). IEEE, pp. 877–880 (2012)
5. Mirobi, G.J., Arockiam, L.: Service level agreement in cloud computing: an overview. In: 2015 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). IEEE, pp. 753–758 (2015)
6. Wu, L., Buyya, R.: Service level agreement (SLA) in utility computing systems. In: Performance and Dependability in Service Computing: Concepts, Techniques and Research Directions. IGI Global, pp. 1–25 (2012)
7. Lu, K., Yahyapour, R., Wieder, P., Yaqub, E., Abdullah, M., Schloer, B., Kotsokalis, C.: Fault-tolerant service level agreement lifecycle management in clouds using actor system. Future Gen. Comput. Syst. **54**, 247–259 (2016)
8. Wu, L., Garg, S.K., Buyya, R.: SLA-based resource allocation for software as a service provider (SaaS) in cloud computing environments. In: Proceedings of the 2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. IEEE Computer Society, pp. 195–204 (2011)
9. Di Martino, C., Sarkar, S., Ganesan, R., Kalbarczyk, Z.T., Iyer, R.K.: Analysis and diagnosis of SLA violations in a production SaaS cloud. IEEE Trans. Reliab. **66**(1), 54–75 (2017)
10. Rajavel, R., Mala, T.: SLAOCMS: a layered architecture of SLA oriented cloud management system for achieving agreement during resource failure. In: Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), 28–30 Dec 2012. Springer, New Delhi, pp. 801–809 (2014)
11. Alhamazani, K., Ranjan, R., Mitra, K., Rabhi, F., Jayaraman, P.P., Khan, S.U., Bhatnagar, V., et al.: An overview of the commercial cloud monitoring tools: research dimensions, design issues, and state-of-the-art. Computing **97**(4), 357–377 (2015)
12. Musa, S.M., Yousif, A., Bashi, M.B.: Sla violation detection mechanism for cloud computing. Int. J. Comput. Appl. **133**(6), 8–11 (2016)
13. Emeakaroha, V.C., Ferreto, T.C., Netto, M.A., Brandic, I., De Rose, C.A.: CASVID: application level monitoring for SLA violation detection in clouds. In: 2012 IEEE 36th Annual Computer Software and Applications Conference. IEEE, pp. 499–508 (2012)
14. Leitner, P., Wetzstein, B., Rosenberg, F., Michlmayr, A., Dustdar, S., Leymann, F.: Runtime prediction of service level agreement violations for composite services. In: Service-Oriented Computing, ICSOC/ServiceWave 2009 Workshops. Springer, Berlin, Heidelberg, pp. 176–186 (2009)
15. Barnes, B.J., Rountree, B., Lowenthal, D.K., Reeves, J., De Supinski, B., Schulz, M.: A regression-based approach to scalability prediction. In: Proceedings of the 22nd Annual International Conference on Supercomputing. ACM, pp. 368–377 (2008)
16. Emeakaroha, V.C., Brandic, I., Maurer, M., Dustdar, S.: Low level metrics to high level SLAs-LoM2HiS framework: bridging the gap between monitored metrics and SLA parameters in cloud environments. In: 2010 International Conference on High Performance Computing & Simulation. IEEE, pp. 48–54 (2010)
17. Jules, O., Hafid, A., Serhani, M.A.: Bayesian network, and probabilistic ontology driven trust model for SLA management of cloud services. In: 2014 IEEE 3rd International Conference on Cloud Networking (CloudNet). IEEE, pp. 77–83 (2014)
18. Adeli, H.: Machine learning-neural networks, genetic algorithms and fuzzy systems. Kybernetes (1999)

19. Burden, F., Winkler, D.: Bayesian regularization of neural networks. In: Artificial Neural Networks. Humana Press, pp. 23–42 (2008)
20. Reiss, C., Wilkes, J., Hellerstein, J.L.: Google Cluster-Usage Traces: Format+Schema, pp. 1–14. Google Inc., White Paper (2011)

A New Methodology for Language Identification in Social Media Code-Mixed Text

Yogesh Gupta, Ghanshyam Raghuwanshi, and Aprna Tripathi

Abstract Nowadays, Transliteration is one of the hot research areas in the field of Natural Language Processing. Transliteration means that transferring a word from one language to another language and it is mostly used in cross-language platforms. Generally, people use code-mixed language for sharing their views on social media like Twitter, WhatsApp, etc. Code-mixed language means one language is written using another language script and it is very important to identify the languages used in each word to process such type of text. Therefore, a deep learning model is implemented using Bidirectional Long Short-Term Memory (BLSTM) for Indian social media texts in this paper. This model identifies the origin of the word from language perspective in the sequence based on the specific words that have come before it in the sequence. The proposed model gives better accuracy for word-embedding model as compared to character embedding.

Keywords Natural Language Processing · Character embedding · Word embedding · Machine learning

1 Introduction

Humans use natural language as their medium for communication. Natural Language Processing (NLP) is an area of artificial intelligence where a machine is trained to process the text to make human–computer interactions in an efficient way. There are various applications of NLP like machine translation, text processing, entity extraction, and so on. Code switching and mixing is a current research area in the

Y. Gupta (✉)
Department of Computer Science and Engineering, Manipal University Jaipur, Jaipur, India
e-mail: er.yogeshgupta@gmail.com

G. Raghuwanshi
Department of Computer and Communication Engineering, Manipal University Jaipur, Jaipur, India

A. Tripathi
Department of Computer Engineering and Applications, GLA University, Mathura, India

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_22

field of language tagging. Language identification (LID) is a primary task in many text processing applications and hence several researches are going on this area especially with the code-mixed data. With the emergence of several social media platforms and the availability of a large amount of text data in them, NLP plays a great role in understanding of that data today. People use social media to discuss their interests, hobbies, views, and so on. At the starting stage, only English was the language for these discussions but later trend was changed and people use multiple languages together. These kinds of languages are called code-mixed languages. An example of Hindi–English code-mixed text is shown below:

Example: Dear, ek idea hai mere mind me

In this sentence, both English and Hindi words are used such as dear, idea, and mind are English words and ek, hai mere, and me are Hindi words but written in Roman script. This is one of the examples of code-mixed language. Therefore, a technique is required to identify languages from code-mixed text used in three social media platforms, namely Facebook, Twitter, and WhatsApp.

This paper presents an innovative framework for capturing information at both word level and context level to output the final tag for language identification in context to the word belonging to which language. For word level, a multichannel neural network (MNN) is used, which is inspired by the modern research in the area of computer vision. Networks related to multichannel have proved the significance in generating the effective results in the NLP-related problems such as classification of complex sentences [1]. For capturing the context, Bi-LSTM-CRF is used.

The remaining paper is structured as follows: Sect. 2 discusses related work in this field. The proposed method is presented in Sect. 3. Section 4 describes the statistics of used dataset. The experimental results and their discussions are shown in Sect. 5. At last, Sect. 6 concludes the paper.

2 Related Work

In this section, some of the recent techniques regarding language transliteration and identification are described. King et al. [2] presented a system which generates the language identifier at word level based on semi-supervised methods. Nguyen and Dogru used CRF model limited to bigrams for identifying the language [3]. Das and Gamback used various features like a dictionary, n-gram, edit distance, and word context for identifying the origin of the word [4]. A shared task on Mixed Script Information Retrieval (MSIR) 2015 was conducted in which a subtask includes language identification of eight code-mixed Indian languages, Telugu, Tamil, Marathi, Bangla, Gujarati, Hindi, Kannada, and Malayalam, each mixed with English [5]. Word-level language identification was performed for English–Hindi using supervised methods developed by Jhamtani et al. [6]. Naive Bayes classifier was used to identify the language of Hindi–English data and an accuracy of 77% was obtained by Ethiraj et al. [7]. Language identification is also performed as a primary step to several other applications. Bhargava et al. [8] implemented a sentiment analysis system which utilized MSIR 2015 English–Tamil, English–Telugu, English–Hindi, and English–Bengali code-mixed datasets. Another emotion detection system was developed for Hindi–English data with machine-learning-based and Teaching-Learning-Based Optimization (TLBO) techniques [9].

Since the code-mixed script is the common trend in the social media text today, many kinds of research are going on for the information extraction from such text. An analysis of the behavior of code-mixed Hindi–English Facebook dataset was done in [10]. A shared task was organized for entity extraction on code-mixed Hindi–English and Tamil–English social media text [11]. Entity extraction for code-mixed Hindi–English and Tamil–English datasets was performed with embedding models. Sapkal et al. have given the approach by the use of SMS which is meant for communicating with others in minimal words [12]. The regional language messages are printed using English alphabets due to the lack of regional keywords. This SMS language may fluctuate, which leads to miscommunication. The focus was on transliterating short form to full form. Zubiaga et al. had mentioned language identification, as the mission of defining the language of a given text [13]. On the other hand, certain issues like quantifying the individuality of similar languages in multilingualism document and analyzing the language of short texts are still unresolved. The below section describes the proposed methodology to overcome the research gap identified in the area of transliterated code-mixed data. Alekseev et al. considered word embedding as an efficient feature and proposed entity extraction for user profiling using word-embedding features [14]. Chaudhary et al. proposed an approach to translate sentences from one language to another language using machine-learning techniques [15]. Samuel explored the character of language in intrapersonal and relational communication. He also defined the complexities of every communicative event [16].

3 The Proposed Methodology

The code-mixed data is the combination of the native script (familiar language) and the non-native script (unfamiliar language), which makes language identification more complex. This incorporates the script specification problem. The proposed framework is presented in Fig. 1, which indicates that the proposed framework contains two components. Initial component trains the MNN at the word level. However, another component is dedicated for the context level using bidirectional LSTM. The proposed framework works like feed-forward network where the output of the first stage becomes the input for another stage.

The proposed framework employs the word-level learning which follows the aforesaid concept. It is due to the flexibility of proposed system with various types of representations that can be used in recognition of various types of NLP problems for identifying the origin of a word in context to the language used in code-mixed data. The proposed network adopted four different channels, the initial three works

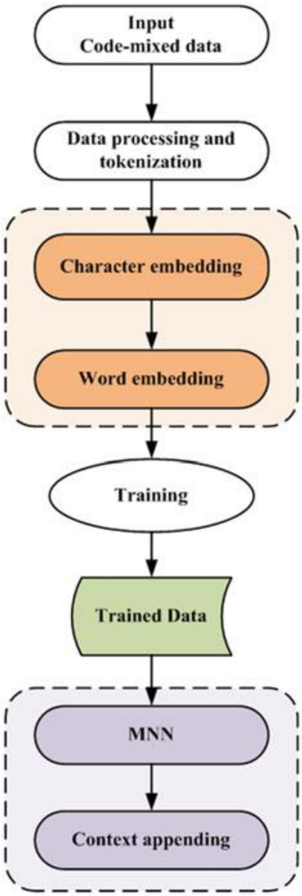


Fig. 1 Proposed framework of Word identification detection

as an input for the Convolution 1D (Conv1D) network; however, the another works as an input for the long short-term memory (LSTM) network [17].

In this work, two systems were developed based on word-based embedding features and character-based context features. For the character-based system, the same procedure as that of word based is done except that the vectors are character vectors. For the embedding to capture the word representation more effectively, additional data apart from the train and test data must be provided to the embedding model. The additional data used here is also a code-mixed Hindi–English social media data collected from other shared tasks. The input for the word embedding will be the train data and the additionally collected dataset. The embedding model generates the vector of each vocabulary (unique) word present in the data. Along with extracting the feature vectors of the train data, its context information is also extracted. The incorporation of the immediate left and right context features with the features of the current word is called 3-gram context appending. 5-gram features were also extracted, which is the extraction of features from two neighboring words before and after the current word. So if the vocabulary size of the training data is $|V|$, and the embedding feature size generated is 100 for each word, then after context appending with 3-gram features, a matrix of size $|V| \times 300$ is obtained. 5-gram appending will result in a matrix of size $|V| \times 500$. The test data was also given to the embedding models. The data were then appended with the 3-gram and 5-gram context information. Further, this data is used for training and testing using machine-learning techniques.

3.1 Word-Based Embedding Model

The word-based embedding model is used to find the feature vectors that are useful in predicting the neighboring tokens in a context. The feature vector for this model is generated using skip-gram architecture of popular Word2vec package proposed by Mikolov et al. [18]. Apart from the skip-gram model, another architecture continuous Bag of Words (cBoW) is also used [18].

Word2vec is a predictive model that is used to produce word embedding from raw text. It exists in two forms, the continuous Bag-of-Words model (cBoW) and the skip-gram model. Algorithmically, these two are similar, except that cBoW forecasts target words from source context words, whereas the skip-gram forecasts source context words from the target words. This gives the flexibility to use skip-gram when we are having a large dataset and one can use cBoW for the smaller dataset. We focused on the skip-gram model for language identification at word level in the multilingual domain to answer (word belongs to which language) in the rest of this paper. When the data is given to the skip-gram model, it maximizes the average log probability, which is formulated as in Eq. (1). In this equation, N is the total number of words in the train data and x is the context size. P is the SOFTMAX probability which is given using Eq. (2).

$$L = \frac{1}{N} \sum_{k=1}^{n} \sum_{-x \le i \le x} \log P(Tk + Ti|Tk) \tag{1}$$

$$P(Ti|Tk) = \frac{\exp(V'Tj(VTk))}{\sum_{w=1}^{w} \exp(V'Tj(VTk))} \tag{2}$$

where w is the vocabulary size; $P(T_j|T_k)$ is the probability of occurrence of the next word; and V' is the output vector representation. The dataset along with the additional dataset collected was given to the skip-gram model. The vector sizes to be generated were fixed as 100. The skip-gram model generates a vector of size 1×100 for each vocabulary word available in the dataset. From this, the vectors for the training data were extracted. The context appending features were then extracted from this file. The final training file for the classifier will consist of the tokens in the train data, their language tag and the 3-gram and 5-gram context feature vectors extracted. Thus, three training files are generated with $|V| \times 101$, $|V| \times 301$ and $|V| \times 501$ dimensions. The test data with its corresponding context appended vectors are fed to the classifier for testing the system.

3.2 Character-Based Embedding Model

The procedure for character embedding is the same as that of skip-gram-based word embedding. Each token in the train data gets split into characters and fed to the system. This will generate a vector for each character. The vector size to be generated was fixed as 100. The vectors generated for each character is used to create vectors for each token as Eq. (3).

$$Y = x + S_h(W, C_{t-k}, \ldots C_{t-k}, C) \tag{3}$$

In regard to above, equation SOFTMAX parameters are denoted by x and S where h is the embedding features of character and word. C is the character vectors and W is the word vectors. $C_{t-k}, \ldots C_{t+k}$ are the characters in the train data.

There are following constraints before applying the proposed algorithm

(1) Each document must consist of words from two languages.
(2) All the documents must be in a single script. The chosen script, in this case, is ROMAN Script.
(3) In the Indian scenario, code mixing is applicable between English and other Indian languages.
(4) The language used in the proposal is English and Hindi, where Hindi is represented using Roman, not Devanagari.

If the Hindi words are presented in Devanagari characters, it is then a simpler task to identify the language. This becomes non-trivial tasks to identify the language as both Hindi and English are written using the same character set.

4 Statistics of Dataset

The dataset used for this work is obtained from POS Tagging task for Hindi–English code-mixed social media text conducted by ICON 2016 [19]. The dataset contains the text of three social media platforms, namely Facebook, Twitter, and WhatsApp. The train data provided contains the tokens of the dataset with its corresponding language tag and POS tag (Table 1).

Proposed framework is tested on the Indian language bodies dataset for language recognition. Further, it is used in the FIRE 2014 and collective problem on transliterated exploration. Classifier is trained and tested on both the languages, i.e., Hindi and English words in the script based on Romanized behavior for Bollywood song

Table 1 Statistics of dataset ICON 2016

Data	Number of sentences		Number of tokens	
	Training data	Testing data	Training data	Testing data
Facebook	772	111	20615	2167
Twitter	1096	110	17311	2163
WhatsApp	763	219	3218	802

lyrics. Complete database of songs consists of 63,000 documents in the form of the text file.

5 Experiments and Analysis

This section discusses the complete experimental part along with results and consequent discussions. The proposed algorithm for retrieving language of the word in code-mixed data is evaluated on the basis of statistical measures and also evaluated using the machine-learning approach. This section provides a complete evaluation based on the statistical model. We performed two separate experiments on the code-mixed data to rationalize the performance of the language, and we have computed code-mixing patterns in the dataset on two metrics. This is being used to know the mixing patterns in the dataset. The proposed system is analyzed and evaluated on the basis of following code-mixing metrics.

MI: It is the abbreviation for multilingual index, which is used for counting the number of words that calculates the distribution variations of language tags in a corpus of languages. Equation 4 defines the multilingual index (MI) as:

$$\mathrm{MI} = \frac{1 - \sum P^2 J}{(k-1)\sum P^2 J} \tag{4}$$

where k denotes the total count of languages, P_j represents the total count of words in the language j over the number of words in the corpus. The value of MI resides between 0 and 1. Value of 0 relates monolingual corpus and 1 relates to the equal number of tokens from each language in a corpus.

CMI: Code-Mixing Index: At the phonetic level, this is calculated by discovering the most frequent language in the utterance and then counting the frequency of the words belonging to all other languages present. It is calculated using Eq. (5).

$$\mathrm{CMI} = \frac{\sum_{i=1}^{n}(wi) - \max(wi)}{n-u} \tag{5}$$

Table 2 MI and CMI values

Language type	MI	CMI
Hindi–English	0.582	22.229

where w_i is the sum of all languages present in the utterance, $\max\{w_i\}$ is the maximum number of words existing from any language, n denotes total number of tokens, and u denotes the number of tokens for other language-independent tags. If an utterance only contains u (i.e. $N = u$) language-independent tokens then its index is considered to be zero. For other utterances, we use the normalization (multiply the value by 100) to acquire the digits in the range of 0–100. The next w_i are the tagged language words and $\max(w_i)$ is the most prominent language words. Applying this equation, we will get CMI = 0 for monolingual utterances because max $(w_i = n - u)$.

To understand the model, consider the following scenario: sentence S_1 contains ten words. Five words are from Language L_1 and remaining 5 words are from Language L_2. Applying Eq. 5, the CMI will be 100 × (1–5/10) = 50. However, another sentence S_2 contains 10 words and each word is from a different language. The CMI = 100 × (1 − 1/10) = 90. It rightly reflects that S_2 is highly mixed as every word belongs to a different language. This CMI value helps us to understand the level of code mixing available in the dataset. Table 2 describes the values obtained for MI and CMI for the used corpus.

Secondly, we computed the similarity score based on the proposed algorithm on the dataset using Eq. (6). It gives significance in labeling the word as either English or Hindi based on the frequency of the word. The proposed algorithm checks the score of the classifier for Language L_j on input W_i as $0 \leq$ score ≤ 1, where score is similarity metrics, sim (W_x, W_y) x and y can be the word in a string. The below section describes the different results obtained on the code-mixed dataset for calculating the similarity score at word level and sentence level.

$$Score(x, y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 y_i^2}} \tag{6}$$

The dataset used for this work is obtained from POS Tagging task for Hindi–English code-mixed social media text conducted by ICON 2016 [19]. The dataset contains the text of three social media platforms, namely Facebook, Twitter and WhatsApp. We use the Hindi–English dataset for the experimental evaluation. The labels used are summarized in Table 3.

The training data contains the tokens of the dataset with its corresponding language tag and POS tag.

(1) **E**—indicating English words—example: This, and, there.
(2) **H**—indicating Hindi words—example: aisa, mera, tera.
(3) **NE**—indicating named entities like Person, Location, and Organization—example: Narendra Modi, India, Facebook.

Table 3 Description of the labels for Hindi–English dataset

Label	Description	Hindi–English %
E	English words only	57.76
H	Hindi words only	20.41
NE	Named entity	6.59
Other	Symbols, emoticons	14.8
Ambiguous	Can not determine whether Hindi or English	0.27
Mixed	Word of Hindi–English in combination	0.08
Unk	Unrecognized word	0.09

(4) **Other**—indicating tokens containing special characters and numbers—example: #, 0–9.
(5) **Ambiguous**—indicating words used ambiguously in Hindi and English—example: is, to, us.
(6) **Mixed**—indicating words of Hindi–English and Number combination—example: madamJi, Sirji.
(7) **Unk**—indicating unrecognized words—example: t.M, @. s, Ss.

All the seven tags are present in the Facebook dataset, where ‘E’, ‘H’, ‘NE’, ‘Other’ are the tags present in Twitter and WhatsApp data as tabulated in Table 3. For generating the embedding vectors, more dataset has to be provided to efficiently obtain the distributional similarity of the data. The additional dataset collected along with the training data will be given to the embedding model. The Hindi–English additional code-mixed data were collected from shared task on Mixed Script Information Retrieval (MSIR), conducted in year 2015 and 2016 [20] and shared task on Code-Mix Entity Extraction task conducted by Forum for Information Retrieval and Evaluation (FIRE), 2016. Most of the data collected for embedding is Hindi–English code-mixed Twitter data. The size of the dataset used for embedding is given in Table 4.

Context appending was done for each Facebook, Twitter, and WhatsApp train as well as test data. These were given to the learning model for training and testing. The cross-validation accuracies obtained for Facebook, Twitter, and WhatsApp with 3-gram and 5-gram features for character-based embedding model and word-based embedding model are presented in below section.

When comparing the overall accuracy obtained for Facebook, Twitter, and WhatsApp, we can see that the accuracy obtained is more with the word-based model

Table 4 Embedding dataset

Number of sentences in the dataset used for embedding (Facebook, Twitter, and WhatsApp)	
ICON 2016	2631
MSIR 2015	2700
MSIR 2016	6139

Table 5 F-measure obtained for Facebook dataset

Embedding type		E	H	NE
Character embedding	3 g	86.45	93.36	65.02
	5 g	85.47	92.55	65.05
Word embedding	3 g	86.99	93.51	67.21
	5 g	85.15	92.47	61.03

Table 6 F-measure obtained for Twitter dataset

Embedding type		E	H	NE
Character embedding	3 g	85.34	93.44	77.12
	5 g	85.38	93.49	80.27
Word embedding	3 g	85.71	93.97	83.94
	5 g	85.42	93.16	78.15

Table 7 F-measure obtained for WhatsApp dataset

Embedding type		E	H	NE
Character embedding	3 g	54.9	80.2	37.7
	5 g	54.3	80.9	31.5
Word embedding	3 g	60.8	81.9	40.2
	5 g	53.7	80.1	40.1

as compared to character-based embedding model. Tables 5, 6, and 7 show the performance of Facebook, Twitter, and WhatsApp Hindi–English code-mixed data. It is clear from these tables that F-measure for language labels E-English, H-Hindi, NE-Named Entity is better using word embedding.

Table 5 also shows that word-embedding 3-gram-based model gives a better score than other models. But Table 6 presents that word-embedding 5-gram-based gets better F-measure values. Similarly, Table 7 shows the results for WhatsApp datasets and clearly show that word-embedding 3-gram-based model gives a better score. This is because the system needs more context information to identify the language. Figure 2 shows F-measure values obtained for Facebook, Twitter and WhatsApp to represent different labels of text as E-English, H-Hindi and NE for Named entity.

6 Conclusion

The intricacy of language identification in code-mixed and code-switched data is governed by the following: data source, code switching and code-mixing manners, and the relation between the languages involved. We find that the code mixing is more

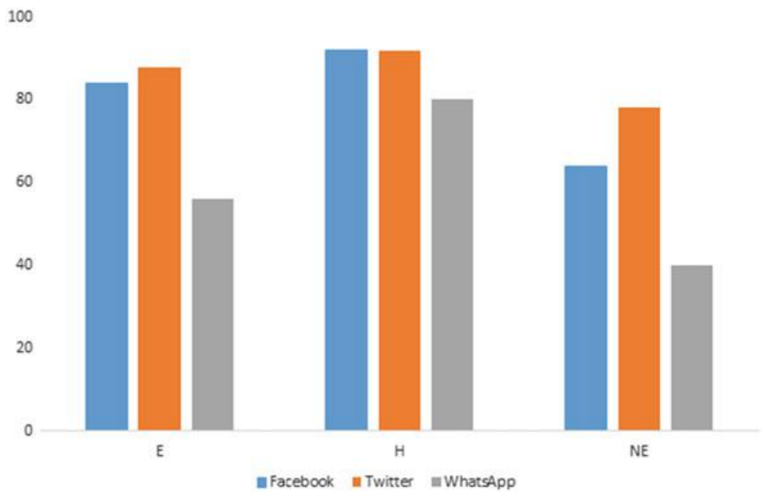


Fig. 2 F-measure for label E, H, and NE

used in social media context as per the evaluation and experiments were undertaken in this work. Code-mixing metrics helps in identifying the code-mixing patterns across language pairs. By analyzing the code-mixing metrics, we conclude that Hindi–English words are often mixed in our dataset. It would be a great idea to investigate the emerging trend of code switching and code mixing to bring conclusion about the behavioral patterns in the data of different sources like lyrics of songs, chat data having different language sets, blog data, and scripts of plays or movies. The results depict that the word embedding is capable to detect the language separation by identifying the origin of the word and correspondingly mapping to its language label.

References

1. Kim, Y.: Convolutional neural networks for sentence classification. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, pp. 1746–1751 (2014)
2. King, B., Abney, S.: Labeling the languages of words in mixed-language documents using weakly supervised methods. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1110–1119 (2013)
3. Nguyen, D., Dogruoz, A.S.: Word level language identification in online multilingual communication. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 857–862 (2013)
4. Das, A., Gamback, B.: Identifying languages at the word level in code-mixed Indian social media text. In: Proceedings of the 11th International Conference on Natural Language Processing, India, pp. 378–387(2014)

5. Sequiera, R., Choudhury, M., Gupta, P., Rosso, P., Kumar, S., Banerjee, S., Naskar, S., Bandyopadhyay, S., Chittaranjan, G., Das, A., Chakma, K.: Overview of FIRE-2015 shared task on mixed script information retrieval, vol. 1587, pp. 19–25 (2015)
6. Jhamtani, H., Bhogi, S.K., Raychoudhury, V.: Word-level language identification in bi-lingual code-switched texts. In: Proceedings of 28th Pacific Asia Conference on Language, Information and Computation, pp. 348–357 (2014)
7. Ethiraj, R., Shanmugam, S., Srinivasa, G., Sinha, N.: NELIS—named entity and language identification system: shared task system description. FIRE Workshop **1587**, 43–46 (2015)
8. Bhargava, R., Sharma, Y., Sharma, S.: Sentiment analysis for mixed script indic sentences. In: Proceedings of International Conference on Advances in Computing, Communications and Informatics, ICACCI, India, pp. 524–529 (2016)
9. Sharma, S., Srinivas, P., Balabantaray, R.: Emotion detection using online machine learning method and TLBO on mixed script. In: Language Resources and Evaluation Conference, vol. 10, no. 5, pp. 47–51 (2016)
10. Bali, K., Jatin, S., Choudhury, M., Vyas, Y.: I am borrowing ya mixing? An analysis of English-Hindi code mixing in Facebook. In: Proceedings of The First Workshop on Computational Approaches to Code Switching, EMNLP, pp. 116–126 (2014)
11. Rao, P.R.K., Devi, S.L.: CMEE-IL: code mix entity extraction in indian languages from social media Text@FIRE 2016—an overview. FIRE Workshop **1737**, 289–295 (2016)
12. Sapkal, K., Shrawankar, U.: Transliteration of secured SMS to Indian regional language. Proc. Comput. Sci. **78**, 748–755 (2016)
13. Zubiaga, A., Vicente, I.S., Gamallo, P., Pichel, J.R., Alegria, I., Aranberri, N., Ezeiza, A., Fresno, V.: TweetLID: a benchmark for tweet language identification. Lang. Resour. Eval. **50**(4), 729–766 (2015)
14. Alekseev, A., Nikolenko, S.: Word embedding for user profiling in online social networks. Computacion y Sistemas **21**(2), 203–226 (2017)
15. Chaudhary, J., Patel, A.C.: Bilingual machine translation using RNN based deep learning. Int. J. Sci. Res. Sci. Eng. Technol. **4**(4), 1480–1484 (2018)
16. Samuel, K.C.: Exploring language; some specificities, complexities and limitations in human communication and social interaction in multi-cultural contexts. Adv. J. Soc. Sci. **5**(1), 26–36 (2019)
17. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
18. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Proceedings of the 26th International Conference on Neural Information Processing System, vol. 2, pp. 3111–3119 (2013)
19. Jamatia, A., Das, A.: Task report: tool contest on POS tagging for code-mixed Indian social media (Facebook, Twitter, and WhatsApp) text. In: Proceedings of ICON (2016)
20. Banerjee, S., Chakma, K., Naskar, S., Das, A., Rosso, P., Bandyopadhyay, S., Choudhury, M.: Overview of the mixed script information retrieval (MSIR). In: CEUR Workshop Proceedings, vol. 1737, pp. 94–99 (2016)

Detecting Influencers in Social Networks Through Machine Learning Techniques

Rishabh Makhija, Syed Ali, and R. Jaya Krishna

Abstract The online social networks have given access to a new way of communication in society. Social networking platforms like Facebook, Instagram and Twitter provide different ways to connect and communicate with people around the world, bringing together the ideas from different parts of the world. In every social network, there would be people who would be influential and can influence other people to their idea. Hence, finding an influential person is very important and helps us to spread information more accurately and to more people. In this paper, we worked with machine learning techniques to identify the most influential nodes in the network, studied different methods to determine the best suitable for the network and understood how information cascading techniques can be applied.

Keywords Influencers · Machine learning · Triadic closure

1 Introduction

Online social networks became a bridge that connects our physical life with the wider world. For example, as of the second quarter of 2019, Facebook is the biggest social network with 2.41 billion monthly active users from all over the world. These networks produce large volumes of data, and its increasing popularity offers the opportunity to study and analyze the behavioral and interaction patterns the data shows.

R. Makhija (✉) · S. Ali
Department of Information Technology, Manipal University Jaipur, Jaipur, India

R. Jaya Krishna
Department of Computer Science, Manipal University Jaipur, Jaipur, India
e-mail: jaya.krishna@jaipur.manipal.edu

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_23

1.1 Graph Theory

Graphs are structures that can represent data visually. This helps us extract valuable information like relationships and allows us to take better data-driven decisions based on them. A graph can be represented as $G = (V, E)$ where V and E are sets of vertices and edges, respectively. Vertices are a set of objects or nodes that are connected together using the edges.

With the help of graphs, we can find influential nodes in a social network. Advertising companies and marketers can analyze and estimate the person who can market their products over their large social networks and allow larger amount of users to be influenced, in turn, increasing the profits.

1.2 Social Networking

Social networking is the use of social media Web sites over the Internet in order to connect with your friends, family, colleagues or even clients. In the last decade, a large number of social networking platforms like Facebook, Twitter, LinkedIn and Instagram have emerged and have population more than most countries. All such platforms have become a huge base for companies to create a client-base and engage customers. Companies aim to increase their brand recognition, loyalty and conversion rates. Social networking helps them to access and be recognizable to more new customers, thereby promoting the brand's name and content.

1.3 Influencers in a Social Network

Influence in any network can be termed as the amount of power required to control the decision of a user based on their position in their network. Thus, social influence occurs when one's opinions, emotions or behaviors are affected by others [1]. It can be used to spread messages widely and quickly in the network. In an online social network, influencers form large social spaces where people follow them in large numbers, thereby expanding their connections. Influencers are, however, present in any network that shows the relationships between the users. For example, we take a dataset of the 9/11 terrorist network containing the graph that shows the interactions among various members of the network. We create a method to extract the five most influential members of the terrorist network that has been discussed in the upcoming sections.

2 Background Overview

We aim to detect the influential nodes in a social network by using a number of concepts, namely triadic closure and various machine learning algorithms. A number of other research papers were referred to obtain the required concepts and resources in our research.

2.1 Triads and Triadic Closure

We can study social influence from a structure level, focusing on the triadic structure as is the simplest one in social networks as well as the cornerstone for studying network formation [2, 3]. A triad is a group of three nodes connected together. Basically, there are two types of triads, open and closed. When direction is taken into consideration, 16 such triads can be formed among three nodes. A triad is a very simple and stable structure whose features can be used to the advantage of social network analysis. Every triad follows the transitive property [4]. This can further be explained with a concept called triadic closure property which states that given three nodes X, Y and Z, if there exists a relationship between X–Y and X–Z, then there must exist a weak or a strong relationship between Y–Z. This property might not be true for every case in a large and complex network but can be helpful as a simple structure to understand networks and predict relationships.

In a referred paper [5], the authors aim to study the social influence from the perspective of triadic structures. The data sets used were one from a micro-blogging Web site, weibo.com and the other from CrossFire which is a shooting game. The method used was ordinary least squares, also known as linear regression. In their experiment, they show that there are as many as 30 kinds of triads possible considering that there are positive and negative neighbors and strong and weak relationships between any three users. The action considered was "retweeting." In the CrossFire dataset, the action considered was if we make free users play games with paying users, would they be influenced to turn into paying users as well. The model was trained using basic features first, and combinations of neighborhood features and triadic features were used. The results of their experiment show that when we add the triadic features in regression techniques, the predictive power increases significantly. Thus, this paper provided enough proof and knowledge from the experiments to believe that the triadic structures in any network can trigger influence among its neighbors and can prove to give out better results when used along with other features in the dataset.

2.2 *Machine Learning Techniques*

Machine learning is an application of artificial intelligence that provides the system with an ability to learn and improve from experience to predict patterns or make better decisions. In the paper referred above [5], it was observed machine learning can help in classifying the more influential nodes in a network, and when they are combined with the right feature sets can produce good results.

We aim to use various classification techniques in order to detect the more influential nodes in a given network.

Logistic Regression. This is a supervised machine learning classification algorithm used to analyze a dataset with one or more independent variables that determine the result. The target variable is categorical, and we use binomial logistic regression, which means the target values has two values, "yes" or "no" [6]. It predicts the probability of an event by fitting the data into a logistic function. The standard logistic function is called the sigmoid function:

$$S(x) = \frac{1}{1 + e^{-x}}$$

K-Nearest Neighbor. K-NN is a supervised machine learning classification algorithm used in pattern recognition and data mining. It can be used to predict both classification and regression problems [7]. Choosing the parameter K is very crucial and it depends on the problem in question. Even with its simple nature, it can give highly competitive results.

Decision Tree. A decision tree is a flowchart-like tree structure where a node represents a feature attribute and the branch represents a decision rule and each leaf node represents the result. It partitions the tree recursively on the basis of the attribute values. These values are based on the attribute selection measures (ASM). ID3 decision tree algorithm uses information gain. C4.5 uses an extension to information gain known as gain ratio to handle the issue of normalizing information gain.

Random Forest. Random forest is a type of supervised machine learning algorithm which can be used for both classification as well as regression producing great results most of the time due to its flexibility [8]. It works by building a "forest" which is a huge collection of multiple random decision trees and combines them together to get more accurate and stable predictions. It adds extra randomness to the model while making trees, by selecting most suitable features from a random subset of features. Random forest also provides an easy way to determine the relative importance of every feature by looking at which we can select which features to drop to avoid overfitting.

2.3 Information Cascading

Information cascading is a phenomenon in which a number of people make the same decision. Finding the influential node can help us in effective cascading of information. If we consider a two case scenario, each case might have a payoff load. This is just the weight factor by which that particular task is influential [9]. For example: If a class of 30 students, everyone has to attend a class. But few students plan to go for a movie. Here, the payoff load for movie would be more. If the students who plan are the most influential node, then the majority of the class would be going for the movie.

3 Methodologies

In order to achieve our aim of detecting influential nodes in a social network and cascading information to a wider range, we use various methodologies as discussed in the coming sub-sections.

3.1 Triadic Closure Analysis

Based on the theory mentioned in Sect. 2.1, we apply the triadic closure analysis on a terrorist network of 9/11. There are 16 different types of standardized triads, as shown in Fig. 1, that we work with for this experiment and the triadic closure analysis helps us to gather which node has which kinds of triads and how many in amount. From this result, we further use it to form a new table and apply probabilistic formula along with weights assigned to the type of triads and calculate an influence index for each node in the network. We sort the results to get the most influential nodes of the network and try to visualize the influential nodes and their reach in the network.

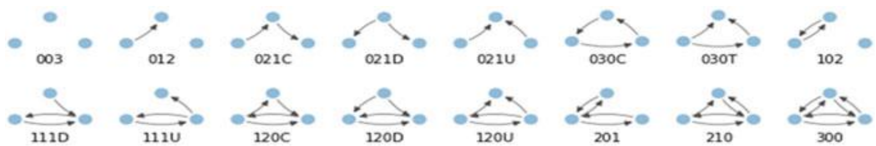


Fig. 1 Different types of triads

3.2 Spammer Detection Using Machine Learning

Based on the theory mentioned in Sect. 2.2, we explore an application of machine learning classification algorithms known as spam detection. We assume that any spammer is likely to have more followees than followers, as it has been observed by the social networks. Due to its ever-growing nature, social networking platforms also end up holding a lot of spam profiles or abandoned profiles that have not yet been removed from their databases. However, focusing on spam profiles, they tend to reflect similar characteristics to that of an influential profile. Thus, to increase the level of security in our research, we explore the detection of spammers in a dataset taken from a micro-blogging site weibo.com.

3.3 Detecting Influencers Using Machine Learning

Based on the theory mentioned in Sect. 2.2, we aim to detect which nodes in a network can be more influential than others. Using a Twitter dataset, we try to classify between two users, which is more influential than the other based on the feature attributes available. To select the most suitable features, we use a forward-backward stepwise selection based on *p*-values of coefficients of linear regression. Next, we create a model function that could train, test, validate and present the results of the various algorithms used. When the dataset along with the classifier object of the algorithm is passed into the function, it undergoes the following steps:

- Split the dataset with splitting percentage = 70% training and 30% test sets
- Train and fit the algorithm
- Make predictions on test dataset
- Draw receiver operating characteristic (ROC) curve
- Cross-validation analysis based on accuracy score and AUC score.

We use logistic regression as our base algorithm and implement it by passing the classifier object and dataset to the model function. We next apply k-NN to classify the more influential nodes, while keeping the value of k as 5. Similarly, we apply decision tree classifiers based on splitting criteria as entropy and Gini index. We also apply random forest classifier keeping its splitting criteria as entropy. Finally, we make a comparison as to which algorithm fared better in classifying the more influential nodes in the dataset. The results obtained by each algorithm show the ROC curve along with other metrics that help us to decide which algorithm classified the problem best.

4 Implementation and Results

Implementations of all our experiments have been carried out in python language on Jupyter notebook.

4.1 Spam Detection on Weibo Dataset

Logistic regression and k-nearest neighbors are used to detect whether a user in a dataset is a spammer or not. The dataset has seven attributes. After analyzing the dataset, and using the graphs in Fig. 2, we make some observations:

- Spammers are more likely to have a lower class than a normal user. The highest frequency of spammers occurs in class 3, followed by class 2 and 1. Therefore, the number of followees in class 3 is higher than any other class. There are no spammers present in class 10 or above.
- Spammer profiles are higher in frequency with the female gender than male gender. This mildly suggests that males create female profiles in order to spam posts.

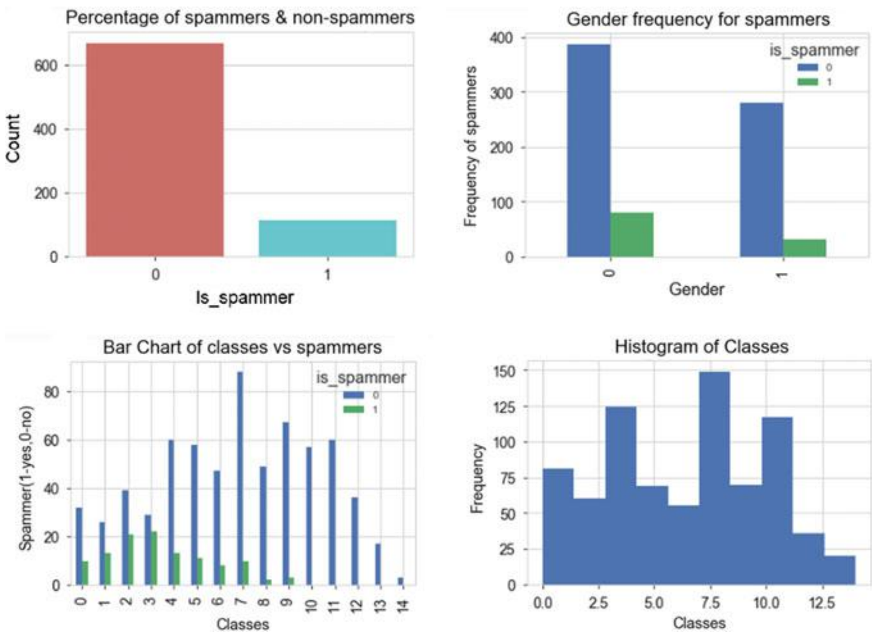


Fig. 2 Weibo dataset visualization

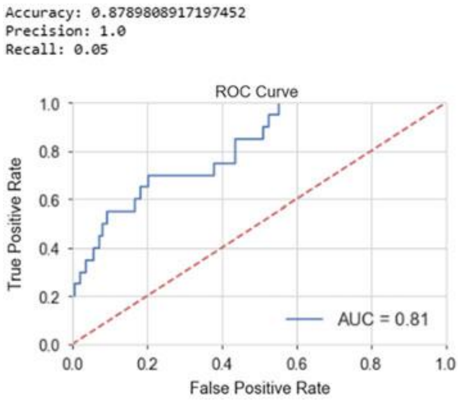


Fig. 3 Logistic regression

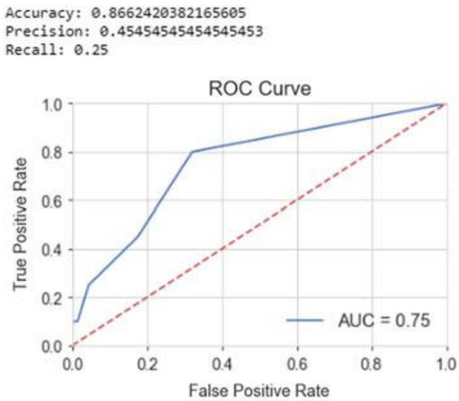


Fig. 4 K-NN algorithm using k = 5

Using gender, number of posts, number of followers and the class the user belongs to as the feature set, through logistic and K-NN algorithms, we compare the results as shown in Figs. 3 and 4.

4.2 Triadic Closure Analysis on 9/11 Dataset

We implemented triadic closure analysis on a data of a terrorist network of 9/11. The dataset had an edge-list with weights assigned based on the level of interactions the nodes had. A second file contained the 'flight' attribute that denotes to which flight was the node related to. We perform the triadic closure on the all the nodes that had the 'flight' attribute and calculate the frequencies of each the 16 triads present in Fig. 4. It is observed that three types of triads, namely 102, 201 and 300 occur in maximum frequencies and an influence index is calculated using the formula below:

Top 5 influencers in the network :

	Node	t-102	t-201	t-300	Influencer-Index
7	Fayez Ahmed	399	42	17	0.557625
19	Ramzi Bin al-Shibh	379	42	11	0.419650
18	Zakariya Essabar	243	31	10	0.347326
4	Lotfi Raissi	237	49	7	0.330964
16	Wail Alshehri	245	33	9	0.330822

Fig. 5 Detection of influencers in 9/11 dataset

$$\mathrm{InfluenceIndex}_i = \sum_{i=1}^{n} \frac{D_i['102']}{D['102']} w['102'] \Bigg] + \frac{D_i['201']}{D['201']} w['102'] \Bigg] + \frac{D_i['300']}{D['300']} w['300']$$

Using this formula while keeping a threshold as 0.2 and calculating the influence index for each node in the network, we can detect the influencers in the network. Figure 5 shows the five most influential nodes extracted from the network.

4.3 Detecting More Influential Nodes in Twitter Dataset

We made use of various machine learning techniques to detect the more influential node among two users, given in a large twitter dataset with 23 attributes. We applied a forward–backward feature selection method based on the *p*-values. Using the resultant feature set, we applied the following machine learning classifiers available on python’s sklearn library, and the results for the same have been shown in Fig. 6.

Among the classification results produced by these five algorithms, random forest outperforms them all. Even though the results do not indicate significant differences, but in a large dataset, the minutest of difference in performances can cause a great deal. However, the random forest does take considerably more amount of processing time as compared to other classifiers, but since our problem statement does not require a real-time analysis of data, and the random forest turns out to be the better-suited algorithm for detecting and classifying the more influential nodes in a dataset.

4.4 Cascading in a Network

For understanding cascading, we would be requiring a directed graph. For this, we are making use of a very common undirected graph, i.e., The Karate club. We have just considered two different activities taking place in the network—Action A and Action B. We assume that initially the complete network is performing activity ‘B’. Let the payoff (A) = a = 6 and payoff (B) = b = 3. By self-inputting few nodes to be working on Activity A. Influence in the node is shown in Figs. 7 and 8.

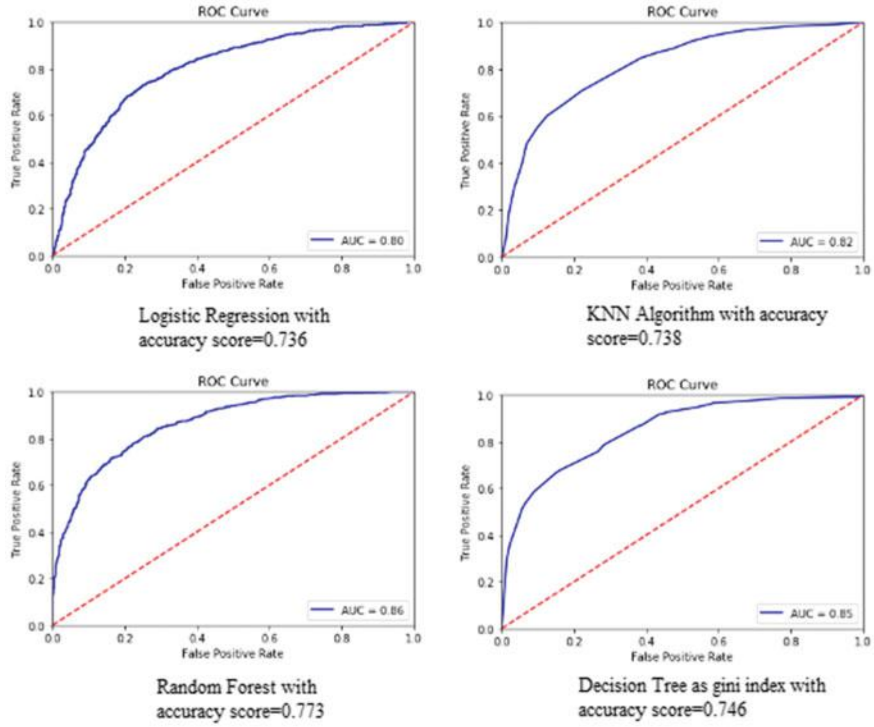


Fig. 6 Results

Fig. 7 Initializing with nodes [0, 2, 3, 4]

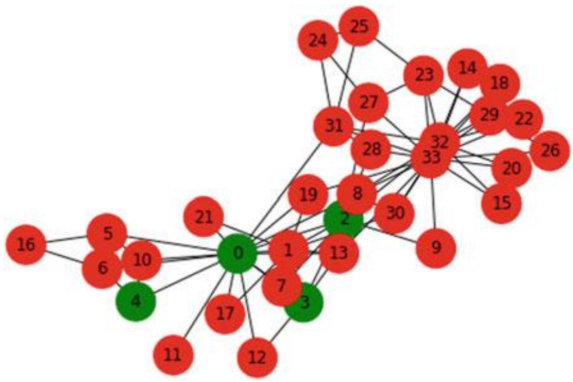
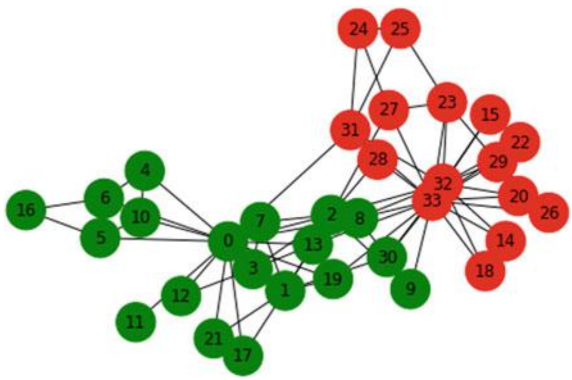


Fig. 8 After cascading



5 Future Work and Conclusion

In terms of related work in the future, we identify the following areas:

- Use of parallelization to try and reduce the computation time.
- Evolution of possible generic laws to govern the formation of networks.
- Work on the centrality methods and understand cascading after k-shell decomposition.

In our research, we explore various methods that can be used to detect influencers in a social network. While implementing triadic closure on a network that is not typically social, and using a probabilistic factor for calculating metrics, we extract the influencers with large interconnectivity in the network. We also use various machine learning techniques to achieve our aim of detecting influencers, and in a comparative analysis conclude that random forest is best suited when it comes to classifying the dataset of a social network that is full of people engaging in random actions. Our experimental results confirm that random forest is a great choice for solving such a classification problem, especially in terms of performance.

We also worked on a cascading method to understand how cascading works in a network, how influential nodes affect the decision and how payoff affect the influence.

References

1. Deutsch, M., Gerard, H.B.: A study of normative and informational social influences upon individual judgment. J. Abnorm. Soc. Psychol. **51**, 629–636 (1955)
2. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: simple building blocks of complex networks. Science 824–827 (2002)
3. Huang, H., Tang, J., Liu, L., Luo, J., Fu, X.: Triadic closure pattern analysis and prediction in social networks. IEEE TKDE **27**, 3374–3389 (2015)
4. Easley, D., Kleinberg, J.: Networks, crowds and markets: reasoning about a highly connected world. Cornell, NY, Cambridge Univ Pr. (2010)

5. Zhang, C., Gao, S., Tang, J.: Learning triadic influence in large social networks. In: IEEE/ACM International Conference on ASONAM (2016)
6. Internet for Logistic Regression. https://towardsdatascience.com/introduction-to-logistic-regression-66248243c148
7. Internet for k-NN and Decision tree. https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623
8. Cutler, A., Cutler, D.R., Stevens, J.R.: Random forests. In: Zhang, C., Ma, Y. (eds.) Ensemble Machine Learning. Springer, Boston, MA (2012)
9. Xu, W., Wu, W., Fan, L., Lu, Z., Du, D.Z.: Influence diffusion in social networks. In: Rassias, T., Floudas, C., Butenko, S. (eds.) Optimization in Science and Engineering. Springer, New York, NY (2014)

Application and Analysis of K-Means Algorithms on a Decision Support Framework for Municipal Solid Waste Management

Narendra Sharma, Ratnesh Litoriya, and Akhilesh Sharma

Abstract Many countries are facing the problem of sustainable solid waste management for a long time. The quantity of generating solid waste increase varies quickly as the rate of the population grows. Now, presently the government has focused on all waste segmentation levels like waste generated evolves and the technologies available to collect and process it. Many new advanced analytical techniques have been proposed for dealing with sustainable solid waste management problems. These advanced tools help the organizations to analyze municipal solid waste raw data, which is stored by local municipal bodies. We have used many innovative analytical tools and technology, such as data mining technology. On these analytical technologies, the clustering data classification technique is one of them. This paper focuses on K-mean clustering algorithms and their working in context with solid waste management. The analysis algorithms have been played an important role in decision making process. This type of algorithm helps the main decision-makers to make the right decisions or helps to improve the efficiency of the decision-making process. An excellent solid waste management framework must include all disturbing factors, including pollution generation, land use, energy use, financial costs, labor needs, monitoring, recycling rates, etc.

Keywords Municipal solid waste management · Data mining · Clustering · K-means · Decision support systems · SPSS Modeler

N. Sharma (✉)
Department of Computer Science, Mewar University, Chittorgarh, India
e-mail: narendra_sharma88@yahoo.com

R. Litoriya
Department of Computer Science & Engineering, Jaypee University of Engineering & Technology, Guna, India
e-mail: litoriya.ratnesh@gmail.com

A. Sharma
Department of IT, Manipal University Jaipur, Jaipur, India
e-mail: akhilesh.shm@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_24

1 Introduction

In population, India is the second-leading country of the world, and its populace of more than 1.35 billion is representing about more than 18% of the world's human population. In urban areas, the people of major cities raised at the rate of 31.8% during the last 10–20 years are 377 million, which is higher than the entire population of many small countries of the world.

A well-designed municipal solid waste (MSW) disposal framework or systems is very supportive for not only the human beings but also environment. Unsuitable solid waste management is very detrimental to public health, causes environmental pollution, dramatically impacts the quality of life of human beings, accelerates natural resources degradation, and causes climate change [1]. The elderly population who lives in isolation needs more care for waste collection and management [2, 3].

An excellent solid waste management framework must include all essential factors like pollution generation, land use, energy use, financial cost factors, labor needs, monitoring in different levels, waste recycling rates, etc. New advanced analytical tools and technology will perform an essential role in the design and development of an advanced municipal solid waste management framework. Data mining technology provides many types of data analysis facilities with the help of different– different algorithms and tools [4–6] that can increase the productivity of the waste management decision-making process. Many researchers define data mining in different–different ways. We can define that data mining is the very most straightforward way to it and is a process used by organizations to analyze raw data sets with using algorithms and tools for generating useful information.

Based on these patterns, organizations get information about their customer's behavior; it is very beneficial for making new marketing strategies, increases the sales of their products, and decrease costs. In the data mining process, all factors like data collection, warehousing, and data analyzing will be included [7]. The availability of various influencing factors of a particular activity leads the process of decision making more complex, which is eventually solved by existing factor interrelationship models or MCDM techniques and using fuzzy-based expert opinion techniques [8–12].

Data mining provides the facility to analyze and evaluate large blocks or sets of information to meaningful patterns. Organizations can use these patterns for different–different purposes, such as database marketing, fraud detection in banking or other sectors, and knowing customer behavior.

Working in data mining is to collect data from various sources and stored data sets in specific database management. These data sets are stored and managed from data warehouses; if the organizations want to report of their company growth and to know about the performance of any particular product and the cost analysis of different–different segments, all these types of activities are performed by various analytical algorithms and tools. Some most usable data mining techniques used for the analysis of data are given below:

I. Statistics: It is based on mathematical formulas and relations, the use of statistics technique to the collection of data sets, categorization of these data, and explanation of data. Some analysts do not consider this technique to a complete data mining technique. But in many research, it is advantageous to discover the patterns and build analytical models [13].
II. Clustering: It is one of the most functional and common classification technique of data mining. In this method, we can be identifying similar behavior data sets that comprise identical properties to each other. Clustering helps the analyst to recognize the characteristics of data sets and which types, and it can be helpful for organizations.
III. Visualization: It is also a very advantageous data mining process. This technique converts easily raw data sets into useful data patterns. We can predict the results in these methods very quickly.
IV. Decision Tree: This technique is based on classification techniques. Predictive models are used for generating a decision tree that looks like a binary tree. In this method, each part of the tree is seen as a hub, and the leaves of the trees are estimated as parcels of the data set identified with that specific arrangement. Decision tree technique was used utilized for information pre-handling, forecast work—also, investigation examination, and other useful analytical works.
V. Association Rules: It helps to find the association between two or more items. With the help of this technique, we anticipate the connection between the various factors in databases. This technic is very useful to find out the unknown patterns in the data sets, which is used in the classification process of the variables and looks with the highest frequencies.
VI. Neural Networks: Individual peoples and organizations have utilized neural network nowadays quickly. It is a significant data mining system used in different kinds of research. It is being used in the primary phases of the data mining innovation. Neural systems are elementary to utilize, provided that the client needs to take a shot at these calculations using algorithms. It need not bother with a lot of learning about the work or database.
VII. Classification: It is the most common and usable technique of data mining. In this technique, we have used some pre-classified data samples to create a model which can be able to classify the large samples of data sets for classifications. Classification beginning significant data of information and metadata (information about information). In this procedure, cluster analysis is performed, and this technique is beneficial for the decision tree or neural network framework [14].

1.1 Working of Clustering Algorithms

This technique is very beneficial for the grouping of like knowing about urban populations, by using geographic location, house type, and other parameters. For information discovery on Web data, this is also very useful. This algorithm divides into two categories, the first category is unsupervised, and the other one is supervised. An unsupervised type algorithm is K-means clustering. It is appropriate for uncategorized data [15].

The main aim of this algorithm is to discover clusters contain similar property, the data sets, number of clusters specified by the variable K. In this algorithm, every data point has allocated iteratively to specific K groups built on the categories that are provided. We know that data points or clusters are formulated based on feature relationships. The analyst have used these algorithms to increase the efficiency of business traditions, for making new business strategies about which types of clusters find out or categorize in unidentified groups or composite data sets. Once the algorithm has been functioned effectively, the clusters are modeled, and it can be shown the valuable information in clusters [16].

2 Literature Review

The use of data mining tools and their associated algorithms are widely used in the past for developing effective decision support frameworks. The utility of data mining applications has now become handy and compact with the availability of the Internet and a large number of Web and mobile apps [17–21]. Singh and Gill [22] have represented their research work in 2013. In this research, they have described and compared the performance of the K-means algorithm. Through this paper, they have to try to uncertain confines of the K-means algorithm by the proposed algorithm. In this research, they have worked about the time taken by K-means on an extensive database and introduced a new clustering algorithm that is efficient compared to the present algorithm.

Oyelade et al. [23] have been doing the work to predict the student's academic performance using the K-means algorithm in 2010. Their research is based on cluster analysis of educational data sets of students. In this paper, they have implemented the K-means clustering algorithm for investigating students' results data. The model was joined with the deterministic model to dissect the understudies' aftereffects of a private Institution in % criteria, which is a decent benchmark to screen the movement of scholastic execution of understudies in higher institution to settle on a viable choice by the scholarly organizers. To develop a suitable method for mobile app development, considerable work has been carried out [24].

Yadav and Sharma [25] presented their research work in 2013, and they have studied the review of the performance of the K-means algorithm. In this paper, they work on the performance of three different modified K-mean algorithms, try to eliminate the limitation of the K-mean algorithm, and increase the productivity of the K-means algorithm. Ali and Kadhum have done the research work of pattern recognition using K-means algorithms. Firstly, they have a focus on the primary working of clustering algorithms on spatial data analysis and image processing. The paper comprises the algorithm and its implementation and focuses on how to use data mining algorithms in pattern recognition. The literature survey reveals that data mining found to be a useful technique for many applications.

3 Role of Data Mining Technique in Decision Support Systems

In Decision Support Systems:-For an accurate decision-making process, the managers require knowledge. In the case of large data amounts, problems may arise in data analysis and to generate essential knowledge patterns. For analyzing the massive data sets, we have performed some analytical tools or techniques for creating useful information, and the entire procedure is called the data mining technique [24].

The data mining tools are very beneficial for discovering knowledge from the large data sets stored in data warehouse and other operative databases. So, we can say that information mining assumes an outstanding job in helping associations to perceive their client's conduct, care, significant customers, deal strategies and stock desires. The central persistence of data mining techniques to analyze large data sets related to different–different fields generates meaningful knowledge or information.

Data mining techniques deal with all standard statistical calculation, from artificial intelligence, and database administration. They are not a substitute for old analytical methods, but the addition of graphical and advanced statistical techniques. Data mining uses various types of classification, mathematical algorithms, fuzzy logic, shape recognition, machine learning, neural networks, genetic algorithms, and data viewing [25].

4 Methodology

For the development of a decision support system, we need to work with various advanced data mining techniques and algorithms such as machine learning, clustering, and other algorithms simultaneously. Depending on the requirements identified in the analysis phase, with the help of all of these technologies, it can be conducive for design reliable decision support system architecture. In this paper, we have studied the K-means clustering approach to decision support systems and data mining. We know that a sound decision support system is beneficial for achieving the targets or organizations' growth. For the development of a decision support system, we

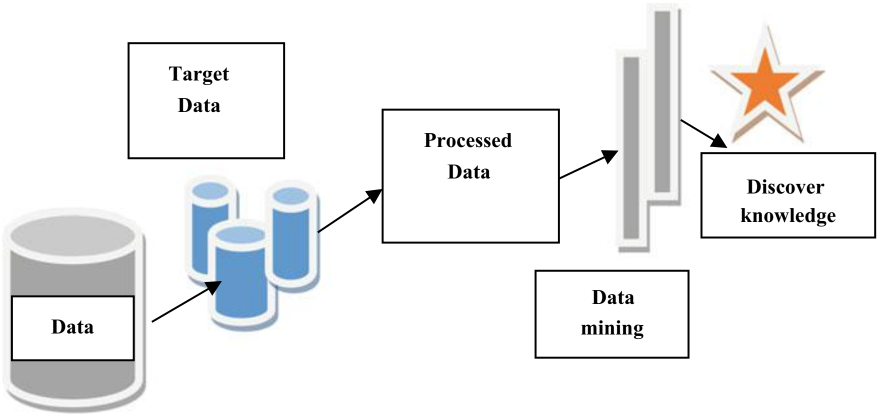


Fig. 1 Basic data mining process model

need to work with various data mining, machine learning, and artificial intelligence algorithms simultaneously.

K-means clustering algorithm provides the facility to identify the same group data, or it can make clusters of those data sets, which have to contain the same behaviors.

4.1 Data sets

We have collected the data from Indian government Web repository. This repository includes all area's database. Other data we have collected from a municipal office deals with solid waste management. These data sets are directly applied to analysis tools, and it generates the result (Fig. 1).

5 Data Analysis and Result

In this first step, we initialize a framework for uploading data sets. In the second step, it can gather information about the attributes. In the next level, it will perform the attribute selection process. Once attributes are selected, we will implement K-means algorithms for making a cluster. We have used the SPSS Modeler tools for data analysis, which provides a vast range of different–different algorithms for data analysis (Table 1).

The application of K-means clustering algorithms is shown as a flowchart. In the SPSS Modeler, firstly, we upload the data set from the database. After that, we analyze the data sets. Figure 2 has shown the clustering analysis of the given data. For the analysis purpose, various types of data set are available. For this research, we

Table 1 Solid waste data sets

Organic	Carbon	Hydrogen	Oxygen	Nitrogen	Sulfur	Ash
Paper	43.5	6	44	0.3	0.2	6
Cardboard	44	5.9	44.6	0.3	0.2	5
Plastics	60	7.2	22.8	0	0	10
Textiles	55	6.6	31.2	4.6	0.15	2.5
Rubber	78	10	0	2	0	10
Leather	60	8	11.6	10	0.4	10
Yard wastes	49.5	6	42.7	0.2	0.1	1.5
Glass	0.5	0.1	0.4	0.1	0	98.9
Metals	4.5	0.6	4.3	0.1	0	90.5
Dirt	26.3	3	2	0.5	0.2	68

https://data.gov.in/

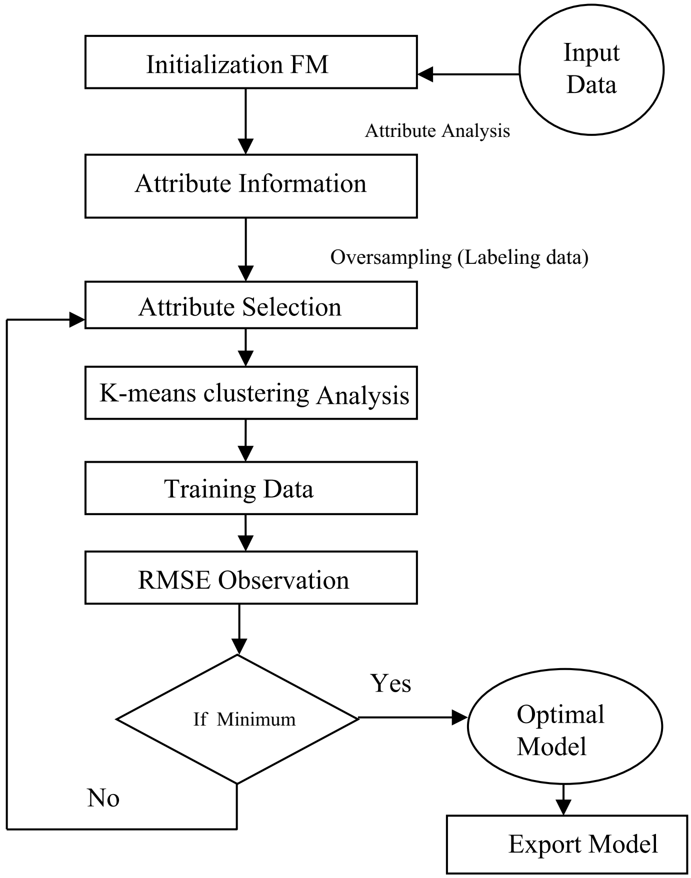


Fig. 2 A data flow diagram of clustering algorithms

have taken an exclusive data set that contains the percentage of different–different organic components for various kinds of solid waste materials (Fig. 3).

We have used IBM® SPSS® Modeler for the clustering analysis of raw data sets. It is the most famous data mining algorithm that provides the facility to make a cluster of similar data sets or grouping the data. It gives the details of many inputs and the quality of clusters. Figure 2 shows the cluster's size and number of iterations performed by the algorithms. We can get the information about all input factors, which have used to make clusters. We can see the number of clusters formed in the data set in Fig. 4. With the help of this analysis, we get information about various transformations performed on given inputs. We also get information about the predictor importance of organic components.

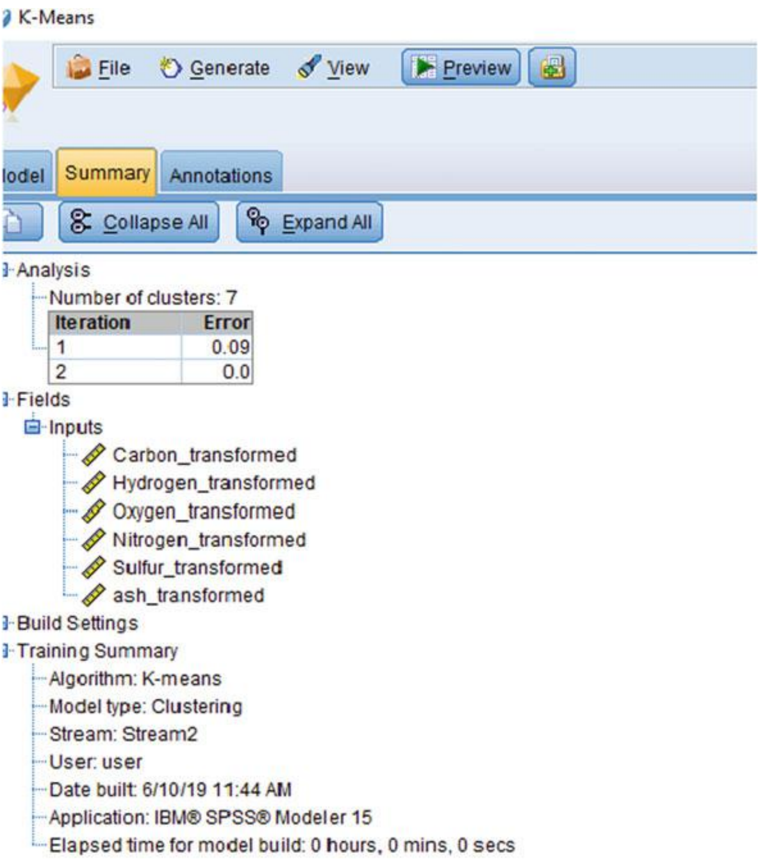


Fig. 3 Summary of solid waste data sets analysis

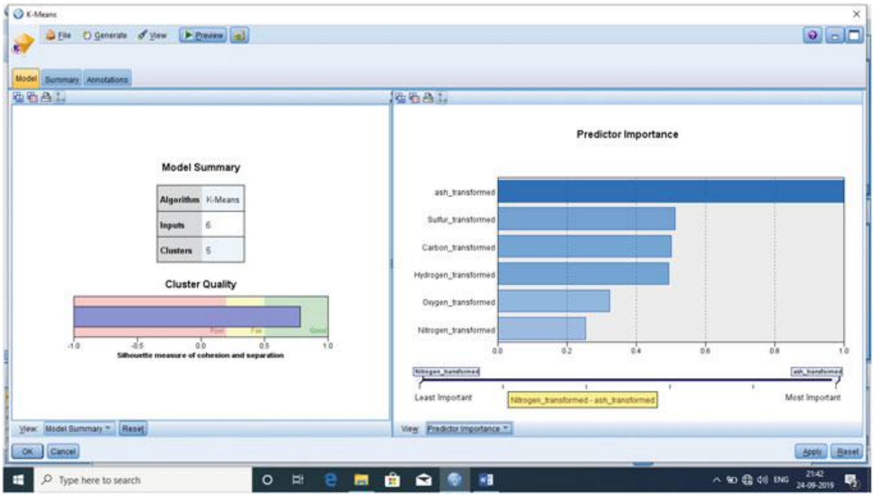


Fig. 4 Cluster summary and prediction importance generated by the K-means algorithm

6 Conclusion and Future

Clustering and other data mining procedures are very predictable for any analytical work. With the help of analysis, work organization knows about consumer behavior; it can be constructive for increasing the growth of the business. Now that time, data mining is used in every field. In our research work, for refining the quality of solid waste management process, an integrated decision support system (DSS) is advantageous, it can be very supportive for improving the efficiency of the decision-making process, and all advanced DSS can work with the cooperation of technologies like data mining algorithms. In this paper, we used K-means clustering algorithms for data analysis. Paper uses SPSS Modeler to provide a variety of modeling algorithms and also work with much artificial intelligence, statistics, and machine learning algorithms. In future work, we will try to use new data mining algorithms of the same data sets and will predict new patterns or useful information. These patterns can help solve the problems of human being lives.

References

1. Mladenic, S., Lavrač, D., Bohanec, N., Moyle, M.: Data Mining and Decision Support, 1st edn. Springer, USA (2003)
2. Pandey, P., Litoriya, R.: An activity vigilance system for elderly based on fuzzy probability transformations. J. Intell. Fuzzy Syst. **36**(3), 2481–2494 (2019)
3. Pandey, P., Litoriya, R.: Elderly care through unusual behavior detection: a disaster management approach using IoT and intelligence. IBM J. Res. Dev. 1–1 (2019)

4. Sharma, N., Bajpai, A., Litoriya, R.: Comparison the various clustering algorithms of weka tools. Int. J. Emerg. Technol. Adv. Eng. **2**(5), 73–80 (2012)
5. Sharma, N., Litoriya, R.: Incorporating data mining techniques on software cost estimation: validation and improvement. Int. J. Emerg. Technol. Adv. Eng. **2**(3), 301–309 (2012)
6. Sharma, N.: An analytical study on the importance of data mining for designing. J. Harmon. Res. (JOHR) **7**(2), 44–48 (2019)
7. Heinrichs, J.H., Lim, J.-S.: Integrating web-based data mining tools with business models for knowledge management. Decis. Support Syst. **35**(1), 103–112 (2003)
8. Pandey, M., Litoriya, R., Pandey, P.: Application of fuzzy DEMATEL approach in analyzing mobile app issues. Program. Comput. Softw. **45**(5), 268–287 (2019)
9. Pandey, M.,Litoriya, R., Pandey, P.: Identifying causal relationships in mobile app issues: an interval type-2 fuzzy DEMATEL approach. Wirel. Pers. Commun. 1–28 (2019)
10. Pandey, M., Litoriya, R., Pandey, P.: An ISM approach for modeling the issues and factors of mobile app development. Int. J. Softw. Eng. Knowl. Eng. **28**(07), 937–953 (2018)
11. Pandey, P., Litoriya, R., Tiwari, A.: A framework for fuzzy modelling in agricultural diagnostics. J. Eur. des Systèmes Autom. **51**, 203–223 (2018)
12. Pandey, M., Litoriya, R., Pandey, P.: Novel approach for mobile based app development incorporating MAAF. Wirel. Pers. Commun. (2019)
13. Chen, C.K.: Decision support system for tourism development: system dynamics approach. J. Comput. Inf. Syst. **45**(1), 104–112 (2004)
14. Rok, K., Matjaz, R., Marjan, K.: Integrating data mining and decision support through data mining based decision support system. J. Comput. Inf. Syst. **47**(3), 89–104 (2007)
15. Kantardzic, M.: Data Mining: Concepts, Models, Methods, and Algorithms. Wiley-IEEE Press (2011)
16. Ng, R., Han, J.: Efficient and effective clustering methods for spatial data mining. In: Proceedings of VLDB-94, pp. 144–155 (1994)
17. Litoriya, R., Kothari, A.: An efficient approach for agile web based project estimation: AgileMOW. J. Softw. Eng. Appl. **06**(06), 297–303 (2013)
18. Litoriya, R., Sharma, N., Kothari, A.: Incorporating cost driver substitution to improve the effort using Agile COCOMO II. In: 2012 CSI Sixth International Conference on Software Engineering (CONSEG), pp. 1–7 (2012)
19. Litoriya, R., Kothari, A.: Cost estimation of web projects in context with agile paradigm: improvements and validation. Int. J. Softw. Eng. **6**(2), 91–114 (2013)
20. Pandey, M., Litoriya, R., Pandey, P.: Perception-based classification of mobile apps: a critical review. In: Luhach, A.K., Hawari, K.B.G., Mihai, I.C., Hsiung, P.-A., Mishra, R.B. (eds.) Smart Computational Strategies: Theoretical and Practical Aspects, pp. 121–133. Springer, Singapore (2019)
21. Pandey, M., Ratnesh, L., Pandey, P.: Validation of existing software effort estimation techniques in context with mobile software applications. Wirel. Pers. Commun. (2019)
22. Singh, N.S., Gill, S.: Analysis and study of K-means clustering algorithm. Int. J. Eng. Res. Technol. **2**(7), 2546–2551 (2013)
23. Oyelade, O., Oyelade, J.: Application of k-means clustering algorithm for prediction of students' academic performance. Int. J. Comput. Sci. Inf. Secur. **7**(1), 292–295 (2010)
24. Pandey, M., Litoriya, R., Pandey, P.: Mobile app development based on agility function. Ingénierie des systèmes d'information RSTI série ISI **23**(6), 19–44 (2018)
25. Yadav, M., Sharma, J.: A review of K-mean algorithm. Int. J. Eng. Trends Technol. **4**(7) (2013)

Android Rogue Application Detection Using Image Resemblance and Reduced LDA

Saket Acharya, Umashankar Rawat, and Roheet Bhatnagar

Abstract Nowadays, the expanding diffusion of Android phones along with the substantial usage of mobile applications is increasing the malware production. Among various malware threats, the rogue applications have expanded their growth in the field of smartphones, especially Android phones. This paper presents an optimal methodology to detect and classify rogue applications using image resemblance and opcode sequence reduction. First, the opcode sequences are extracted, and then, they are converted into gray images. After this, Linear Discriminant Analysis (LDA) is applied in two stages. LDA is a supervised probabilistic method that is used for class separation and size reduction. In the first stage, the image sizes are reduced by selecting only the optimal features using LDA. The main objective of this stage is to increase the accuracy rate by reducing the size of opcode sequences. In the next stage, LDA is applied to test and train the dataset samples for separating rogue and benign apps. The experimental results on the rogue application families and unknown rogue apps show that the proposed methodology is efficiently able to identify rogue apps with an accuracy rate of 96.5%.

Keywords Machine learning · Image similarity · Classification · Android malware · Rogue application · Android security · Linear discriminant analysis

1 Introduction

Today, smartphones have turned out to be a multitasking device rather than simply a communication device. In other words, people carry smartphones not only for communicating with each other but they also store their personal and confidential data in their phones. Android mobile phones are becoming popular day by day due to the availability of huge number of Android applications for almost each and every category like music, gaming, office apps, location-based apps, chatting apps and so

S. Acharya (✉) · U. Rawat · R. Bhatnagar
Manipal University Jaipur, Jaipur, India
e-mail: saket.acharya@jaipur.manipal.edu

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_25

on. Hence, the Android phones have become a good motivation for the attackers to target these devices. Though there are several types of attacks possible on Android devices, the most recent and concerned security threat in the area of smartphones is rogue application. Rogue application, as the name implies, is a malware type that misleads or deceives the Android users. These malicious applications pretend to be a good app to remove device malware, but they steal crucial user data like login credentials, private messages and photos, etc, without the consent of the user. Also, the rogue apps sometime open pornographic advertisements in the browser and display malicious ads on the mobile screen.

In general, there are two approaches to detect Android malware: static approach and dynamic approach [1, 11]. The static methods focus on reverse engineering, i.e., analyzing source codes and resources, whereas the dynamic methods try to analyze the app behavior by running the app in a sealed Sandbox environment. There are not much studies done on Android rogue app detection. Therefore, identifying Android rogue app is a new area of study and has a vast future scope.

This paper aims to identify an optimal method for detecting Android rogue application using image similarity and dimension reduction. The Android APK files are first transformed to gray-scale images, and then, their dimensions are reduced. The proposed method is capable of identifying rogue applications with 96.5% accuracy.

2 Related Work

The researchers have done a lot of research work in detecting malware. Most of the studies have suggested three methods to detect Android malware: static, hybrid and dynamic methods. Static and dynamic methods are briefed above. Hybrid methods combine the features of static and dynamic techniques. Besides, studies are also done in detecting and classifying Android malware using machine learning algorithms. The related studies are mentioned as follows:

Khan and Sahs [6] classified malwares with the help of Control Flow Graphs (CFG) and permissions. They used SVM algorithm for classification. Yerima et al. [10] obtained features like APIs, native code from AndroidManifest, XML file, user- and kernel-level permissions and then calculated entropy information to classify the features. They used Bayesian classification algorithm for classification. Jerome et al. [2] obtained the opcode sequences from .dex files and convert them into N-gram binary sequences. These sequences were used to extract important features. The authors used SVM algorithm for classification. Their study showed that the false negative samples contain threats for the system.

Wei et al. [7] proposed a machine learning-based tool named Androidetect to detect Android malwares. The authors identified the relationship between critical permissions, system components and critical APIs. Then, they constructed Eigen vectors using the above relationship, and finally, they compared the proposed method with the machine learning algorithms: NaÏve Bayes, random forest, CART and J48 decision tree. They concluded that Androidetect performs faster and has better accuracy.

Xiaoyan et al. [8] obtained permissions from AndroidManifest file and put them in a binary sequence vector. The authors used Principle Component Analysis (PCA) to identify the best features and linear SVM algorithm for the classification. Further, they compared the outputs with other machine learning classifier algorithms like NaÏve Bayes, random forest, CART, J48 decision tree, etc., and concluded that SVM performs better.

Boojoong et al. [8] applied machine learning techniques on N-opcodes. Zhang et al. [12] converted n-grams opcodes of size 2 into gray images, and further, they applied KNN and PCA algorithms to classify the malwares.

Nataraj et al. [5] converted binary files into images and classified malwares using image processing method. Keyaung et al. [11] converted opcode blocks into a color picture using a combination of image processing techniques. Further, the authors used similarity measure to identify the similarity between test cases.

Yang et al. [9] proposed a topic-specific methodology to classify malwares using data flow information. First, the authors used topic modeling algorithm like LDA for app clustering, and then, they identified information gain ratio for the topic-specific clusters.

Zhuo et al. [4] presented a hybrid method for detecting Android malwares based on the machine learning algorithms. First, the authors constructed CFG of the Android app to extract APIs. Based on the API, they created time, frequency and Boolean datasets. Using these datasets, an ensemble model is created to detect malwares.

Li et al. [3] proposed a malware detection technique named SIGPID, based on permissions utilization analysis. Rather than extracting and identifying all Android permissions, the authors created three pruning levels by dissecting the permissions data and finding the most relatable permissions that can be used in differentiating genuine and malicious applications. After this, machine learning techniques are applied to classify the malware families.

3 Proposed Methodology

Figure 1 demonstrates the flow of proposed method. The important phases in the proposed methodology are to identify Rogue applications using LDA and opcode image generation. In the stage, the APK samples are disassembled, and then in the next stage, the opcode sequence (with length two) is obtained for every sample. In the subsequent stages, every opcode sequence is given a value using probability method, and then, on the basis of these values, gray-scale images are generated for each. In the next stages, LDA is applied. First, a weight is provided to every feature, and then in the next phase, LDA algorithm is applied on the train samples, and the final outcome is obtained.

The detailed explanation of above-mentioned steps is discussed further.

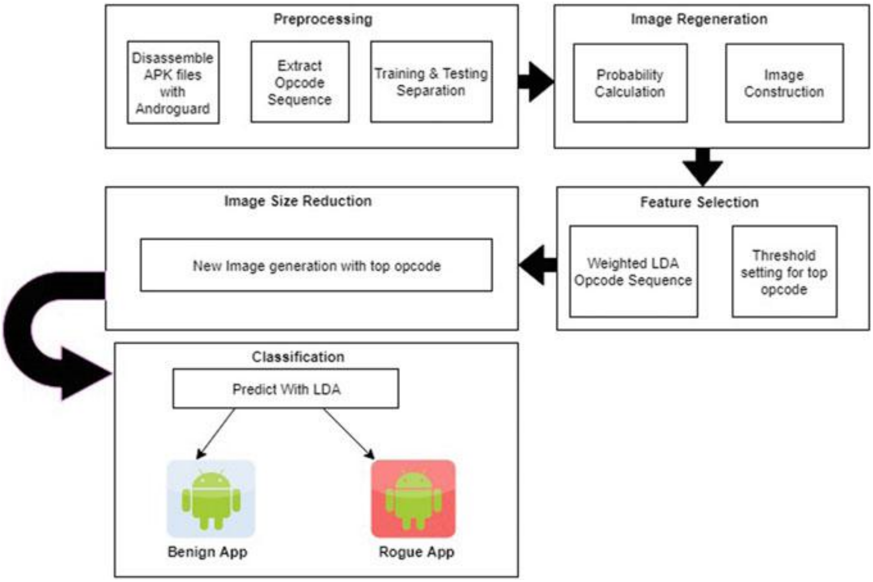
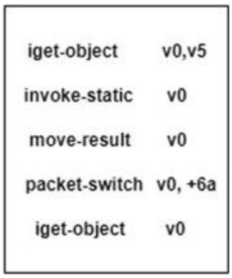


Fig. 1 Proposed methodology

3.1 Preprocessing

Every APK file contains atleast one dex file and several other resources in a compressed format. To extract the Dalvik Opcode, there are several tools available online. Androgurad [web link] is one of important tools used to disassemble the APK file. After obtaining the class files, an opcode sequence of size two is generated for each APK. Sample Dalvik opcode sequence is shown in Fig. 2. Then, the training and testing sets are generated from the dataset. In this situation, k-cross-fold is used.

Fig. 2 Sample opcode sequence



3.2 Image Creation

Generating an image for every sample is achieved using opcode sequence and based on the probability function (1):

$$V(Seq_i, Seq_j) = P(Seq_i, Seq_j) \tag{1}$$

Every image is demonstrated by an opcode sequence matrix, where every image pixel is described by a sequence of size 2 and its value. The image generated with distinct sample is shown in Fig. 3. It is clear from the figure that it can identify the difference between rogue application and benign application and also the similarity between distinct rogue apps.

3.3 Feature Selection

As the Dalvik contains 256 opcodes, an image of size 256 × 256 pixels is generated for every sample. However, due to this large size image, the accuracy rate will be less due to the presence of additional data. The objective of this section is to choose the significant opcodes that will help in increasing the accuracy and separating the class of rogue apps and benign apps from each other. The studies that have been done so far in this field have used information gain or PCA. In this paper, feature extraction is performed using LDA. LDA is a technique used in pattern identification, statistical analysis and machine learning to determine several feature combinations which distinguish at least two class objects among each other. The features must be selected on the basis of the ratio 'R.' R defines the ratio of the inside-class variances and neighboring-class variances. Using LDA, this step is performed by calculating matrices of inside-class scatter and neighboring-class scatter.

(a) BeanBot Binary Image (b) Jifake Binary Image (c) Benign Binary Image

Fig. 3 Original images

The inside-class scatter matrix is computed by Eq. (2)

$$S_i = \sum_{x=1}^{c} S_x \tag{2}$$

In Eq. (2), 'c' implies total number of classes and c = 2 in our case. S_x is the xth class inside-class scatter, and it is shown in Eq. (3):

$$S_x = \sum_{feD_i}^{k} (i - v_i)(i - v_i)^T \tag{3}$$

In Eq. (3) S_x denotes the result of inside-class scatter of the ith feature. 'f' denotes the value of ith feature, and vi is the mean-vector which is determined by using Eq. (4) given below:

$$v_i = \frac{1}{k_i} \sum_{feD_i}^{k} F_j \tag{4}$$

In Eq. (4) Fj denotes jth feature values of the kth sample, and k_i represents the total samples inside class. The neighbor-class scatter is computed using Eq. (5):

$$S_n = \sum_{i=1}^{c} N_i (m - m_i)(m - m_i)^T \tag{5}$$

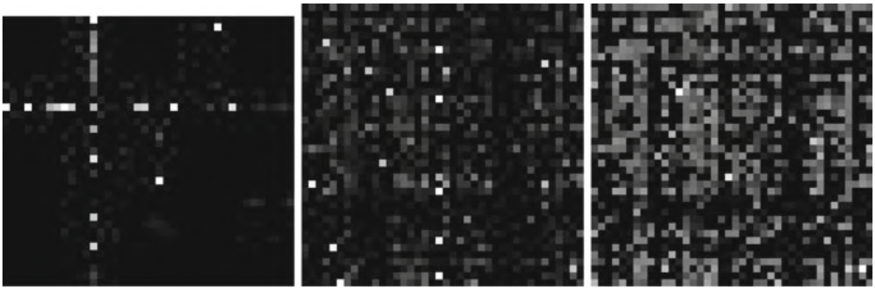
In Eq. (5), overall mean is denoted by 'm.' N_i denotes class size, and m_i denotes sample mean size. The values of N_i and m_i are computed for all the features, and every value is assigned weight using Eq. (6)

$$S(T) = \frac{S_n}{S_x} \tag{6}$$

The features that are having more similarities for inside-class scattering than neighbor-class scattering have a larger S_i, and hence, $S(T)$ is larger too. During this point, a threshold value has to be set for identifying the significant features, and this value is identified by experiments.

3.4 Image Size Reduction

Now, using the opcode sequences obtained in the previous step, a novel image matrix is generated such that the image only consists of the opcode which was determined in the previous phase. By doing this, smaller images can be generated for the samples. Image size reduction is depicted in Fig. 4.



(a) BeanBot Reduced Image (b) Jifake Reduced Image (c) Benign Reduced Image

Fig. 4 Reduced images

3.5 Classification

This section discusses the image conversion process. Every image is transformed to a vector such that every item depicts the pixel value, and (i, j) values of all the vectors depict P(seqi, seqj). Since the images are of same size, the vector lengths are the same too. Now, taking training sample datasets, LDA is applied, and then, testing process is done to predict the behavior.

4 Implementation

This section discusses the implementation part and results obtained from the experiment. First, we discuss the experimental setup, second, the datasets used and then finally the evaluation results.

4.1 Experiment Setup

To perform reverse engineering and obtaining opcode, we used Androguard [web link] inbuilt libraries based on Python. For LDA classification, we used scitkit-learn [web link] inbuilt Python library. The system on which the examination is performed has the following specifications: core-i7 Processor (2.5 GHz), 24 GB RAM and Ubuntu 18.04.

4.2 Dataset

We have used two datasets in our research, i.e., rogue app dataset and normal app dataset (containing benign apps only). The dataset samples are extracted from VirusTotal [web] and Andrototal [web]. Both of these provide open APIs. The rogue app samples contain malware families related to FlashingPuma CardinalFall, Triada, etc., and several unknown rogue apps that were listed during 2016–2018. Our dataset contains more than 100 K samples.

VirusTotal API labeled the samples according to their malicious intents. The suspicious file was first uploaded on VirusTotal, then, the file was scanned using 60 distinct anti-malware engines, and report was generated. There were three cases:

– If more than ten anti-malware had identified that the file belongs to a rogue application family, we marked it as malicious rogue family.
– If the family was not identified but the file was recognized as malicious rogue app, we labeled the file as unknown malicious rogue app.
– For non-malicious samples, if the detection rate is less than ten, we labeled them as non-malicious.

The dataset stats are given in Table 1.

4.3 Computation

To calculate the accuracy, there are following four criteria: True positive (T_p); total rogue app samples which are clearly classified. False Positive (F_p); total non-malicious samples which are classified by mistake. False Negative (F_n); total rogue app samples which are classified by mistake. True Negative (T_n); total non-malicious samples which are clearly classified. In every fold, the above items are computed, and fold accuracy is computed by Eq. (7):

$$F_{acc} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{7}$$

Equation (8) gives the overall accuracy.

Table 1 Dataset samples

Dataset sample app family	Sample count
Jifake	10,000
BeanBot	15,000
Unknown	45,000
Non-malicious (benign)	50,000
Total	125,000

$$Overall_{acc} = \frac{1}{N}\sum_{i=0}^{k} F_{acc_i} \quad (8)$$

4.4 Results

This section discusses the results of the study. As mentioned in the previous section, the dataset samples include rogue app families and collection of unknown rogue apps. For rogue app families, k-fold cross validation (with k = 10). Every fold consists of several samples for testing and training. In our experiment, the threshold is taken 40, and the overall accuracy achieved is 96.5% using diminished LDA. The accuracy is demonstrated in Fig. 5.

The execution details in k-fold (k = 10) are shown in Table 2. It is clear that the Tp rate is nearly 100%, which means that the proposed technique can detect almost all rogue apps in the dataset. The various image sizes taken at distinct threshold levels are demonstrated in Fig. 6.

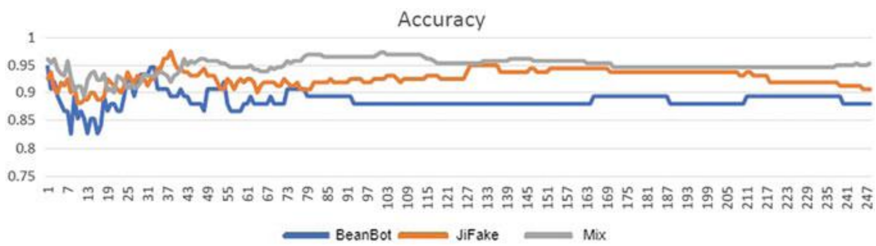


Fig. 5 Accuracy measure with different threshold

Table 2 K-fold cross validation

K-fold cross validation	True positive ratio (TPR) (%)	True negative ratio (TNR) (%)
K = 1	100	100
K = 2	100	0
K = 3	100	0
K = 4	100	0
K = 5	93	0
K = 6	98	0
K = 7	100	60
K = 8	100	0
K = 9	94	0
K = 10	100	0

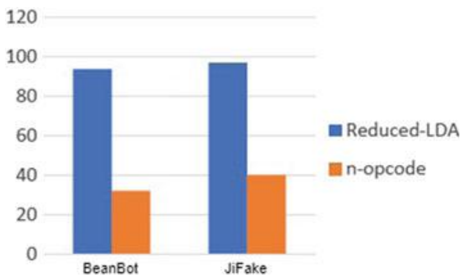


Fig. 6 Accuracy for N-opcode analysis and LDA with dimensional reduction

Table 3 Features number comparison

	N-opcode analysis	LDA with dimensional reduction
BeanBot	789,256	1200
Unknown rogue	834,255	1150
JiFake	792,252	1100

For BeanBot dataset, k-fold (with k = 5) cross validation is used. For unknown apps sample, the benign samples and rogue samples along with their families were classified with 96.5% accuracy (taking threshold value as 40). Total number of features required for training and testing phase is shown in Table 3.

It is clear that by reducing the number of features, the time and memory are reduced too. It can be seen that our proposed method can detect known rogue apps, as well as new zero day rogue apps.

5 Conclusion and Discussions

The identification of rogue application on smartphones has become a very crucial research area since the past few years. In this paper, a technique is proposed to decrease the opcode sequence vector length and classify the rogue applications using LDA (with dimensionality reduction). The proposed technique has the capability of detecting unseen samples of rogue applications with a good accuracy rate as discussed in Sect. 4. Future study can improve the probability function in order to decrease the false positive rate (FPR). PCA can be used if the size of the samples is very small. In that case, first PCA can be used, and then, LDA can be applied in order to get a very good accuracy rate.

References

1. Faruki, P., Bharmal, A., Laxmi, V., Ganmoor, V., Gaur, M.S., Conti, M., Rajarajan, M.: Android security: a survey of issues, malware penetration, and defenses. IEEE Commun. Surv. Tutor. **17**(2), 998–1022 (2014)
2. Jerome, Q., Allix, K., State, R., Engel, T.: Using opcode-sequences to detect malicious android applications. In: 2014 IEEE International Conference on Communications (ICC), pp. 914–919. IEEE (2014)
3. Li, J., Sun, L., Yan, Q., Li, Z., Srisa-an, W., Ye, H.: Significant permission identification for machine-learning-based android malware detection. IEEE Trans. Ind. Inform. **14**(7), 3216–3225 (2018)
4. Ma, Z., Ge, H., Liu, Y., Zhao, M., Ma, J.: A combination method for android malware detection based on control flow graphs and machine learning algorithms. IEEE Access **7**, 21235–21245 (2019)
5. Nataraj, L., Karthikeyan, S., Jacob, G., Manjunath, B.: Malware images: visualization and automatic classification. In: Proceedings of the 8th International Symposium on Visualization for Cyber Security, p. 4. ACM (2011)
6. Sahs, J., Khan, L.: A machine learning approach to android malware detection. In: 2012 European Intelligence and Security Informatics Conference, pp. 141–147. IEEE (2012)
7. Wei, L., Luo, W., Weng, J., Zhong, Y., Zhang, X., Yan, Z.: Machine learning-based malicious application detection of android. IEEE Access **5**, 25591–25601 (2017)
8. Xiaoyan, Z., Juan, F., Xiujuan, W.: Android malware detection based on permissions (2014)
9. Yang, X., Lo, D., Li, L., Xia, X., Bissyandé, T.F., Klein, J.: Characterizing malicious android apps by mining topic-specific data flow signatures. Inf. Softw. Technol. **90**, 27–39 (2017)
10. Yerima, S.Y., Sezer, S., McWilliams, G., Muttik, I.: A new android malware detection approach using bayesian classification. In: 2013 IEEE 27th International Conference on Advanced Information Networking and Applications (AINA), pp. 121–128. IEEE (2013)
11. Zachariah, R., Akash, K., Yousef, M.S., Chacko, A.M.: Android malware detection a survey. In: 2017 IEEE International Conference on Circuits and Systems (ICCS), pp. 238–244. IEEE (2017)
12. Zhang, J., Qin, Z., Yin, H., Ou, L., Xiao, S., Hu, Y.: Malware variant detection using opcode image recognition with small training sets. In: 2016 25th International Conference on Computer Communication and Networks (ICCCN), pp. 1–9. IEEE (2016)

An Indexed Non-probability Skyline Query Processing Framework for Uncertain Data

Ma'aruf Mohammed Lawal, Hamidah Ibrahim, Nor Fazlida Mohd Sani, and Razali Yaakob

Abstract Today, we are in the era of making multi-criteria decisions based on analysis of available data collected from autonomous databases that usually contain uncertain data. Skyline query technique returns a set of interesting objects (skylines) to the user by eliminating objects that are dominated by other objects within the database. Obviously, without doubt streamlining the process of processing skylines in providing answers to user-specified queries is inevitable. In this paper, we proposed *SQUiD* framework that combines an index-based technique with a prediction method to reduce the computational time for computing skylines over uncertain high-dimensional data. Through experimentations, results obtained clearly demonstrate the superiority of *SQUiD* framework over *SkyQUD* framework and *BBIS* algorithm.

Keywords Skyline query · Non-probability approach · Indexing technique · Uncertain data

1 Introduction

Modern computing has been largely responsible for influencing the growth of data been generated or captured [1]. Nowadays, databases containing tens or thousands of objects with hundreds of dimensions are been captured daily. The quest for processing

M. M. Lawal (✉)
Department of Computer Science, Faculty of Physical Sciences, Ahmadu Bello University, Zaria, Kaduna, Nigeria
e-mail: mmlawal@abu.edu.ng

M. M. Lawal · H. Ibrahim · N. F. Mohd Sani · R. Yaakob
Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Selangor, Malaysia
e-mail: hamidah.ibrahim@upm.edu.my

N. F. Mohd Sani
e-mail: fazlida@upm.edu.my

R. Yaakob
e-mail: razaliy@upm.edu.my

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_26

RB ID	Name	Monthly Rent Rate ($)	Bed Space	No. of Bathrooms	Square Feet (SqFt)
3000139	3 Bedroom	4699	3	2	1268
1384458	2 Bedroom	2900 – 4100	2	2	1016
3156011	2 Bedroom	3300	2	1.5	857
1633031	1 Bedroom	2630 – 3420	1	1	718
1516553	Studio	2956 – 2999	Studio	1	648
3419201	1 Room	2675	Studio	1	890
1974530	1 Bedroom	2780 – 2959	1	1	830
3261442	2 Bedroom	3395	2	1	790
3000139	3 Bedroom	4699	3	2	1268

Fig. 1 Data set for apartments at Cambridge, Massachusetts in the USA

query over high-dimensional data has resulted in the introduction of a number of preference query processing algorithms, among the most widely studied is the skyline query processing technique [1]. This technique applies the preferences of every user equivalently, such that objects which are dominated by another object are removed from the set of skyline objects [2]. Skyline algorithms are found useful in processing multi-objective query operations as they require no user-defined ranking function and as such are found applicable in multi-criteria decision support and recommender systems [3–6].

Uncertain data are data with uncertainty information, which exist widely in database applications. Due to the abundant volume of these uncertain data, accessing data for supporting many practical applications becomes tedious, and hence, an efficient technique for processing skyline objects is required [3–9]. In this work, uncertain data are defined as those data with continuous range values, in which their exact values are not explicitly represented in the database. Figure 1 is a snapshot of Rental Beast.com Web site for Cambridge, Massachusetts apartments in the United States of America (http://www.rentalbeast.com). From this snapshot, examples of continuous range values are 2780–2959, 2630–3420, 2956–2999, and 2900–4100 as presented under the monthly rent rate dimension.

A careful look at the dimension for monthly rent rate in Fig. 1, the values along this dimension are not comparable as the dimension contains both exact and continuous range values. With this kind of data representations, evaluating dominance relation between the set of objects becomes complicated and laborious. In addition, existing solutions proffered for addressing this problem employ a probability approach as in the case of [4], while [8] employed the median value calculations to map a continuous range value to an exact value. However, these approaches are not with their limitations. The probability-based approach involves many probability calculations, while the median value calculation is not realistic as the exact value for a continuous range value may not likely be at the center.

To address the uncertain data associated problem, we proposed skyline query processing on uncertain data (*SQUiD*), an indexed non-probability based framework, for evaluating skyline queries over such data. The framework employs the *X*-tree, an indexing structure to assist the skyline query process by grouping similar objects (in terms of space) into the same node of the tree. To adequately return exact values, we employ the *Fuzzy* algorithm [10] to avoid probability calculations. The algorithm identifies approximate value for each continuous range value and as such the conventional skyline algorithm could be utilized. Without loss of generality, we assumed that we are dealing with a certain database.

The rest of the paper is organized as follows: An overview of existing approaches for skyline query computation over uncertain data is presented in Sect. 2. The preliminaries to fundamental concepts of skyline query processing are presented in Sect. 3, followed by the discussion of our framework for handling skyline queries on uncertain data in Sect. 4, while experiment and results are presented in Sect. 5. The last section concludes this paper.

2 Related Works

Research works on skyline computation can be categorized into two distinct groups based on the approach followed in designing the algorithm, namely index and non-index-based approaches. The index-based approach requires preprocessing operations, namely sorting, categorizing, and indexing in order to compute skyline objects [8, 11–13]. While non-index-based approach requires no presorting, instead a threshold value is used to efficiently prune off dominated objects before returning the skyline results [4–7, 9, 14]. A lot of works on skyline query processing have been proposed by adopting the index-based approach as no probability computation is required before skyline results are returned. The index-based approach has been seen to be quite significant and helpful in speeding up skyline query processing as it plays a major role in minimizing the computational time for processing a query by streamlining the number of comparisons made when computing the skyline objects.

Among the index-based approach is the nearest-neighbor (NN) proposed by [15]. The skyline algorithm proposed by [14] called Bitmap utilizes an *m*-bit vector to encode each tuple. Similarly, [11] proposed the *SSPL* algorithm which is based on a pre-constructed data structure for supporting efficient skyline query computations. Another index-based algorithm is the divide-and-conquer (D&C) algorithm proposed by [2] for handling skyline query. With the D&C algorithm, the data space is partitioned along the dimensions repeatedly into two halves. The binary predicates for these two sets of results are combined and further processed to return skyline objects. The block nested loop (BNL) algorithm, a naïve, generic algorithm for computing skyline objects is proposed by [15]. The idea behind the BNL algorithm is that each tuple is sequentially scanned and compared with every other tuple in the primary memory. The dominated tuples are discarded, while non-dominated tuples are maintained in the buffer.

The work by [12] is the first to propose the branch and bound skyline (BBS), an NN-based algorithm which utilizes R-tree for supporting the evaluation of skyline objects. This pioneer work has inspired several other works on skyline query processing that adopted the index-based approach for computing skylines over uncertain data. Among the work that employed the index-based approach for computing skyline over uncertain data is the work of [8]. The *BBIS* algorithm in [8] employs the median value calculations to compute an exact value for each continuous range value. However, assuming that the exact value always lies at the center of the continuous range value is unrealistic. In addition, the *BBIS* algorithm seems inefficient, as the performance of the *BBIS* algorithm deteriorates as the number of dimensions increases beyond six dimensions.

The work by [9] proposed p-Skyline, a probability-based skyline query approach for evaluating the skyline of uncertain data. To achieve p-Skyline, first it derives the probability of each uncertain object to be a skyline and consequently uses a probability threshold value p to return the skyline by pruning off objects whose chances of being a skyline object are very slim (with probability $< p$). The only set back arises when the probability threshold value specified by the user is low or high. Specifying the permissible probability threshold value that will yield correct result remains an issue. Meanwhile, another work by [6] proposed *SkyQUD* framework that was inspired by the work of [9] employs a non-index-based approach for computing skyline over uncertain data. In this framework, a probability calculation method was utilized just like in the works of [7, 9, 14] to address the problem associated with skyline query processing over uncertain data in a dimension. Even though this work employs a method that captures the uncertainty associated with continuous range values, the *SkyQUD* algorithm seems inefficient for processing skyline query over uncertain high-dimensional data as many probability calculations need to be performed before skyline results are returned.

3 Preliminaries

To make our framework for processing skyline queries over uncertain data explicitly clear, definitions and notations used are presented in this section.

Definition 1 (*Dominance*) Given a database D with d-dimension, $d = \{d_1, d_2, \ldots, d_d\}$ and two objects p and $q \in D$. Assume minimum value is preferred for all the dimensions and without loss of generality we also assume that the value of p_i and q_i are non-negative ($p_i, q_i \geq 0$), where p_i and q_i denote the values of the ith dimension. Object *p dominates* object *q*, formally written as $p < q$ iff $\forall d_i \in d,\ p_i \leq q_i$ and $\exists d_j \in d,\ p_j < q_j$.

Definition 2 (*Skyline Query*) Given the database D, an object $p \in D$ is said to be a *skyline object* if and only if there does not exist any other objects $k \in D$ which dominates p. Then, the skyline on D is the set of all skyline objects in D.

Table 1 Example of D_u

Objects	Dimension 1	Dimension 2
A	110–120	50
B	120	25
C	60–83	45
D	78	105
E	100	50
F	120–160	40
G	145	40
H	85	45
I	120	40
J	160	45

Definition 3 (*Uncertain Database*) Given a database D with d-dimension, $d = \{d_1, d_2, \ldots, d_d\}$. The value of an object $p \in D$ for the ith dimension is said to be in a continuous range form if $p_i = [v_l - v_u]$, where v_l is the lower bound value and v_u is the upper bound value of p_i. The value of an object $p \in D$ is said to be an exact value if $p_i = c$ where c is a constant value. The database D is said to be uncertain, D_u, if it contains at least an object $p \in D$ where its value is in the form of $[v_l - v_u]$, otherwise D is said to be certain, D_c.

The example provided in Table 1 will be used throughout the paper to clarify the steps. The dominance relation and skyline query are easy and straightforward to apply to a set of objects whose values along its dimensions are all exact. For instance, it is obvious that object I dominates object J as I is better than J in both dimensions. However, this may not be the case when values along the dimensions can be a continuous range or an exact value. For example, comparing the objects C and D, we could not conclude that C is better than D or otherwise as the value of Dimension 1 for C is better than D if it falls below 78, and worst if it falls toward the end of the upper bound value.

In our work, we translate the above problem to the following: how to map a given uncertain database, D_u, to a certain database, D_c, in such a way that the conventional skyline query and dominance relation can be employed. This is further formalized as follows: Given a database D with d-dimension, $d = \{d_1, d_2, \ldots, d_d\}$. For each $p_i = [[v_l - v_u]$, its exact value $p_i = c$ is determined. In [8], c is the median value of v_l and v_u. In our work, c is identified based on the nearest neighborhood value of *p,* and thus, the value of c is not standardized to the center for every continuous range value as assumed in [8]. Our approach is in line with the real-world setting as similar objects will most likely have similar features.

4 The SQUiD Framework

Primarily, the essence of computing skylines is to identify objects which are not dominated by any other objects. In addition to the definition of basic concept of dominance relation introduced by [2] and skyline query problem mentioned in Sect. 1, the methodology for this is briefly expressed in this section. The *SQUiD* framework for evaluating skylines over uncertain data is presented in Fig. 2. It is made up of two phases, namely *data preprocessing* and *real-time*.

4.1 Data Preprocessing Phase

The purpose of the *data preprocessing* phase is essentially meant to build an *X*-tree for organizing the uncertain data in order to facilitate the reduction of computational cost. The *X*-tree clusters close distance objects together in a minimum bounding rectangle (MBR). Rather than comparing an object to every other object, we localize the evaluation of dominance relation among objects within each MBR. Using the

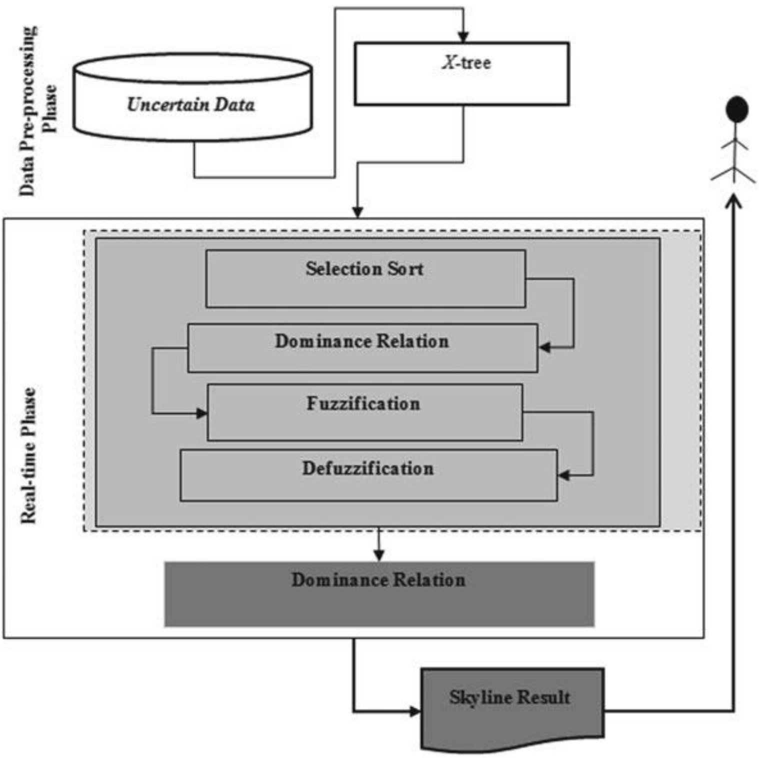


Fig. 2 Proposed *SQUiD* for processing skyline over uncertain data

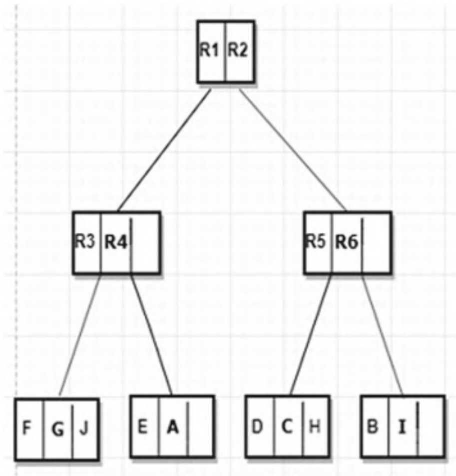


Fig. 3 X-tree structure of sample data set in Table 1

working example in Table 1, the objects are organized into six MBRs, namely *R*1, *R*2, *R*3, *R*4, *R*5, and *R*6 as depicted in Fig. 3. *R*3 contains objects *F*, *G,* and *J*; *R*4 contains objects *E* and *A*; *R*5 contains objects *D*, *C*, and *H*; while *R*6 contains objects *B* and *I*. With this structure, each object is compared to the objects within its MBR to identify the region candidate skyline objects.

4.2 Real-Time Phase

This phase is responsible for addressing the uncertain data associated problem in order to seamlessly process skyline query over uncertain data in which the *selection sort*, *dominance relation*, *fuzzification* and *defuzzification* methods were introduced.

Selection Sort. The *selection sort* technique employs a sorting algorithm that sorts an array by repeatedly finding the minimum element (considering ascending order) from unsorted part and putting it in the array. The objects within each MBR derived from the previous phase are entered into an array, and their index positions are sorted in ascending order. The sorted array list will serve as an input for the dominance relation method. Objects having the same value along the same dimension are ranked equally, while incomparable objects are not considered as they are automatically returned as candidate skyline objects.

Referring to the MBRs derived from the *data preprocessing* phase, four arrays are produced as depicted in Fig. 4. For example, the objects of *R3* are assigned an object indexed position (OIP) based on the order of entry into the array. Thus, objects *F, G*, and *J* are assigned OIP 1, OIP 2, and OIP 3, respectively. Where minimum values are preferred in both dimensions, these index positions are then used to rank the objects along each dimension. For instance, based on Dimension 1, *G* is ranked 2 and *J* is ranked 3, while *F* is not ranked as it is not comparable with other objects. Similarly,

R3

Objects	OIP	OID 1	OID 2
F	1	120 – 160	40
G	2	3	3
J	3	2	2

R5

Objects	OIP	OID 1	OID 2
D	1	3	1
C	2	60 – 83	45
H	3	1	3

R4

Objects	OIP	OID 1	OID 2
E	1	2	1
A	2	1	1

R6

Objects	OIP	OID 1	OID 2
B	1	1	2
I	2	1	1

Fig. 4 Sorted arrays based on the selection sort method

for Dimension 2, objects *G* and *J* are ranked 2 and 3, respectively, while object *F* is not considered for ranking. This process is repeated for all the MBRs, and the final arrays produced by the selection sort are depicted in Fig. 4.

Dominance Relation. The *dominance relation* technique is employed for identifying candidate skyline objects in the array produced by the selection sort. This is solely meant to remove dominated objects that would not contribute to the skyline result from the array. For instance, referring to the array of MBR *R*3 in Fig. 4, object *G* clearly dominates object *J* since its ranking index positions in both dimensions are better than the ranking index position of *J*. Thus, *J* is removed from the array, while the incomparable object *F* is returned together with *G* as the candidate skyline objects of *R3*. This process is repeated for each of the array. Figure 5 shows the results of the candidate skyline objects for each sorted array produced by the selection sort method.

Fuzzification. The *fuzzification* method [16] conducts a fuzzy analysis in order to quantify and handle vagueness, ambiguous associated with the uncertain data. The continuous range value found in the candidate skyline objects is decomposed into several segments in order to realize a fuzzy membership function (value within the range of 0 and 1). This method ensures that the exact value obtained is computed based on the neighborhood values, rather than using the median value.

R3

F	120 – 160	40
G	145	40
J	160	45

Candidate skylines, *R3* = {*F*, *G*}

R5

D	78	105
C	60 - 83	45
H	85	45

Candidate skylines *R5* = {*D*, *C*, *H*}

R4

A	110 - 120	50
E	100	50

Candidate skyline, *R4* = {*E*}

R6

B	120	25
I	120	40

Candidate skyline *R6* = {*B*}

Fig. 5 Candidate skyline objects of each sorted array

With *fuzzification* method, first for all the candidate skyline objects, the center points for each uncertain value are calculated, their continuous range value is portioned into smaller segments CRS_i, and their membership function is computed using the following function

$$Membershipfunction(Mf) = \frac{1}{1 + d_i} \sum_{i=1}^{k} CRS_i \tag{1}$$

where d_i is the distance from the segment d_i to all k nearest points and $(0 \leq Mf \leq 1)$. These points are obtained by utilizing the *knn* algorithm. This process is repeated to obtain a new center point until $0 \leq \varepsilon = (p_n - p_o) \leq 0.01$. When $\varepsilon = (p_n - p_o) \leq 0.01$, the $Membershipfunction$ value obtained becomes the final membership function values for the continuous range value. Using the working example, with objects *D*, *C*, *H,* a fuzzy membership value of 1.64785708836537 is obtained for object *C*.

Defuzzification. Primarily, the *defuzzification* method at this stage is to convert all fuzzy membership functions into exact values using the *centroid* (center of gravity—COG) *defuzzification* method. Our choice of *centroid defuzzification* technique is buttressed by the work of [16] which realized the *centroid* and *bisect* techniques as the best among other *defuzzification* methods. Thus, the corresponding computed exact value for each continuous range value is utilized to perform the skyline dominance relation. Employing the centroid defuzzification technique, the exact value is computed as

$$\text{Exact value} = \frac{\sum Area_i * ds}{\sum TotalArea} \tag{2}$$

So, we realize 72.64785708836537 as the exact value for the continuous range value (60–83) of object *B*.

Dominance Relation. The *dominance relation* method is again employed to compute the final skyline result by evaluating the dominance relation among the candidate skyline objects derived, as values of these objects are now comparable. Applying dominance relation on the exact values obtained in previous step, object *B* dominated objects *C* and *D* to realize objects *B* and *H* the final skyline objects for the set of objects presented in Table 1.

5 Experiment and Results

The *BBIS*, *SkyQUD,* and *SQUiD* skyline algorithms were implemented using Java (jdk-8u181-windows-x64) programming language running on a 64 bit Windows 7 operating. Our experiments were carried out on an Intel core i5 PC with 3.20 GHz processor and 16 GB main memory. The *BBIS* and *SkyQUD* algorithms were implemented based on the original algorithm specified by [6, 8]. This research work makes

Table 2 Experimental parameters

Parameter	Values	
Data set type	Synthetic	Real
Size of data set	1 M	21961
Data distribution	50	
Number of dimensions	3, 6, 10, 15, 20	6, 10, 15, 17

use of two data sets, namely synthetic and real data sets. The experimental parameters are presented in Table 2.

The synthetic data set includes correlated, anti-correlated, and independent data sets. We used the same generator as used by [5] to generate these synthetic data sets. To this effect, a *d*-dimensional data set of 1,000,00 objects, where *d* is varied from 3 to 20, and each object represents a uniform random variable generated between 1 and 100 was generated. For each object, we set one of its dimensions to represent the uncertain data, where the distribution between exact values and continuous range values is set equal to 50%. The length of each dimension of the continuous range value is from 1 to 100. To evaluate the performance analysis of these algorithms over real data set, we used the National Basketball Association (NBA) statistic (www.basketball-reference.com). The NBA data set consists of 21,961 objects from 1946 to 2009. Each object has 16 dimensions that represent various statistic values associated with NBA data set. The NBA records are made up of exact values, but we explicitly added another dimension to the NBA data set to represent an uncertain dimension. To measure the performance of these algorithms, the number of dimensions is varied, while CPU processing time is measured.

The results presented in Figs. 6, 7, 8 and 9 show the performance of the skyline algorithms with varying number of dimensions. Though there was a stiff competition between the *SkyQUD* and the *SQUiD* algorithm when the number of dimensions is

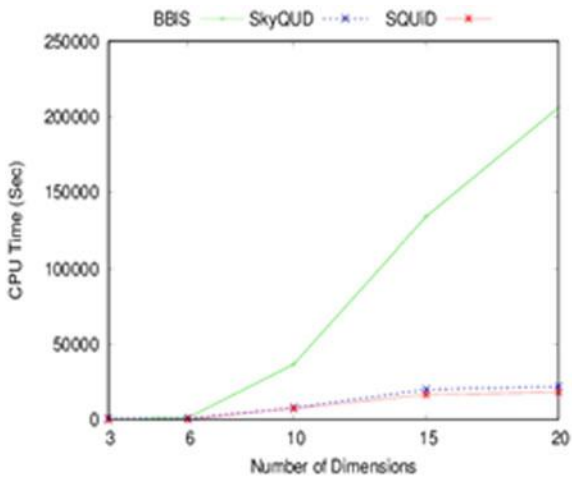


Fig. 6 CPU processing time (Sec) for anti-correlated data set

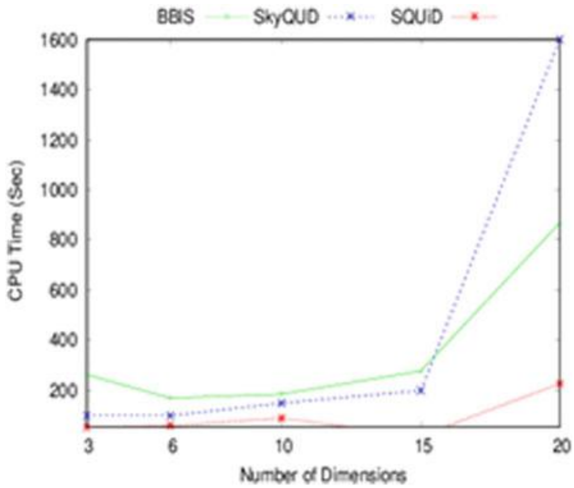


Fig. 7 CPU processing time (Sec) for correlated data set

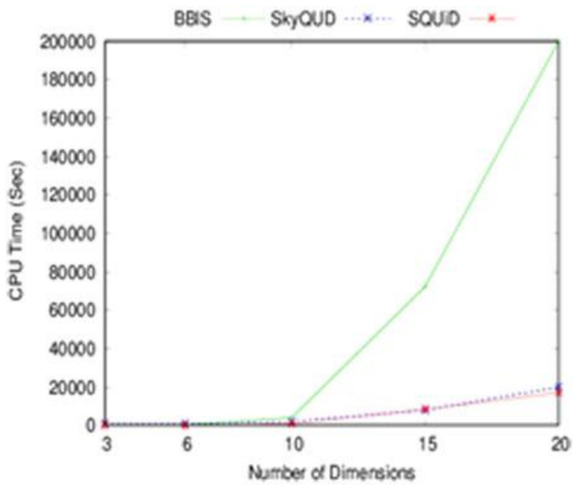


Fig. 8 CPU processing time (Sec) for independent data set

varied between 3 and 10, the superiority of the *SQUiD* algorithm is clearly demonstrated when the number of dimensions is varied beyond 10. Generally, by varying the number of dimensions, a steady deterioration of *SkyQUD* and *BBIS* performance due to dimensionality curse is noticeable. This translates to an increase in the CPU processing time due to high number of pairwise comparisons as every object is compared to other objects in SkyQUD algorithm. Similarly, in BBIS algorithm, the dimensionality curse resulted in an increase in the CPU processing time, due to high overlaps among MBRs of the *R**-tree indexing structure used.

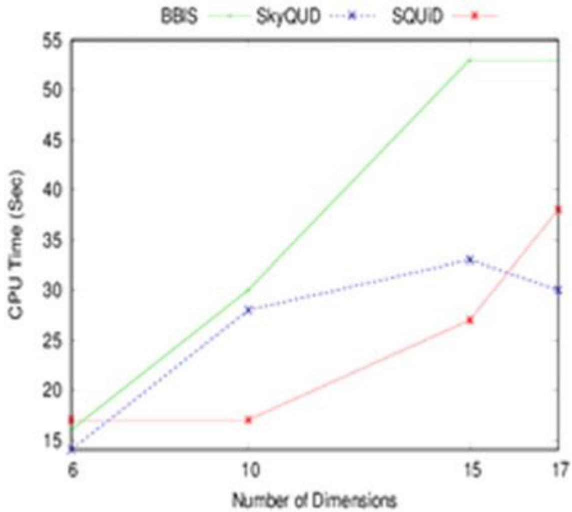


Fig. 9 CPU processing time (Sec) for real data set

6 Conclusion

In this research work, we realize the *SQUiD* framework that addresses the associated problem of processing skyline on uncertain data. The *data preprocessing* phase utilizes the *X*-tree indexing structure to organize the uncertain high-dimensional data in order to facilitate the process of processing skyline queries. In the *real-time* phase, we combine the *selection sort, dominance relation, fuzzification* and *defuzzification methods* to derive skyline results. To achieve an efficient technique for computing skylines over uncertain data, the continuous range values were mapped to exact values using the fuzzy algorithm so that the dominance relation and skyline query can be easily applied to return skyline results. Through experimentations, the superiority of our proposed framework against the *SkyQUD* and *BBIS* algorithms for computing skyline query over uncertain data efficiency has been demonstrated.

References

1. Lawal, M.M., Ibrahim, H., Mohd Sani, F., Yaakob, R.: Skyline query algorithms for computing uncertain databases: a survey, a chapter in a book: The landscape of Computing and Informatics Research Edition (2016–2017), pp. 173-191 (2018)
2. Börzsönyi, S., Kossmann, D., Stocker, K.: The skyline operator. In: Proceedings of the International Conference on Data Engineering, pp. 421–430 (2001). https://doi.org/10.1109/icde.2001.914855
3. Gothwal, H., Choudhary, J., Singh, D.P.: The survey on skyline query processing for data-specific applications. In: Proceedings of the 3rd International Conference on Internet of Things and Connected Technologies (ICIOTCT), pp. 26–27 (2018). https://doi.org/10.2139/ssrn.3168566

4. Saad, N.H.M., Ibrahim, H., Alwan, A.A., Sidi, F., Yakoob, R.: A framework for evaluating skyline query on uncertain autonomous database. Proceedings of the International Conference on Computational Sciences, pp. 1546–1556 (2014). https://doi.org/10.1016/j.procs.2014.05.140
5. Saad, N.H.M., Ibrahim, H., Alwan, A.A., Sidi, F., Yakoob, R.: Computing range skyline query on uncertain dimension. In: Hartmann, S., Ma, H. (eds.) Lecture Note in Computer Science: vol. 9828, pp. 377–388. Springer-Verlag. Database and Expert System Applications, (2016). https://doi.org/10.1007/978-3-319-44406-2_31
6. Saad, N.H.M., Ibrahim, H., Sidi, F., Yaakob, R.: Skyline probabilities with range query on uncertain dimensions. Proceedings of the Advances in Computer Communication and Computational Sciences, pp. 225–242 (2019). https://doi.org/10.1007/978-981-13-6861-5_20
7. Khalefa, M.E., Mokbel, M.F., Levandoski, J.J.: Skyline query processing for uncertain data. Proceedings of the International Conference on Information and Knowledge Management, pp. 1293–1296 (2010). https://doi.org/10.1145/1871437.1871604
8. Li, X., Wang, Y., Li, X., Wang, G.: Skyline query processing on interval uncertain data. Proceedings of the 15th International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing Workshops, pp. 87–92 (2012). https://doi.org/10.1109/isorcw.2012.26
9. Pei, J., Jiang, B., Lin, X., Yuan, Y.: Probabilistic skyline on uncertain data. Proceedings of the International Conference on Very Large Database, pp. 15–26 (2007)
10. Zadeh, L.A.: On fuzzy algorithms. Proceedings of the International on Fuzzy sets, Fuzzy logic, and Fuzzy systems, pp. 127–147 (1996). https://doi.org/10.1142/9789814261302_0010
11. Han, H., Li, J., Yang, D., Wang, J.: Efficient Skyline Computation on Big Data. IEEE J. Trans. Knowl. Data Eng. **25**(11), 2521–2535 (2013). https://doi.org/10.1109/TKDE.2012.203
12. Papadias, D., Tao, Y., Fu, G., Seeger, B.: Progressive skyline computation in database systems. Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 41–82 (2003). https://doi.org/10.1145/1061318.1061320
13. Swidan, M.B., Alwan, A.A., Turaev, S., Gulzar, Y.: A Model for Processing Skyline Queries in Crowd-sourced Databases. Indones. J. Electr. Eng. Comput. Sci. **10**(2), 798–806 (2018). https://doi.org/10.11591/ijeecs.v10.i2
14. Tan, K.L., Eng, P.K., Ooi, B.C.: Efficient progressive skyline computation. Proceedings of the International Conference on Very Large Database, pp. 301–310 (2001)
15. Kossmann, D., Ramsak, F., Rost, S.: Shooting stars in the sky: an online algorithm for skyline queries. Proceedings of the International Conference on Very Large Database, pp. 275–286 (2002)
16. Tóth-Laufer, E., Takács, M.: The effect of aggregation and defuzzification method selection on the risk level calculation. Proceedings of the 2012 IEEE 10th International Symposium on Applied Machine Intelligence and Informatics (SAMI), pp. 131–136 (2012). https://doi.org/10.1109/sami.2012.6208943

Analysis of Operational Control Mode of Intelligent Distribution Grids with Multi-microgrid

Yi Zhao, Jilai Yu, Danyang Guo, and Mingfei Ban

Abstract In the future, the comprehensive sensing multi-microgrids energy-efficiency management based on the ubiquitous electric Internet of things will promote the application of multi-microgrids. Moreover, the participation of microgrids in the electricity market has diversely enhanced their initiatives and abilities. Based on the analysis of the coupling relationship of the intelligent electrical distribution system of multi-microgrids, this paper reorganized the microgrid types from the differences in microgrid holder, grid feature, the role of distribution and grid-connected service. In the end, different types of microgrids operational control modes were put forward.

Keywords Multi-microgrid · Intelligent distribution system · Operational mode

1 Introduction

With the intervention of high-density distributed microgrid, it is imperative for the distribution grid to open its electricity sales to social capital. Multi-channel cultivation of the main players in the market for electricity sales will help more users to have the right to choose which can improve the quality of electricity sales service and the level of energy used. The reform of the electricity sales side is coordinated with the electricity price, the trading system and the power generation plan to forming effective competitive market structure and market system, promoting the optimal allocation of energy resources, improving energy-efficiency and clean energy consumption and improving the safety and reliability of power supply [1–5].

In the current smart distribution grid system, the microgrid is connected into the distributed power source (energy) on its own initiative, which makes it become a producer as a source (energy) provider instead of just a consumer like it used to be, and it can also be a prosumer which can both consume and produce power. At the same time, the effect of multi-core, high-permeability networks connected to

Y. Zhao (✉) · J. Yu · D. Guo · M. Ban
Department of Electrical Engineering, Harbin Institute of Technology, Harbin, China
e-mail: reef614@163.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_27

the network on energy distribution managers will be also on conventional energy, resizing, node voltage change or consequences arising such as rebuilding the grid and improving reactive energy. The solution is not only the optimal scheduling algorithm which can find the aim of optimizing and optimal consequence of multi-microgrid grid connection. As the diversify of multi-microgrid energy types; the increase of storage capacity, power source, and the connection of electric vehicles will be asked during the operation and management of grid distribution in the future. This appeal of benefit is not only about the physics and component but also concern with the profit of multi-microgrid when it participates in management and service of distribution as a competitive entity. The different roles of the microgrid in the distribution system will affect the control mechanism and operation mode of the entire system, prompting new forms, new formats, and new mechanisms for the distribution grids. Therefore, diverse modes, diverse service requirements and diverse service forms will make the power supply grid more flexible and more complex. How to allocate and control depends on the aim of your behavior, in some situation like commodity exchange/transaction with the main goal of profit, operation/control with the main goal of physical system security and quality control and social interaction/sharing with the main goal of information interaction and confidentiality or wisdom promotion. For different operational conditions, how to choose appropriate management modes in the process of coordinated operation between the distribution grid and the multi-micro grid will be a problem which needs to be solved [6–10].

2 Diversification of High-Permeability Distribution Mode of Multi-microgrid

The connection of the high-permeability distribution of multi-microgrid will lead to great changes in distribution management and service mode and awareness. The function of distribution is no longer a one-way charge collection for load users, nor centrally or hierarchically schedule as it the main grid. This is due to the characteristic of multi-microgrid such as the battery. Later, the comprehensive sensing multi-microgrids energy-efficiency management which based on the ubiquitous electric Internet of things will give the multi-microgrids soul and thought, and the participation in the electricity market has been significantly enhanced their initiative and ability while also reflecting certain diversity. These diversifications are mainly embodied in differences of microgrid holders, grid features, the role of distribution and grid-connected service. The division of the microgrid type is shown in Fig. 1 [11–15].

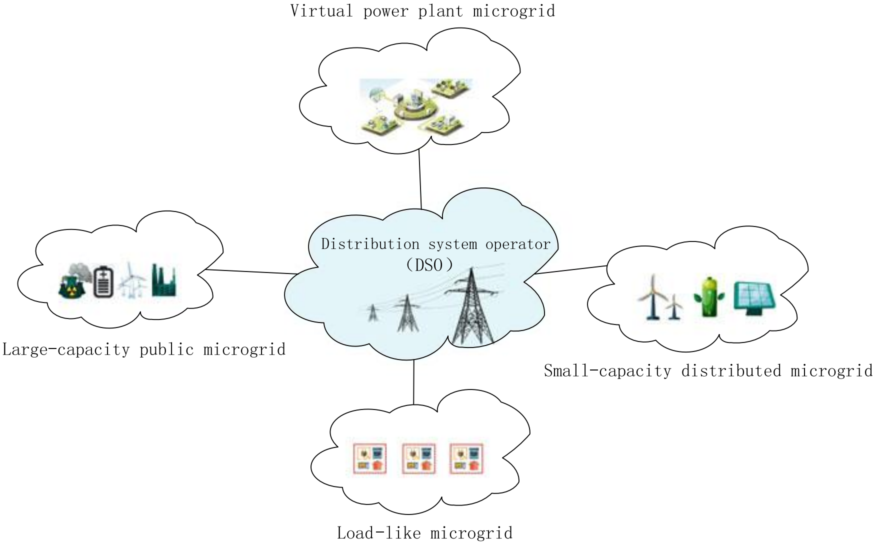


Fig. 1. Microgrid division

2.1 Virtual Power Plant Microgrid

Through subtly control methods and energy management, the virtual power plant integrates various distributed power supply, packs multiple small power supplies and outputs relatively stable output. In management, it is more focused on top-down management and control. The major characteristics of microgrid are similar to power plant when it is connected to the grid, and some of the distributed power supply, energy storage and controllable loads can be managed centrally. Among them, power supplies or loads with large randomness such as electric vehicles can be managed hierarchically. At present, they are studying how to maximize the distribution company's revenue by configuring capacity, coordinating operation, controlling follow-up of specified output and participating in market bidding. However, due to the uncertainty of renewable energy generation, virtual power plant microgrid needs to cooperate with the distribution companies to reduce the cost of electricity purchases, increase the revenue of virtual power plants and lead renewable energy generation corporations to improve their forecasting accuracy [16–20].

2.2 Large-capacity Public Microgrid

The distributed power supply capacity and energy storage capacity of the microgrid are large. Under the conditions of consume the local load first, they may participate in the power generation most time of a day to satisfy the market's needs. This demand is two-way. Inform your needs to the distribution manager while complying with the constraints of them. The microgrid with this kind of behavior mostly belongs to government (public welfare) which has the right to operate the distribution grid. Because its own energy-efficiency management system can optimize its operation, it is highly controllable and suitable for the combination of centralized and hierarchical control. The top-level control of the distribution grid is to centrally manage the electricity of distributed power generation, achieving reasonable energy distribution through communication commands. The bottom layer mainly coordinates and controls the energy interaction and local consumption of the distributed power supply, achieving stable operation in the same level microgrid by the two-way interacting of nearby communication commands [21–25].

2.3 Small-capacity Distributed Microgrid

The owner of this kind of microgrid is small-scale electricity sellers, and its participation, scope, scale or needs in the market may be diverse. Meanwhile, with different degrees of limitations, this kind of microgrid is the main player in the future distribution market. It is numerous and unwilling to publish their internal management information. The participation of market based on private information protection is clearly reflected in the competition in the nearest or certain region. This kind of microgrid is suitable for hierarchical coordinated control which can satisfy the local load. Because of the different control objectives during coordinated operation, the distribution grids system needs to be divided into three levels of coordination control frameworks: distribution grid level, microgrid group level and microgrid level. Different levels of targets and operating entities lead to different operating modes. On the premise that microgrid operates safely, stably and reliably, simplifying communication requirements, reducing power generation costs and improving distribution revenue are the optimal conditions for hierarchical coordinated control.

2.4 Load-like Microgrid

In most cases of small networks that contain a small amount of the distributed power source, it can be considered as a load but rarely displays the characteristics of the power supply to participate in the operation of the distribution network. However, it can participate in network operation as a service provider, for example, some intermittent loads (including central air conditioning systems, electric vehicle charging / replacing systems, etc.), and in most cases the energy can be (electricity) consumers, sometimes can participate in grid peaking or valley filling services and can also participate in grid emergency backup services (or precise load shedding services in emergency situations), and some of the operation only needs to be managed by the microgrid control center (energy management system). However, some operation needs to be reported to the management center of the distribution management center or even the higher voltage level grid (for example, to serve the provincial grid emergency control problem).

The large-capacity virtual power plant microgrid is prone to bully the weak in the market competition, and we can schedule the power plant with the method of centralized optimization which is beneficial to assist the operation of main grid. The small-capacity load microgrid market has insufficient competition desire and ability, and therefore, we can regard it as the traditional load. Large-capacity public welfare microgrid operation control can be realized by constructing multi-agent which is hierarchical control of multiple microgrids and centralized control of agent layers. There are a number of small-capacity distributed microgrids, it is scattered, and it is impossible and unnecessary to have a global optimal control for the distribution. Because it will still be a one-way flow in a certain area, and the distribution manager can give up part of the market space which can let the microgrids have a relatively free moderate game.

3 Research on Multi-microgrid High-Permeability Control Mode

A single centralized or hierarchical operation control mode cannot provide an effective solution. Therefore, it is necessary to build a multi-mode coexistence, multi-mode reasonable switching of the distribution integrated energy-efficiency operation management mode which the target is multiple microgrid's profit and the constraint is safe operation to enhance the distribution service capabilities. On this basis, the type of microgrid has been re-divided from the perspectives of microgrid, microgrid holder, typical features, main unit, role positioning, main service type and operation control mode as shown in Table 1.

4 Conclusion

This paper proposed different types of microgrid operational and control modes which need to be adopted in the future, and it is conducive to the improvement of microgrid's own energy-efficiency management capability and its trend of large-scale development. Power grid enterprises and the electric consumer will explore the construction of source-grid-load friendly interaction system together. The system can effectively stabilize the fluctuations caused by renewable energy, enhance the source-load of grid's complementary ability and provide an effective solution for the large-scale replacement of fossil-fueled energy by renewable energy.

Table 1. Microgrid types and main features

Microgrid types	Holder	Typical characteristics	Main load	Role		Main business type	Operational mode
				Production consume	Service		
(1) Virtual power plant microgrid	Government, power plant, power grid company's power sales company	Small quantity, large power supply area, large capacity, high grid voltage level (high voltage distribution network)	Distributed power supply, energy storage system, controllable load, electric vehicle	Product consume	Suppress the randomness of distributed power supply and participate in the main network to eliminate peaks and fill the valley	Directly trade with power users or pay full amount	Centralized control
(2) Large-capacity public benefit microgrid	Social capital investment in electricity sales companies (with distribution network operation rights)	Small quantity, large power supply area, large distributed power supply capacity and large energy storage capacity	Distributed power supply, energy storage system, controllable load, electric vehicle	Product consume	Participate in distribution network adjustment services	Non-marketization (or mixed market and non-market) operational/control behaviors with physical system security and quality control as the main objectives	combination of centralized control and layered control

(continued)

Table 1. (continued)

Microgrid types	Holder	Typical characteristics	Main load	Role		Main business type	Operational mode
				Production consume	Service		
(3) Small-capacity distributed microgrid	Independent sales company (does not have the right to operate the distribution network, does not undertake the guarantee service)	Large number, mainly in the park, small distributed power and energy storage capacity, low grid voltage level (medium voltage distribution network)	Distributed power supply, energy storage system, controllable load, electric vehicle	Product consume	Social interaction/sharing behavior with information interaction and confidentiality or wisdom (quality of life or comfort) as the main goal	Commodity exchange/transactions with market-oriented interests as the main target	Stratified operation control and price game in the nearest region
(4) Load microgrid	Independent sales company (does not have the right to operate the distribution network, does not undertake the guarantee service)	The vast majority, the park is dominated, a small number of distributed power sources, and the grid voltage level is low (medium voltage distribution network)	Partially interruptible load (including central air conditioning system, electric vehicle charging/replacement system, etc.)	Consume	Participate in power grid peak-filling services and grid emergency backup services (or precision load-cutting services in an emergency situation)	some of the operation only needs to be managed by the microgrid control center (energy management system) some operation needs to be reported to the management center of the distribution management center or even the higher voltage level grid (for example, to serve the provincial grid emergency control problem)	Traditional load management is the mainstay, and the peak operation and valley filling and emergency operations are under centralized control

References

1. Kai, Y., Chongbo, S., Haoran, J., Yi, S., Guanyu, S., Peng, L.: Coordinated control method for voltage and reactive power of active distribution network based on difference-of-convex programming. Proc. CSU-EPSA **30**(09), 108–113+150 (2018)
2. Wenlin, G., Guang, S., Jiaqing, Z., Jiang, T., Shiming, S., Bowen, W.: Unit distribution network based operating and controlling function design in new-generation distribution automation system. Autom. Electr. Power Syst. **42**(19), 163–170 (2018)
3. Jiaqi, L., Jian, C., Wen, Z., Yicheng, Z., Guannan, B.: Integrated control strategy for battery energy storage systems in distribution networks with high photovoltaic penetration. Trans. China Electrotech. Soc. **34**(02), 437–446 (2019)
4. Ning, Jiang, Pan, Zhang, Wang Yuqian, Wu, Liping, Tang Ping, Leijiao, Ge: Construction of core index of dispatching control system in smart distribution network. Proc. CSU-EPSA **30**(05), 86–91 (2018)
5. Chengshan, W., Peng, L., Hao, Y.: Development and characteristic analysis of flexibility in smart distribution network. Autom. Electr. Power Syst. **42**(10), 13–21 (2018)
6. Xilin, X., Yiqun, S., Liangzhong, Y., Zheng, Y.: Source-load-storage distributed coordinative optimization of and (part I): consensus based distributed coordination system modeling. Proc. CSEE **38**(10), 2841–2848+3135 (2018)
7. Zhou, S., Heping, T., Weixian, W., Mingyu, P., Lu, Z.: Coordinated control strategy of energy storage systems in distribution network with the integration of new energy. Mod. Electr. Power. **35**(01), 19–25 (2018)
8. Xiaobo, D., Limin, C., Chunhua, N., Xiangmei, D., Pudong, G., Zaijun, W.: Multi-level dispatching and control of active distribution network for virtual cluster of distributed photovoltaic. Autom. Electr. Power Syst. **42**(03), 21–31 (2018)
9. Xiaobo, D., Limin, C., Chunhua, N., Xiangmei, D., Pudong, G., Zaijun, W.: Research on grid-connected control strategy of distributed photovoltaic power supply. Power Sources **43**(04), 637–640+657 (2019)
10. Yueqiang, W., Wenbo, H., Weibin, L., Jiajie, Z., Yuan, Y., Feifei, Z.: Optimal allocation for distribution generator considering energy storing and randomness of photovoltaic. Electr. Measurement Instrumentation **56**(09), 91–96+134 (2019)
11. Qingrong, Z., Han, L., Xianyun, L., Shuzheng, W.: Research on active distribution network and network effects on non symmetrical voltage fault conditions with PV. Electr. Measurement Instrumentation **56**(04), 101–106 (2019)
12. Jian, L., Jianli, N., Yonghui, D.: Distribution automation system. Water and Power Press, Beijing, China (2003)
13. The guide of planning and design of distribution network: Q/GDW 1738—2012[S]. (2012)
14. Wen, L., Hongwei, D., Qia, D., et al.: Designand key technologies of optimal dispatch for smart distri-bution network. Autom. Electr. Power Syst. **41**(3), 1–6, 88 (2017)
15. Canizes, B., Soares, J., Vale, Z., et al.: Optimal approach for re-liability assessment in radial distribution networks. Autom. Electr. Power Syst. **11**(3), 1846–1856 (2017)
16. Mishra, A.: The review of trends in smart grid and its applications. Int. J. Cloud-Comput. Super-Comput. **4**(1), 27–34 (2017)
17. Ravi Kumar, T.: Easy to access the data storage in cloud computing. Int. J. Priv.E Cloud Comput. Environ. Manag. **4**(2), 1–6 (2017)
18. Bagwari, A., Kanti, J., Tomar, GS.: Multiple antennas based sensing technique. Int. J. Urban Des. Ubiquitous Comput. **4**(2), 41–50 (2016)
19. Wang, C., Shan, N., Lei, H.: The construction of 3D virtual driving scene of police wheeled armored vehicle. Int. J. Internet Things Big Data. **1**(1), 11–20 (2016)
20. Seungwoo, KIM., Jangju, KIM., Dongwook, HA., Yeonseung, RYU.: Carnegie Mellon University's CERT dataset analysis and suggestions. Int. J. Adv. Res. Big Data Manag. Syst. **1**(2), 1–6 (2017)

21. You, S.-H., Moon, J.-S., Yoon, M.-Y.: A protective and antioxidant and whitening effect of B16F10 melanoma cell induced by ultraviolet a by Rubus coreanus Miquel extract. Int. J. Bio-Sci. Bio-Technol. **10**(3), 21–26 (2018)
22. Tao, H., Tianshe, S., Kai, S.: Accuracy and comfort analysis of digital shoe design. Int. J. Multimed. Ubiquitous Eng. **13**(2), 13–20 (2018)
23. Lala, A., Bhaskar, A., Chakrabarti, P.: Review on energy efficient approaches on dynamic source routing protocol in MANET. Int. J. Smart Device Appl. **5**(1), 1–10 (2017)
24. Adinarayana, E.: Edge detector for spectral image using new approach. Int. J. Digit. Contents Appl. Smart Devices **3**(1), 9–18 (2016)
25. Tran, T.A., Han, K.S.: Factors affecting the consumer resistance to innovations on using the internet banking services—focused on vietnam. Int. J. Secur. Technol. Smart Device **4**(2), 1–6 (2017)

Engineering and Applied Sciences

Skin Lesion Classification: A Transfer Learning Approach Using EfficientNets

Vandana Miglani and MPS Bhatia

Abstract This paper studies the ability of deep convolutional neural networks (DCNNs) to classify skin lesions belonging to seven different categories. Two pre-trained state-of-the-art architectures for computer vision tasks ResNet-50 and Google's recently proposed, EfficientNet-B0, were fine-tuned for the classification task on the HAM10000 dataset. The dataset comprises 10015 dermatoscopic images belonging to seven classes of skin cancer melanocytic nevus, melanoma, benign keratosis, basal cell carcinoma, actinic keratosis, vascular lesions, and dermatofibroma. The aim of the study was to establish how well the EfficientNet family of models (which result in up to 8.4× parameter reduction and 16× FLOPS reduction) transfers to the skin classification task in comparison with the ResNet architecture. Overall, it was found that the EfficientNet-B0 model, with fewer parameters, outperformed the ResNet-50 model. EfficientNet-B0 model produced better ROC AUC values for each classification category and also achieved higher macro and micro averaged AUC values for the overall classification, 0.93 and 0.97, respectively (in comparison with, 0.91 and 0.96 of the ResNet-50 model).

Keywords Skin cancer · Transfer learning · Deep learning · Lesion classification · EfficieNtnet · ResNet

1 Introduction

Skin cancer is the uncontrolled growth of abnormal skin cells, caused particularly by excessive exposure to ultraviolet radiations. The past 10-year period, from 2008 to 2018, has witnessed a drastic increase in the annual occurrences of melanoma cases, primarily affecting the fair-skinned populations [1, 2]. More than 10,000

V. Miglani (✉) · M. Bhatia
Netaji Subhas University of Technology, New Delhi, India
e-mail: vandana.miglani.12@gmail.com

M. Bhatia
e-mail: bhatia.mps@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_29

people are killed each year in the USA attributing to the same [3]. Early diagnosis of the malignancy accompanied by treatment greatly improves survival chances [4]. If detected within the first year of its development, survival chances are as high as 97%. However, detection in the later stages lessens survival rates to 15% [5–7]. It follows therefore that an improvement in the accurate and timely detection of the condition could significantly impact the number of patients that can survive the tumor.

The usual process for detection involves visual observation of the suspicious region by a dermatologist. Diagnostic accuracy, however, is strongly related to the professional experience of the dermatologist. It has been concluded through various past studies that dermatologists, without additional support, can detect cancer with an accuracy of roughly 65–80% [8]. In doubtful cases, high-resolution dermatoscopic images are used to aid visual inspection. These images are obtained with the help of a high-resolution magnifying camera. The control of lighting reduces reflections on the skin making deeper skin layers visible. With the assistance of these images, the overall accuracy of detection reaches roughly 75–84% [9, 10].

Artificial intelligence and deep learning techniques have, in the past, proved their prowess in solving challenging problems. With the help of large datasets, these techniques have achieved results comparable with humans and in many cases, surpassing them. Further, there has been increasing use of deep learning techniques in medical imaging applications to assist physicians and doctors with detection, classification, and segmentation tasks. Deep convolutional neural networks (DCNNs), however, do not require any such prior feature engineering. Without the assistance of a human expert, DCNNs can extract features from images with the help of several convolutional layers. As the network gets deeper, finer features are extracted. Fully connected layers are finally used for the classification task. To accelerate the training procedure and achieve higher accuracy, transfer learning is a widely adopted technique in deep learning today. Pre-trained networks which have learned millions of parameters/weights on a different dataset are fine-tuned to a particular dataset, to learn specific features relevant to that dataset.

This paper aims at comparing two deep learning architectures ResNet [11] and EfficientNet [12]—fine-tuned for a multi-class classification consisting of seven different classes in the HAM10000 dataset [13], comprising 10015 dermatoscopic images. While the ResNet architecture has been used extensively in the past in image classification tasks, the EfficientNet architecture is a novel scaling model proposed by Google in 2019. The rest of the paper is organized as follows: Sect. 2 presents previous research and work done in the field, Sect. 3 elaborates the methods adopted in the study, and Sect. 4 presents the results obtained. Section 5 presents the discussion, and finally, Sect. 6 concludes the study.

2 Literature Review

Past research on skin lesion classification using machine learning methods required the transformation of input data/images into important handcrafted features which were application-specific. Features extracted were then fed into classifiers for the actual classification tasks. Almansour et al. [14] developed an algorithm for distinguishing between melanoma versus non-melanoma images using feature extraction from color and texture of images. These handcrafted features were fed into support vector machines for the classification task. Abbas et al. [15] classified tumors by extracting color texture properties and developed an algorithm AdaBoostMC, an adaptive boosting multi-label algorithm.

With the increase in computation power and influx of a large amount of data, however, deep learning techniques became popular, wherein networks were capable of learning which features are important for classification, without human involvement. Deep learning algorithms through convolutional neural networks have demonstrated great success in computer vision and image processing tasks. The breakthrough in computer vision came in 2012 when Krizhevsky et al. [16] trained a deep convolutional network capable of classifying 1.3 million high-resolution images belonging to 1000 classes in the ImageNet dataset. In the field of skin cancer research, Esteva et al. [6] in 2017 demonstrated the usage of a pre-trained Inception v3 CNN model to classify 129,450 clinical skin cancer images. The model was also tested against 21 board-certified dermatologists, and it was found that the CNN model achieved performance at par with human experts, demonstrating the competence of AI in medical imaging tasks. Yuexiang et al. [17] in 2018 proposed a deep learning framework comprising two fully convolutional residual networks which were evaluated on ISIC dataset 2017. The approach showed promising results. Haenssle et al. [18] in 2018 trained Google's Inception v4 architecture using dermatoscopic images and compared their diagnostic performance against a board of 58 certified dermatologists. Results showed that in most cases the CNN model outperformed the professionals. This study once again demonstrated that human experts and dermatologists could be aided in the task of classifying skin lesions. Outcome measures were sensitivity, specificity, and area under the curve (AUC) of receiver operating characteristics (ROC) values. Dorj et al. [19] in 2018 developed a deep convolutional neural network with an ECOC SVM classifier for the classification of four categories of skin cancer. For the extraction of features, the neural network used in their approach was the pre-trained AlexNet. Han et al. [20] in 2018 also fine-tuned the ResNet-152 architecture with images from the Asan dataset, MED-NODE dataset, and atlas site images to classify images belonging to 12 skin diseases.

This paper proposes to study how well EfficientNets (a family of models, proposed by Google in ICML 2019) transfer to the skin lesion classification task by comparing its performance to the backbone architecture of computer vision tasks—the ResNet architecture. To the best of our knowledge, no previous research has attempted to explore how well EfficientNets transfer to this kind of medical imaging dataset. While the ICML paper did demonstrate the ability of EfficientNets to transfer well for five out of eight popular datasets, none of the eight datasets were related to image processing in the medical sector.

3 Methods

3.1 Dataset

In this research, the human against machine with 10,000 training images (HAM 10000) dataset is used. It comprises 10,015 dermatoscopic images of pigmented lesions belonging to seven different classes of skin cancer, namely

– Melanocytic nevus
– Melanoma
– Benign keratosis
– Basal cell carcinoma
– Actinic keratosis
– Vascular lesion
– Dermatofibroma

Images' representative of the different classes of cancer is released as a training set and available publicly through the ISIC archive. Colored images present in the dataset originally have dimensions of 450 X 600.

Introductory qualitative analyses of the dataset revealed that the greatest number of training images belonged to melanocytic nevus class (6705 images) with the least number of images belonging to the dermatofibroma class (115 images). ISIC training images are representative of a balanced dataset in terms of the sex distribution—with 53.97% of the images belonged to males and 45.45% to females. Age distribution of the affected individuals depicts that the dataset comprises greater instances of patients having ages from 30 to 60. While melanoma has been proven to be age- and sex-dependent, these factors have not been taken into consideration by this study.

3.2 Background

This research aims to compare the performance of EfficientNet on the HAM10000 dataset for skin lesion classification, in contrast to the performance of the popular and widely used ResNet architecture. Proposed EfficientNet architecture outperforms other ConvNets on the ImageNet dataset while training faster and using fewer parameters. More importantly, in the context of the study, EfficientNets have also transferred well and achieved state-of-the-art accuracy in comparison with other ConvNets on five out of eight publicly available datasets: CIFAR-100, BirdSnap, Flowers, FGVC Aircraft, and Food-101. In this work, ResNet and EfficientNet have been used and evaluated.

ResNet Architecture Residual networks (ResNet) are one of the commonly used architectures for computer vision tasks. ResNets allow us to train deep neural networks of 150+ layers successfully by accounting for the vanishing gradient problem, using ‘skip connections’ or ‘identity shortcut connections.’ The ResNet-50 architecture, consisting of over 23 million parameters, is shown in Fig. 1.

EfficientNet Architecture In Google’s ICML paper 2019, ‘EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,’ they introduced a family of models called EfficientNets. Their study proposes a principled approach to scaling models and improving CNN performance. It is common knowledge that increasing CNN performance requires scaling in dimensions—depth, width, or resolution. Conventionally, in practice, to improve model accuracy, the dimensions are experimented with arbitrarily. After studying the impact of scaling in different dimensions, their study proposes adopting a more principled approach toward model scaling—by scaling each dimension uniformly with fixed scaling coefficients. After determining the fixed scaling coefficients, the baseline network is scaled to achieve higher accuracy and improved efficiency.

As the efficiency of the network depends not only upon scaling but also on the underlying baseline network, they also developed a new baseline network optimizing both accuracy and efficiency (FLOPS). The resulting architecture uses a slightly larger version of the mobile inverted bottleneck convolution (MBConv), which is then scaled up to obtain the family of models known as EfficientNets. Figure 2 shows the architecture of EfficientNet-B0 which is used in this study. While the ICML paper has demonstrated the effectiveness of using EfficientNet-B0 and its improved accuracy on the ImageNet dataset, we intend to establish how well this proposed architecture transfers to skin lesion classification dataset. In comparison with ResNet-50 model, EfficientNet-B0 has only 5.3 million parameters (reduction in number of parameters by a factor of 4.9). Further, it reduces FLOPS by an order of 11×. EfficientNet-B0, therefore, is much faster and efficient in comparison with state-of-the-art ResNet architecture.

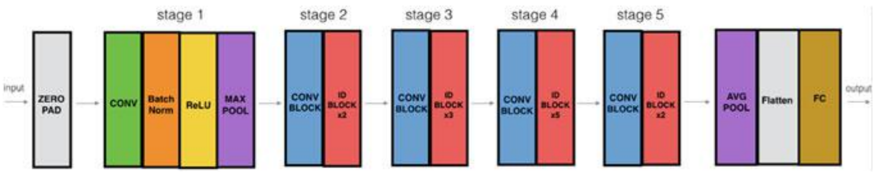


Fig. 1 Resnet-50 architecture

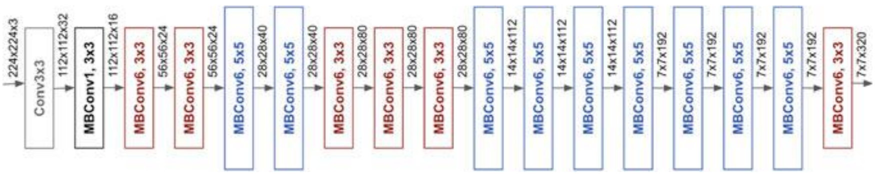


Fig. 2 EfficientNet-B0 architecture

3.3 Experiments

Experiments were conducted on a computer equipped with a Core i5 7th Gen processor, 8 GB DDRAM, and NVIDIA Geforce N16S graphic card. The images were resized from the original dimensions of 450 × 600 to the standard size of 224 × 224 to make them compatible with the ResNet architecture. The dataset images are augmented by rescaling, adjusting the rotation range (10), zoom range (0.1), width shift range (0.1) as well as the height shift range (0.1). For the classification task, the last layer from both the models was removed and replaced by a softmax activation layer with seven classes to enable multi-class classification of the skin lesions.

Both experiments were performed with the same training parameters to ensure uniformity and a fair comparison basis. Images were trained in batch sizes of 64 images for 50 epochs with an initial learning rate of 0.001. The learning rate adjustment was achieved with the help of Keras's callback—ReduceLROnPlateau, with parameter 'factor' set to 0.2 and 'patience' number of epochs set to 5. The dataset was split as 70% training set, 15% as the validation set, and 15% as the testing set to evaluate the performance of two architectures.

4 Results

The following parameters have been used to evaluate the performance of models:

$$Precision(PPV) = TP/(TP + FP)$$

$$Recall(TPR) = TP/(TP + FN)$$

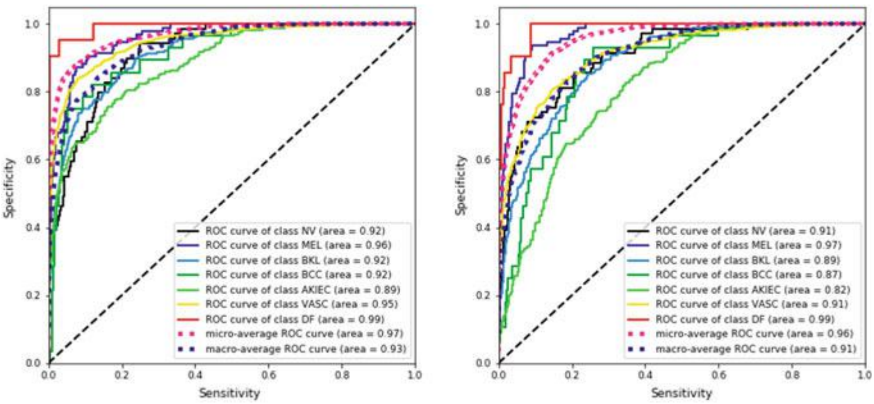


Fig. 3 ROC curves and AUC values for each and overall diagnostic category of EfficientNet-B0 and ResNet-50 (left to right)

Table 1 ROC AUC values for both algorithms on each and overall diagnostic category

Architecture	NV	MEL	BKL	BCC	AKIEC	VASC	DF	Micro	Macro
EfficientNet—B0	0.92	0.96	0.92	0.92	0.89	0.95	0.99	0.97	0.93
ResNet—50	0.91	0.97	0.89	0.87	0.82	0.91	0.99	0.96	0.91

Table 2 Micro and macro averaged precision, F1—score, and ROC AUC for each architecture

Architecture	Precision micro–macro	*F1*-Score micro–macro	ROC AUC micro–macro
EfficientNet-B0	0.89 – 0.85	0.89 – 0.86	0.97 – 0.93
ResNet—50	0.87 – 0.81	0.88 – 0.82	0.96 – 0.91

$$F1 - Score = 2 * TP/(2 * TP + FP + FN)$$

Here, *TP* refers to True Positive, i.e., the number of positive cases labeled correctly, FP refers to False Positive, i.e., the number of positive cases labeled falsely, and *FN* refers to False Negative, i.e., the number of negative cases labeled falsely. *F1* Score is a simple metric that takes into account both precision and recall, by computing their harmonic mean. This is particularly also used as the dataset is unbalanced. To visualize the trade-off between sensitivity and specificity graphically, ROC curve and associated AUC values are also used. The performance of the two proposed models is shown in Fig. 3, which represent the area under the curve (AUC) values for receiver operator characteristic (ROC) for each category of the two models. As the dataset is unbalanced, macro and micro averaged ROC AUC values are computed. As can be seen, EfficientNet-B0 demonstrates higher performance with micro and macro averaged ROC AUC values of 0.97 and 0.93, respectively, in comparison with micro and macro averaged ROC AUC values of 0.96 and 0.91 with the ResNet-50 architecture. Table 1 shows the ROC AUC values obtained from both the models for each diagnostic category. It can be seen that the EfficientNet-B0 model produced higher ROC AUC values for all categories except melanoma. Dermatofibroma category depicted the highest results for both the models. Table 2 shows the micro and macro average precision score, *F1*-score, and ROC AUC values. Precision is indicative of the proportion of positive identifications that were correct. *F1*-score achieves its best value closest to 1. Results here show that EfficientNet-B0 performed better than the ResNet-50 architecture. Further, the confusion matrices for both the models have also been shown in Fig. 4. Values on the diagonal show the recall value of each category. Looking at the matrices, it can be understood that both models confused melanoma and melanocytic nevus with each other. Vascular lesions category gave the best results for both models. For other classes, EfficientNet-B0 model showed higher recall values in comparison with the ResNet-50 model.

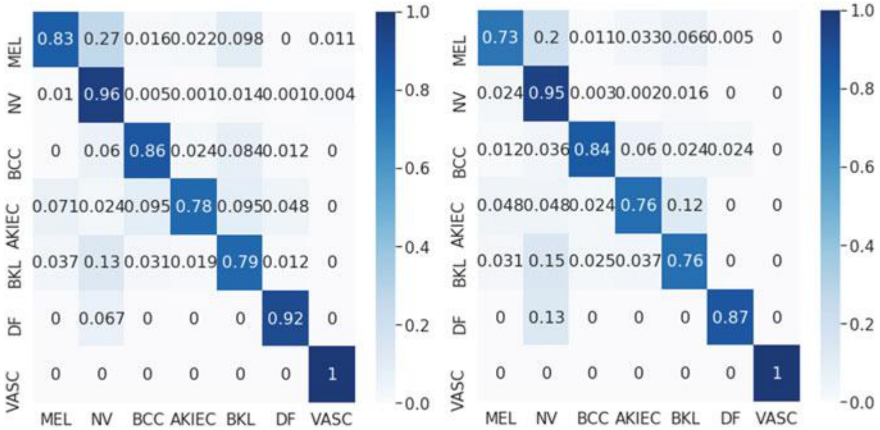


Fig. 4 Confusion matrix of EfficientNet-B0 model and ResNet-50 model (left to right)

5 Discussion

With one in five Americans developing skin cancer by the age of 70, it is imperative that accurate diagnosis takes place in the early stages when survival rates are greater than 90% [22]. As discussed earlier, the usual methods of inspection by physicians and dermatologists have certain inherent hurdles which can be assisted with the involvement of intelligent systems in the classification process. There has been increasing use of artificially intelligent systems aiding doctors and professionals in assisting healthcare solutions. Previously conducted studies have shown the capability of DCNNs in medical imaging applications such as identification of cardiovascular abnormalities, detecting fractures, diagnosis of neurological diseases, and many more.

In this study, we explored the ability of DCNNs of two different architectures to classify skin lesions in seven different classes using transfer learning. Using specific metrics, both architectures were evaluated. Insight was also gained about how EfficientNet-B0 transfers for the HAM10000 dataset. The ROC AUC values for the different classes are discussed as follows: for melanocytic nevus, benign keratosis, basal cell carcinoma, actinic keratosis, vascular lesions, and dermatofibroma, EfficientNet-B0 depicted higher ROC AUC values. For class melanoma, ResNet-50 performed marginally better with a ROC AUC value of 0.97 in comparison with 0.96 of the EfficientNet-B0 architecture. Overall, it can be concluded that EfficientNet-B0 outperforms the ResNet-50 architecture in classification accuracy with fewer parameters and FLOPS. When compared with previous research in the field, as per our knowledge, we are the first ones to demonstrate how well EfficientNets transfer to the skin classification lesion task. As EfficientNets use a far lesser number of parameters and computation power while achieving higher accuracy, they could potentially represent state-of-the-art architecture for computer vision tasks. Finally, it is also worth noting that previous studies [23] have established that highly experienced dermatologists are able to classify skin lesions belonging to categories of melanoma with ROC AUC value of 82.86% and basal cell carcinoma with ROC AUC value of 88.82%. As can be seen from our results, DCNNs are capable of achieving classification accuracy which is comparable to that of human experts.

6 Conclusion and Future Work

With the great success of DCNNs for different applications, it is no surprise that convolutional networks can assist in the healthcare sector as well. This study has demonstrated how pre-trained convolutional neural networks through transfer learning can be fine-tuned to extract representational features from images that can classify skin lesions with an accuracy which surpasses that of human experts. This can have far-reaching implications for the medical community as the life-saving diagnosis can be made easily available to patients which otherwise might not have easy access to it. Dependable diagnosis in the early stages can greatly impact the survival rates of patients. This study has also elucidated the most recent state-of-the-art DCNN, a family of models called the EfficientNets. It has also been found that EfficientNet-B0 transfers well for solving the skin lesion multi-class classification challenge. Future work in the area should also aim at including other metadata information about patients when classifying skin lesions as it had been established that melanoma is age- and sex-dependent.

References

1. Nami, N., Giannini, E., Burroni, M., Fimiani, M., Rubegni, P.: Teledermatology: state-of-the-art and future perspectives. Expert. Rev. Dermatol. **7**(1), 1–3 (2014). https://doi.org/10.1586/edm.11.79
2. Fabbrocini, G., Triassi, M., Mauriello, M.C., Torre, G., Annunziata, M.C., De Vita, V., Pastore, F., D'Arco, V., Monfrecola, G.: Epidemiology of skin cancer: role of some environmental factors. Cancers (Basel) **2**(4), 1980–1989 (2010). https://doi.org/10.3390/cancers2041980
3. Melanoma research gathers momentum. Lancet. **385**(2323) (2015). https://doi.org/10.1016/S0140-6736(15)61087-X
4. Ouhib, Z., et al.: Aspects of dosimetry and clinical practice of skin brachytherapy: the American Brachytherapy society working group report. **14**(6), 840–858 (2015)
5. Dorj, U.O., Lee, K.K., Choi, J.Y., Lee, M.J.M.T. and Applications.: The skin cancer classification using deep convolutional neural network. 1–16 (2018)
6. Esteva, A. et al., Dermatologist-level classification of skin cancer with deep neural networks. **542**(7639), 115 (2017)
7. Ruiz, D., Berenguer, V., Soriano, A., SNchez, B.J.E.S.W.A.: A decision support system for the diagnosis of melanoma: a comparative approach. **38**(12), 15217–15223 (2011)
8. Argenziano, G., Soyer, H.P.: Dermoscopy of pigmented skin lesions: a valuable tool for early diagnosis of melanoma. Lancet Oncol. **2**(7), 443–449 (2001)

9. Ali, A.R.A., Deserno, T.M.: A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data. Proc. SPIE. Int. Soc. Opt. Eng. **28**(8318), 1–6 (2012). https://doi.org/10.1117/12.912389
10. Fabbrocini, G., De Vita, V., Pastore, F., D'Arco, V., Mazzella, C., Annunziata, M.C., Cacciapuoti, S., Mauriello, M.C., Monfrecola, A.: Teledermatology: from prevention to diagnosis of nonmelanoma and melanoma skin cancer. Int. J. Telemed. Appl. **2011**(17), 125762 (2011). https://doi.org/10.1155/2011/125762
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016)
12. Tan, M., Le, Q.V.: Efficientnet: rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946 (2019)
13. Tschandl, P., Rosendahl, C. Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data **5**(180161), (2018). https://doi.org/10.1038/sdata.2018.161
14. Almansour, E., Jaffar, M.A.J.I.I.J.C.S.N.S.: Classification of dermoscopic skin cancer images using color and hybrid texture features. **16**(4), 135–139 (2016)
15. Abbas, Q., Celebi, M.E., Serrano, C., GarcA, I.F., Ma, G.J.P.R.: Pattern classification of dermoscopy images: a perceptually uniform model. **46**(1), 86–97 (2013)
16. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012)
17. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A.: Automated melanoma recognition in dermoscopy images via very deep residual networks. **36**(4), 994–1004 (2017)
18. Haenssle, H., et al.: Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. **29**(8), 1836–1842 (2018)
19. Dorj, U.O., Lee, K.K., Choi, J.Y., Lee, M.J.M.T.: The skin cancer classification using deep convolutional neural network. 1–16 (2018)
20. Han, S.S., Kim, M.S., Lim, W., Park, G.H., Park, I., Chang, S.E.J.J.o.I.D.: Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. (2018)
21. Skin Cancer Homepage. https://www.aad.org/media/stats/conditions/skin-cancer. Accessed 10 Aug 2019
22. Skin Cancer Facts and Statisctis. https://www.skincancer.org/skin-cancer-information/skin-cancer-facts. Last accessed 10 Aug 2019
23. Rezvantalab, A., Safigholi, H., Karimijeshni, S.: Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms. (2018)

Change Footprint Pattern Analysis of Crime Hotspot of Indian Districts

Mohd Shamsh Tabarej and Sonajharia Minz

Abstract Crime is the ubiquitous feature of all modern cities. Different cities face a different crime rate. Crime may be committed by a person who belongs to the same locality. In the crime analysis, identification of hotspot is helpful. Hotspot of the crime is the region where more criminal activities occur as compared to other regions. To better understand the criminal activity, temporal analysis of hotspot is necessary. In this paper, crime done under the Indian Penal Code (IPC) from 2001 to 2014 is used for the analysis. Geo-coding is done on the crime data for spatial processing. Kernel density estimation is used to find the hotspot of crime. Maps were created using ArcGIS. To find the footprint of the hotspot, grid overlay is constructed over the study region. Grid cell having point from the hotspot is considered as the footprint of the hotspot. A connected component of the hotspot is found to identify the number of hotspots. To find the temporal pattern, a number of hotspots and the footprint of the hotspot are plotted against the year. Characteristic vector is defined, which identifies the direction of change. Similarity Index is defined based on characteristic vector which identifies similarity between two patterns.

Keywords Special data · Hotspot · Footprint · Crime · Grid overlays · Kernel density · Characteristic vector · Similarity Index

1 Introduction

Crime is one of the ubiquitous features of all modern cities, but not all neighbouring cities face the same crime rate. It has certain inherent characteristics. Whenever a crime happens, it happens at a certain location, certain time, and it is done by someone who belongs to a certain place. The location of the crime may be the place where the

M. S. Tabarej (✉) · S. Minz
Jawaharlal Nehru University, New Delhi 110067, India
e-mail: mstabarej@gmail.com

S. Minz
e-mail: sonaminz@mail.jnu.ac.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_30

person who committed the crime resides or a nearby areas [1]. Therefore, place and time play a crucial role in understanding and tackling crime. Crime ranging from residential burglary to homicide is clustered in space and time, forming crime hotspot [2]. There are so many data available to build a model to analyze crime. For instance, the occurrence of crime, the location and time of the crime can be monitored as well as victim's and perpetrator's economic data can be recorded.

One of the techniques to analyze the crime is spatial hotspots. Crime hotspots are the region where occurrence of crime is more as compared to other regions. For example, for the data points shown in Fig. 2a, hotspots are shown in Fig. 2b, in red colour.

Crimes are clustered in space and in a neighbourhood, but its rate of occurrence varies from place to place. Prior studies suggest that the relationship between crime, their location, population, the business type is stationary, but it may not be the case. Crime is highly clustered, and neighbourhood of a crime may even have area with very low or zero crime. It is found that a very small percentage of crime location contains significant percentage of crime.

An extensive understanding of spatial and temporal variation within crime hotspot can greatly impact the crime prevention, apprehending of criminals, planning and resource allocation, crime modelling and forecasting, crime control programs and evolution of crime prevention [3]. Hotspot model assumes the predicted crime cases occur within the same locality. There are several techniques and models to analyze hotspot like, spatial histogram, mixture model, clustering, scan statistics, density-based estimation. These methods only show the crime type, location and time of the crime. It does not show any relation between space, footprint and density with time. So, there is a need to model the spatio-temporal pattern of crime.

This paper finds the hotspot of the crime using kernel density estimation. The map is created using ArcGIS for visualization. Grid overlays are constructed over the study region to find the footprint of the hotspot. A connected component of the hotspot defines the number of hotspots. To find the change pattern of the hotspot, footprint and number of hotspot were plotted for the year 2001–2014. Similarity Index was defined which gives the similarity between two trends.

Rest of the paper is organized as follows: in Sect. 2, the related work is discussed, in Sect 3, the related concepts is discussed, in Sect. 4, the proposed work that is used for finding change footprint of the hotspot is discussed, in Sect. 5, the experiments and results are discussed, in Sect. 6, the conclusions and future research directions are discussed.

2 Related Work

There are various methods to find the hotspot (i.e. denser region). For example, density-based clustering [4] like DBSCAN, DENCLUE, CLIQUE is able to find a denser region of arbitrary shape. Another method to find the denser region is histogram. In paper [5], Sharif, Maad, Halim and Derasit found the hotspot of the road accidents by dividing the road into small sections. They used nearest neighbour hierarchical clustering and spatial-temporal clustering for hotspot detection and visualize on ArcGIS. In paper [6], Bappee, Junior and Matwin proposed a machine learning model for crime prediction using geospatial feature of various categories of crime. They used HDBSCAN to find the hot point which is the centre of the hotspot. This study predicts the relationship between geographical region and criminal activity. In paper [7], Chaikaew, Tripath and Souries analyzed the pattern of diarrhoea in Chaing Mai province in Thailand to find their geographic distribution and hotspot pattern.

In paper [3], Herrmann examined the shift in the hotspot by an hour of a day and day of the week. In analyzing the different level of hotspot, dual shift in time and day is crucial. Nearest neighbour hierarchical clustering (NNHC) was used to find the hotspot. NNHC from CrimeStat was used for the analysis. In paper [8], Khalid, Wang, Shakeel and Nan found the hotspot of street crime for developing the crime control strategy.

The methods found in the literature does not show any relation between pattern of change in the footprint of the hotspot. Therefore, the objective of this paper is to find the hotspot of the criminal activity, identify its region, i.e. footprint, and find change pattern of the footprint over a period of time.

3 Preliminaries

Definition 1 A **data point** $\mathbf{x} = (x, y, a_1, a_2, ...a_k)$ is a vector that comprises the spatial and non-spatial attributes. The spatial attributes (x, y) represent a pre-decided coordinate values of a location such as city, a district centre. The non-spatial attributes $(a_1, a_2, ...a_k)$ values describe the attribute at that location.

Definition 2 A **study region S** is defined as an administrative region or the minimum bounding polygon covering the occurrence of the crime.

Definition 3 A **hotspot H** [9–11] of an activity is the location or region which is densely located in a geographical space. Therefore, given a set of observations, $D = \{x_1, x_2, ...x_n\}$ of a certain activity. Each subset of D may be a candidate hotspot. The hotspot may be mathematically defined as a Boolean function H with the domain as the power set of D and the binary set $\{T, F\}$ as the range, $H : 2^D \rightarrow \{T, F\}$. Candidate hotspot $H_i \subseteq D$ and density of the hotspot H_i are $den(H_i)$.

$$H_i = \begin{cases} T, & \text{if } den(H_i) \geq \delta \\ F, & otherwise \end{cases} \tag{1}$$

3.1 Kernel Density Estimation

Kernel density estimation (KDE) [12] is a nonparametric way to estimate the probability density function of the given data point. Given an arbitrary data set $x_1, x_2..., x_n \in D$, the idea is to place a kernel function on each data point $x_i \in D$. The kernel density estimate is given as:

$$f(x) = \frac{1}{Nh} w_i \sum_{i=1}^{N} K\big(\frac{x - x_i}{h}\big) \tag{2}$$

where $K()$ is the kernel function and x is a data point on which kernel density is to be calculated, h is a parameter for controlling the bandwidth of the density estimate, w is a weight associated with data points and is defined as

$$weight(w_i) = \text{number of crime cases at a perticular point} \tag{3}$$

Commonly used kernels are Gaussian, rectangular, triangular, Bi-weight. An illustration of the density estimation is shown in Fig. 1b.

Definition 4 Grid Overlay The study region S is divided into a two-dimensional grid. The grid G represents the graticule, which overlays the study region. The dimension of the grid is $N \times M$, N being the rows and M being the column, N and M are users defined. Thus, G is a two-dimensional array of grid cells $g_{i,j}$ where i is the ith row and j is the jth column. A grid cell may or may not contain a data point. Set of data point $T_{i,j}$ in a grid cell $g_{i,j}$ is given as:

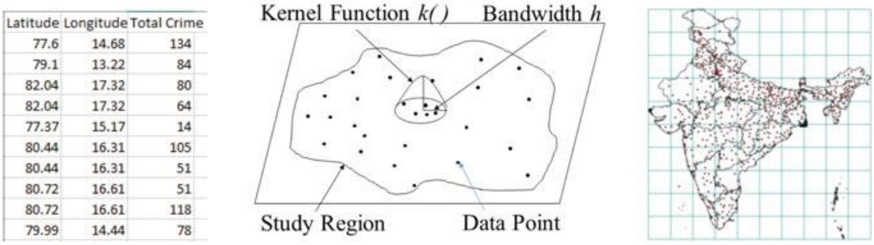
$$T_{i,j} = \begin{cases} \{x | x \in g_{i,j}\}, & \text{if } |g_{i,j}| > 0, \\ \phi, & \text{otherwise} \end{cases} \tag{4}$$

Example 1 Grid over the study region is shown in Fig. 1c. For simplicity, $10 \times 10 = 100$ grid cell is shown. The number of data points is 724, only 10 out of 724 data points are shown; each data point is associated with a weight w, which is shown in Fig. 1a.

Definition 5 Footprint [10] of a certain activity or pattern is defined as the extent or geographical coverage. In this work, footprint of a data point is considered to be the grid cell $g_{i,j}$ in which a data point lies. Footprint of a hotspot H is defined as the set of grid cell $g_{i,j}$ in which data points lie and those cells are also qualified for the hotspot. If f is a function to find the footprint of the hotspot, H then $f(H) = \xi$, where ξ denotes the footprint of the hotspot.

Measure of Footprint

– For a data point $x \in D$, $f(x) = \{g_{i,j} | x \in g_{i,j}\}$



Latitude	Longitude	Total Crime
77.6	14.68	134
79.1	13.22	84
82.04	17.32	80
82.04	17.32	64
77.37	15.17	14
80.44	16.31	105
80.44	16.31	51
80.72	16.61	51
80.72	16.61	118
79.99	14.44	78

(a) Data points (b) Illustration of KDE (c) Grid over S

Fig. 1 An illustration of KDE and grid over the study region S

- For the set of data points $\{x_1, x_2, ...x_k\} \in D$, $f(x_1, x_2, ...x_k) = \{\cup_{i=1}^{k} g_{i,j} | x_1, x_2, ...x_k \in g_{i,j}\}$
- For the hotspot H, $f(H) = \{\{g_{i,j}\} | x_i \in g_{i,j}\}$

Example footprint of the hotspot as shown in Fig. 2e is the extent or region covering the hotspot.

3.2 Change Pattern Analysis

Time series data is the series of observations that a variable takes over a period of time. In the analysis of time series data, changes in values are as important as the value itself. In view of this, the series which are changing in the same manner are similar to each other regardless of the degree of change. In order to find the change pattern, based on the value obtained at the fixed intervals, a novel algorithm using binary string representation is used [13]. The proposed algorithm has several advantages. First, the proposed algorithm finds the common structure between different objects. Second, it represents the common structure by using binary system; this helps in fast processing.

Grid over the study region is created, and the grid which contains the points belonging to a hotspot is included in the footprint of the hotspot.

i Suppose at time t_i dataset D_i contains h_n number of hotspots. Let $\tau_1, \tau_2, ...\tau_n$ is the number of grid cells of the hotspot of a crime $h_1, h_2, ...h_n$, respectively.

ii Let σ_i denotes the sum of grid cell in the hotspot at time t_i and is given by

$$\sigma_i = \sum_{i=1}^{n} \tau_i \tag{5}$$

iii For the sequence of n observation, **difference vector** is defined as $< \delta_1, \delta_2, ...\delta_n >$ where δ_i denotes the difference and is given as: $\delta_i = \sigma_i - \sigma_{i-1}$

iv **Characteristic vector** of change [13] is defined as $\alpha =< \lambda_1, \lambda_2, ...\lambda_{i-1} >$ where each λ_i is

$$\lambda_i = \begin{cases} 1, & \text{if } \delta_i > 0 \\ 0, & \text{otherwise} \end{cases}$$

v **Change Similarity** Given two time series $x =< x_1, x_2, ...x_n >$ and $y =< y_1, y_2, ...y_n >$. Let $< \delta_{x1}, \delta_{x2}, ...\delta_{xn-1} >$ and $< \delta_{y1}, \delta_{y2}, ...\delta_{yn-1} >$ be their corresponding difference vector and $< \lambda_{x1}, \lambda_{x2}, ...\lambda_{xn-1} >$ and $< \lambda_{y1}, \lambda_{y2}, ...\lambda_{yn-1} >$ be their corresponding characteristic vector. Then, x and y are said to have similar trend if for some $1 \leq i \leq n$, $\lambda_{xi} = \lambda_{yi}$.

vi **Similarity Index** Similarity index measures the similarity between a pair of time series. This index may be useful in clustering time series based on characteristics vector of change. Similarity between two time series D_1 and D_2 is given as:

$$Sim(D_i, D_{i+1}) = \frac{p+q}{p+q+r} \tag{6}$$

where $p = \sum_{i=1}^{n}(\lambda_{xi} = \lambda_{yi} = 1), q = \sum_{i=1}^{n}(\lambda_{xi} = \lambda_{yi} = 0)$ and $r = \sum_{i=1}^{n}(\lambda_{xi} \neq \lambda_{yi})$ such that $p + q + r = n$.

The value of the similarity index is $0 \leq Sim(D_1, D_2) \leq 1$.

Example 2 For the two time series $t_i =< 4, 2, 6, 1, 8, 3, 7, 9 >$ and $t_{i-1} =< 5, 3, 8, 5, 9, 1, 2, 4 >$, difference vector is given as $< -2, 4, -5, 7, -5, 4, 2 >$, $< -2, 5, -3, 4, -8, 1, 2 >$ and characteristic vector $< 0, 1, 0, 1, 0, 1, 0 >$ and $< 0, 1, 0, 1, 0, 1, 0 >$, respectively. Magnitude of the two time series is not equal, but the characteristic vector are bitwise equal, so the two time series are similar in trend. Similarity index $Sim(t_i, t_{i-1}) = 1$.

4 The Proposed Work

The methodology for the experiments is divided into three phases, namely hotspot detection, hotspot analysis and change analysis. Each phase is described below:

Hotspot detection: Step 1 of the hotspot detection algorithm describes the preprocessing phase. Spatial attribute c_i is linked to the non-spatial attribute of each p_i to result into P for spatial processing. A point vector of the form $P = (45.1222, 43.2315, 1, 5, ..., 9)$ be the output of this phase. In step 2, hotspot is identified, $den(P_i)$ calculates the density of each point (lines 5–7), SD is the standard deviation of the density of all points (line 8), δ defines the threshold for a point to be included in the hotspot (line 9), and finally, the hotspot H is identified (lines 10–15).

Hotspot Analysis: Input to the hotspot analysis phase is H. Step 1 identifies the footprint of the hotspot. Grid G is created over the study region S (line 1); a simple example is shown in Fig. 1c. Grid cell $g_{i,j}$ which contains points in the hotspot is identified and included in the footprint (lines 2–9).

Algorithm 1 Hotspot Detection

Input:

- Dataset d and C
- Study region S

Output: Hotspot H
Algorithm:
Step 1: Data Preprocessing
1: Let $p_1, p_2, ...p_n \in d$ and $p_i = (a_1, a_2, ...a_k)$ where $a_1, a_2, ...a_k$ are the attributes of p and d is the data set containing p_i
2: Let $c_1, c_2, ..., c_n \in C$, where each $c_i \in C$ is the geographic coordinates of p_i, i.e. $c_i = (x_i, y_i)$ where x_i, y_i are the latitude, longitude, respectively.
3: Geo-coding: $P_1, P_2, ...P_n \in D$, where $P_i = (c_i, p_i)$
4: Point vectors $\mathbf{P_i} = (x_i, y_i, a_1, a_2, ...a_n)$
Step 2: Hotspot Identification
5: **for** $i \leftarrow 1$ to n **do**
6: $\quad den(P_i) = f(P_i)$
7: **end for**
8: Compute $SD = std(den(P_i))$
9: Threshold $\theta = \frac{1}{\alpha} \times SD$
10: Hotspot $H = \phi$
11: **for** $i \leftarrow 1$ to n **do**
12: $\quad$ **if** $den(P_i) \geq \delta$ **then**
13: $\quad\quad H = H \cup P_i$
14: $\quad$ **end if**
15: **end for**

In Step 2, the number of hotspots n_H is to be found. A function called $connected\_component()$ is defined which finds the connected components of the grid. Finally, connected grid cell is counted which gives n_H.

Change Analysis Input to the change analysis phase is the number of hotspots and footprint of the hotspot for the temporal dataset $D_2, D_2, \ldots D_k$. Trend similarity for the footprint of the hotspot and the number of the hotspots was found as described in Sect. 3.2 for different datasets. ξ_{ci} represents the footprint of the hotspot of a particular of type c like Murder, and n_{Hci} represents the number of hotspots at time t_i. Similarity Index describes the similarity of the crime (line 2–3).

5 Experiments and Results

5.1 *Study Area and Data Collection*

Data set that are used in this study work is obtained from www.data.gov.in, which contains the district-wise crime data committed under the Indian Penal Code (IPC) for the years 2001–2014. The crime which is included in this data set contains murder, attempt to murder, rape, kidnapping and abduction of women's and girls, cheating,

Algorithm 2 Hotspot Analysis

Input:

– Hotspot H
– Number of rows M and number of columns N

Output:

- Footprint of the hotspot ξ
- Number of Hotspots n_H

Algorithm:
Step 1: Footprint identification

```
Create grid G = grid(S, M, N)
Footprint ξ = φ
while G ≠ φ do
    for i ← 1 to count(H) do
        if H_i ∈ g_{i,j} then
            ξ = ξ ∪ g_{i,j}
        end if
    end for
end while
```

Step 2: Number of Hotspots

```
Number of Hotspots n_H = connected_component(H)
```

Algorithm 3 Change Analysis

Input: For the data set $D_1, D_2, ...D_k$

– Time series of number of hotspot $n_{H1}, n_{H2}...n_{Hk}$
– Footprint of the hotspot $\xi_1, \xi_2, ...\xi_k$.

Output: Change Similarity Matrix
Algorithm:

```
for i ← 1 to k do
    similarity index of footprint = Sim(ξ_ci, ξ_ci+1)
    similarity index of the number of hotspots = Sim(n_Hci, n_Hci+1)
end for
```

dowry death, etc. Out of which only murder, rape, kidnapping and dacoity are chosen for the analysis. Geo-coding is done by adding district coordinate of each district to the data set for spatial processing. Figure 2a shows the location of the cities of the crime.

5.2 *Evaluation Parameters*

To observe the results of the experiments carried out in this work, the following evaluation parameter has been used:

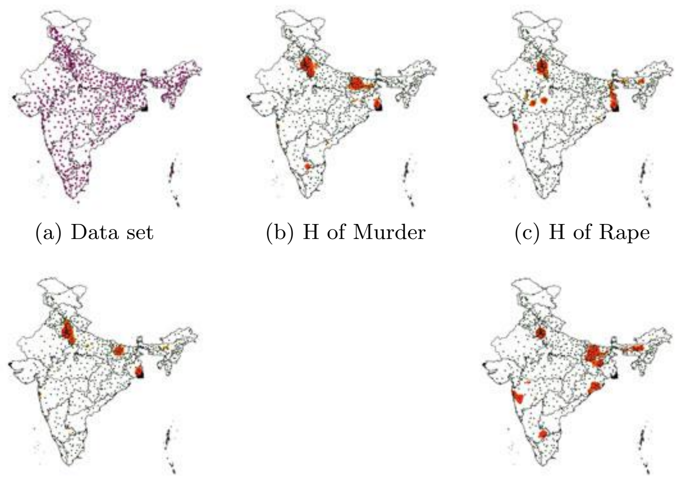
1. Number of grids: the coverage of the footprint is the polygon drawn out by the continuous grids. Therefore, the number of grids is used as the measure of the coverage.
2. Number of hotspots: from year to year, the number of hotspots could also vary. Therefore, the number of hotspots is also considered as the evaluation parameter.

5.3 Results

Crime hotspot for the year 2014 is shown in Fig. 2. Hotspot for murder, rape, kidnapping and dacoity is shown in Fig. 2b, c, d, e, respectively. Hotspots are arbitrary in shape and size, mostly centred to a place. Figure 2b hotspot of murder in which, there are eight hotspots were found out of the three are most denser and spread over larger area one in Delhi, one in Bihar and one in West Bengal.

Figure 2c shows the hotspot of rape. The denser hotspot was found in Delhi, West Bengal and in Madhya Pradesh. Figure 2d shows the hotspot of kidnapping. One hotspot covers the larger region including Delhi, UP West and Haryana, and two other major hotspots were found in Bihar and West Bengal. Figure 2e shows the hotspot of dacoity. The major hotspot was found in Delhi, Bihar, Jharkhand, Assam, Orissa, Maharastra and Karnataka. Out of these states, Bihar is the most affected state for dacoity.

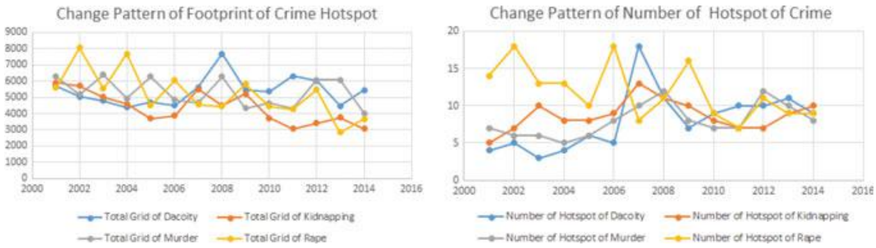
Figure 3a shows the change pattern of the footprint of the crime hotspot over the year 2001–2014. The curve goes on decreasing from 2001 to 2007. In 2008, all the crime shows an increase in the footprint of crime. After 2008, again there is a



(a) Data set (b) H of Murder (c) H of Rape

(d) H of Kidnapping (e) H of Dacoity

Fig. 2 Hotspot H of different crime cases for the year 2014



(a) Change pattern of footprint of crime hotspot 2001-14

(b) Change pattern of number hotspot of crime 2001-14

Fig. 3 Change footprint pattern of the hotspot

Table 1 Similarity index for the footprint of the hotspot for all the cases

Similarity index	Murder	Rape	Kidnapping	Dacoity
Murder	1	0.3076	0.3846	0.5384
Rape	0.3076	1	0.6153	0.3076
Kidnapping	0.3846	0.6153	1	0.38460
Dacoity	0.5384	0.3076	0.38460	1

decrease. Also, there is a slight increase and decrease but overall there is a decrease in the footprint of crime. Similarly, Fig. 3b shows the change pattern of the number of hotspots. It also shows a slight increase and decrease in the number of hotspots, but overall there is a decrease in the number of hotspots from 2001 to 2014. It is found in the study that crime is high, i.e. occurrence of crime, a number of hotspots and area covered by hotspot (footprint) is maximum in the year 2008.

Characteristic vector of change for each the cases is given as: Murder $< 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0 >$, Rape $< 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1 >$, Kidnapping $< 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0 >$, Dacoity $< 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1 >$ The value 1 shows the increase in crime, and value 0 indicates either no change or decrease in the crime. Similarity index for the footprint of the hotspot, for all the criminal cases, is shown in Table 1. For instance, similarity between footprint of the rape and murder is 0.3076 for the period 2001 to 2014. This means in the study period in 30.76% of the cases murder and rape occurs together. In the same way, Table 1 can be analyzed. It is found in the study that kidnapping and rape are linked with 61.53% and dacoity and murder are 53.84%.

6 Conclusions and Future Research Directions

This paper deals with detection of the hotspot H, determining the number of hotspots and measuring the coverage of the hotspot for crime for the years 2001–14 as per the www.data.gov.in. Hotspots are irregular in shape and vary in size. Change pattern of the footprints of the hotspot of various crimes over the period of 14 years was also analyzed. The analysis of pattern of change based on the characteristics vector of change and the similarity index is a novel contribution of this paper.

References

1. Chainey, S., Ratcliffe, J.: GIS and crime mapping. Wiley. (2013)
2. Short, M.B., Brantingham, P.J., Bertozzi, A.L., Tita, G.E.: Dissipation and displacement of hotspots in reaction-diffusion models of crime. Proc. Natl. Acad. Sci. **107**(9), 3961–3965 (2010)
3. Herrmann, C.R.: The dynamics of robbery and violence hot spots. Crime Sci. **4**(1), 33 (2015)
4. Loh, W.K. Park, Y.H.: A survey on density-based clustering algorithms. In: Ubiquitous information technologies and applications, pp. 775–780. Springer, (2014)
5. Shariff, S.S.R., Maad, H.A., Halim, N.N.A., Derasit, Z.: Determining hotspots of road accidents using spatial analysis. Indones. J. Electr. Eng. Comput. Sci. **9**(1), 146–151 (2018)
6. Bappee, F.K., Junior, A.S., Matwin, S.: Predicting crime using spatial features, (2018). arXiv:1803.04474
7. Chaikaew, N., Tripathi, N.K., Souris, M.: Exploring spatial patterns and hotspots of diarrhea in chiang mai, Thailand. Int. J. Health Geogr. **8**(1), 36 (2009)
8. Khalid, S., Wang, J., Shakeel, M., Nan, X.: Spatio-temporal analysis of the street crime hotspots in faisalabad city of pakistan. In: 2015 23rd International Conference on Geoinformatics, pp. 1–4. IEEE, (2015)
9. Eftelioglu, E., Tang, X., Shekhar, S.: Geographically robust hotspot detection: a summary of results. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW), pp. 1447–1456. IEEE, (2015)
10. Tabarej, M.S., Minz, S.: Rough-set based hotspot detection in spatial data. In: International Conference on Advances in Computing and Data Sciences, pp. 356–368, Springer, (2019)
11. Xie, Y., Shekhar, S.: A nondeterministic normalization based scan statistic (nn-scan) towards robust hotspot detection: a summary of results. In: Proceedings of the 2019 SIAM International Conference on Data Mining, pp. 82–90, SIAM, (2019)
12. Tran, T.N., Wehrens, R., Buydens, L.M.: Knn-kernel density-based clustering for high-dimensional multivariate data. Comput. Stat. Data Anal. **51**(2), 513–525 (2006)
13. Abughali, I.K., Minz, S.: Binarizing change for fast trend similarity based clustering of time series data. In: International Conference on Pattern Recognition and Machine Intelligence, pp. 257–267, Springer, (2015)

Itemset Mining Based Episode Profiling of Terrorist Attacks Using Weighted Ontology

Saurabh Ranjan Srivastava, Yogesh Kumar Meena, and Girdhari Singh

Abstract Itemset mining is a prominent research problem of data mining domain. It is extensively utilized for knowledge discovery in domains dealing with multi-component records or itemsets. Terrorism is a similar domain where every terrorist attack carries attack attributes such as location, target and attack type as components. Treatment of terrorist attack episodes as itemsets can facilitate effective pattern analysis and forecast of future attack episodes. This paper introduces a novel approach of mapping three major attributes of terrorist attacks taken place in a region in a single weighted ontology. The weighted ontology is later employed to discover and forecast useful information about possible attack episodes in the future.

Keywords Itemset mining · Episode · Ontology · Weighted ontology · Terrorism · Terrorist attack

1 Introduction

An ontology can be termed as a visual representation of the existence and relationships of various abstract entities belonging to a common domain of subject [1]. Developing an ontology tailored to the requirements of specific projects is more an art than a science in terms of implementation. The experience of the ontology designers plays a crucial role in capturing the complexity and diversity of the domain of subject for which the ontology has to be designed. Till date, numerous ontologies have been devised for domains such as network security, text analytics, e-commerce and so on [2]. Capturing adequate depth as well as breadth of the domain concept

S. R. Srivastava (✉) · Y. K. Meena · G. Singh
Department of Computer Science & Engineering, Malviya National Institute of Technology, Jaipur, India
e-mail: saurabhranjansrivastava@gmail.com

Y. K. Meena
e-mail: ymeena.cse@mnit.ac.in

G. Singh
e-mail: gsingh.cse@mnit.ac.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_31

always remains the primary challenge of an ontology designer. This means that balancing the coverage of each minute concept against the desired level of detail is always a challenge. Ontologies represent relationships of abstract classes belonging to their domain as combination of multiple units. Such combinations of units or items can be termed as itemsets in data mining applications. An itemset is practically an unordered assembly of one or more values of same or different attributes. On treating various specifications of a terrorist attack like location and attack type as attributes or items of a single itemset, modeling the profile of terrorist attacks and discovering patterns about upcoming attacks can be effectively performed by the use of ontologies. Specifically, by measuring the repetitions of various attributes in different terrorist attacks, ontological structures with weighted descriptions can be developed to address the problem of terrorist attacks. This paper works on the development of ontological techniques for weighted profiling of terrorist attacks.

2 Literature Survey

Exploration of relations or patterns of associations among collections of items (itemset) has remained a topic of frequent research. Multiple works on the discovery of association rules from closed itemset lattices [3], [4] are available. These works exploit graph-based structures of items called itemset lattices [5] to mine relations among items.

Complex and covert networks such as terrorist organizations can be efficiently analyzed by using graph structures and network analysis. Terrorist networks can be plotted as graphs representing their constituent groups or actors as nodes, while their links can be specified as edges of the graph. Various graph-based approaches for pattern extraction [6] over criminal and terrorist data have been proposed.

Graph or network analysis has been also used to develop solutions for predicting the behavior of terrorist organizations by treating their key actors [7] as component nodes of a graph. A terrorist network can be analyzed on the basis of features or attributes of each of its component nodes. Another method for analysis of terrorist networks [8] is the structure-based analysis that evaluates graphs on the basis of their semantic configuration, size and connectivity. A work of extracting terrorist events from print media [9] uses ontologies for comparing linguistic attributes on a similarity feature of term frequency–inverse document frequency (TF-IDF). This approach of terrorism analysis, based on terrorism ontological linguist features, showcased a recall rate of 85.15% over Thailand terrorism dataset.

Our method analyzes terrorist attacks against a set of attack attribute values. The occurrence of each attribute adds to the proximity of repetition of attack in the future.

3 Terms and Parameters

In this paper, we propose a novel approach for profiling and analysis of a set of values (itemsets) by mapping them as ontological graphs of a subject domain. Our approach works by treating these itemsets as systematic structures called episodes and mining information patterns from them.

3.1 Definitions

Class Label. A categorical or nominal value that serves as a class or container for pre-sampled values of a database making them suitable for analysis can be termed as a class [5]. The class label is similar to the term 'concept' in terminology of 'concept hierarchy' in data mining applications that groups items or values under predefined labels.

Itemset. It is a structural collection of items or values belonging to a certain class label [5]. Set of items or values that appear frequently together in a transaction data set is referred to as a frequent itemset. For example, consider a shopping store having the following three classes of items, namely fruits, dairy and bakery:

Fruits: {**A**pple, **B**anana}
Dairy: {**M**ilk, **Y**ogurt, **C**heese}
Bakery: {**D**oughnut, **P**astry}

Suppose, the only constraint on the purchase transactions from this store is that every transaction will have exactly one item from each class. This implies that every itemset originating from such transactions will be of exact length 3 having one item of every class. Samples of such transactions can be itemsets like {**A**pple, **M**ilk, **P**astry} and {**B**anana, **C**heese, **D**oughnut} and so on.

Ontology. An unambiguous specification of a set of classes relevant to a common domain can be termed as an ontology [1]. Ontologies are utilized for modeling of item—relationship problems in various domains of interests such as e-commerce, agent behavior modeling, natural language generation, text-based information extraction and knowledge management systems [2].

3.2 Weighted Ontology

Now, we propose our concept of weighted ontology that forms the base of our work. An ontology that presents the properties and the relations of its component set of concepts or class labels with a numerical value of importance or weightage can be termed as a weighted ontology. To illustrate this concept, we continue our example of shopping store. For a transaction database given in Table 1, we can generate weighted ontologies for **A**pple and **B**anana itemsets as given in Fig. 1.

Table 1 Sample transactions from shopping store transaction database

TId	Transactions with apple item in itemset	TId	Transactions with banana item in itemset
T1	{Apple, milk, doughnut}	T4	{Banana, yogurt, doughnut}
T2	{Apple, cheese, doughnut}	T5	{Banana, milk, pastry}
T3	{Apple, yogurt, pastry}	T6	{Banana, cheese, pastry}

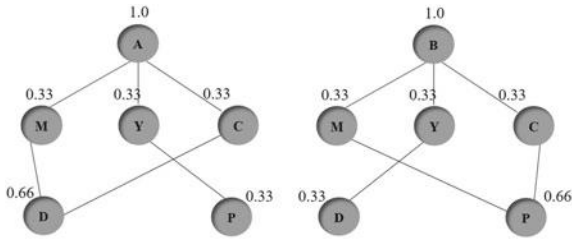


Fig. 1 Weighted ontologies for apple and banana itemsets

3.3 Episode Profile Modeling

Any occurrence of certain events in a combination is often referred to as an episode [10]. As discussed earlier, an itemset is a structural collection of labeled values, whereas an episode can be termed as an assembly of events preferably in an order of execution. If we ignore the order of occurrences of components, it implies that a transaction composed of an itemset can be also considered as an episode. In continuation of our shopping store example, both itemsets {**A**pple, **M**ilk, **P**astry} and {**B**anana, **C**heese, **D**oughnut} can be considered as purchase episodes on ignoring that which item was bought first.

Here, it should be noted that multiple episodes related to a domain can attain one or more events or components in common. For example, transactions T1 and T2 in Table 1 have **A**pple and **D**oughnut in common. Our approach maps the connected occurrence of every component of an episode to generate weighted patterns for analysis. In this paper, we assume that every transaction itemset (E: episode) is composed of multiple events (A: attribute). By utilizing these basics, now we define parameters that form the basis of our work proposed in this paper.

The occurrence of an attribute value of a class C against total occurrences of all attributes of its class is referred to as attribute support. In above example, out of T4, T5 and T6, milk, yogurt and cheese each have appeared once out of total three transactions leading to an attribute support of 1/3 = 0.33 for each. Similarly, both Apple and Banana appear in three out of three transactions of their sets, leading to an attribute support of 3/3 = 1.0 for both.

$$Attribute\ Support\ \sigma \mathrm{A}_{ci} = \frac{\text{occurance of attribute } A \text{ at class level } C}{\text{sum of occurances of all } i \text{ attributes at class level } C}$$

The sum of all component attribute supports of an episode E is termed as episode support. The episode support score of {Apple, Yogurt, Pastry} will be the sum of 1.0, 0.33 and 0.33 totaling 1.66, while to the episode support of {Banana, Milk, Pastry} will sum up to 1.99.

$$Episode\ Support\ \sigma\mathrm{E}_k$$
$$= \text{sum of } k \text{ component attribute support of an episode E at each class level } C$$

The possibility or proximity of occurrence of an episode can be computed as the total of attribute support of the relevant episode divided by the sum of all attribute supports at its class level C. Proximity score maps the expectancy of occurrence for an episode in terms of percentage.

$$Proximity\ Component\ Score\ \rho\mathrm{E} = \sum_{E=1}^{n} \frac{\text{attribute support of an episode E}}{\text{sum of all attribute supports at class level C}}$$

Now in the next section, we elaborate on the types of episodes that can be mined on the basis of the parameters discussed above.

3.4 Types of Episodes

On the basis of occurrence and repetition, very episode (transaction) can be considered as either occurred already or possible in future. For exploring the pattern of occurrences among episodes and generating forecasts from them for upcoming episodes, we identify and label them under one of the four specific categories.

Observed Episode ($\mathrm{E_O}$): Itemsets or episodes that have been previously observed are put under this category. This implies that the attribute support of all the connected nodes of the episode E is above zero ($\sigma A_{CE} > 0$).

Projected Episode ($\mathrm{E_P}$): Episodes that are not present in the transaction database, but are visible in the ontology graph without any additional connectivity of any attribute nodes, fall under this category. Episodes not observed yet but depicting a high proximity score are denoted under $\mathrm{E_P}$.

Drawn Episode ($\mathrm{E_D}$): If connecting existing attribute nodes in the ontology graph draws new possible episodes, then such episodes will be classified as drawn episodes ($\mathrm{E_D}$). This category includes all the possible episodes that may evolve with existing set of attributes under given class labels.

Appended Episode ($\mathrm{E_A}$): Appending a new attribute node under a given class label and connecting it to remaining attribute nodes of other classes generates appended episodes.

It is category of rare and nonexistent episodes which in part or entirety are observed in other transaction databases. Generally, attribute nodes imported from other transaction databases of similar domain become the part of appended episodes. We propose every episode of any subject domain to belong to at least one of the above-mentioned categories. Examples of each of these episode categories are discussed ahead in detail.

4 Weighted Ontologies of Terrorist Attack Episodes

Every terrorist attack attains a set of attributes that generate a pattern such as location of attack, target or victim, type of attack performed and so on. Mapping of certain attack attributes into an itemset or episode can depict a terrorist attack of any of the above-mentioned categories and can be employed to mine useful patterns. Our approach considers both frequent as well as rare itemsets for profiling of terrorist attack episodes. To implement our proposed approach, we have used a subset of Global Terrorism Database (GTD) [11] data. From GTD, we have fetched data of terrorist attacks by Al-Qaeda in three countries, namely Mali, a sub-Saharan African nation, Yemen, a Middle Eastern country of North Africa and USA. We utilized the attack episode data of two most terror inflicted regions of Mali and Yemen each: Gao, Timbuktu, Abyan and Shabwah for the time span of 1990–2017 to implement our proposed approach. We also partitioned the attack episode data of USA into 70% of training data and 30% test data to forecast attack episodes in New York for a time span of 3 years (2016–2018). Referring to the attack attributes of GTD with highest weightage [12], we have used location, target and attack type attributes of terrorist attacks to showcase our approach. The terrorist attacks occurred in these four regions have been mapped into their individual weighted ontologies of terrorist episodes.

These weighted ontologies are employed to place every terrorist episode in its relevant category. Subsequently, support values and proximity scores are computed to forecast the possibility of occurrence or repetition of these attack episodes. Treating location, target and attack type as concepts of attributes or class labels, to present the attribute nodes in the weighted ontologies, we have assigned respective abbreviation code to every attribute node as presented in Table 2. Every attack episode is presented as a chain of attribute nodes with a relationship label (AT, ON, BY) over the connecting edge. This implies that a suicide attack on military on a checkpost location in Abyan can be depicted as presented in Fig. 2.

Out of the total 69 terrorist attacks occurred in Yemen during the time span under analysis of this work, nine attacks occurred in Abyan, leading to a support value of 0.13. Among those nine attacks, two have taken place on military check post locations resulting into a support score of 0.22. But, in addition to check posts, attacks on military have taken place on other locations also totaling to five attacks that cause a higher support value of 0.55. Finally, out of these nine attacks, three suicide attacks result into a support score of 0.33. All these attribute support values sum up to an episode score of 1.23. Similarly, all episodes of each region are mapped into their respective weighted ontology graphs as shown in Fig. 3.

Now, we analyze the episodes of terrorist attacks by utilizing these weighted ontologies and categorize them under the four categories discussed earlier.

Table 2 Codes for various class labels of terrorist episode ontologies

CODE	Location	CODE	Target/victim	CODE	Attack types
CKP	Check post	BSN	Business	AMB	Ambush
OFC	Office	CVL	Civilian	ASL	Assault
OPN	Open	FWR	Foreign worker	ASN	Assassination
PUB	Public	GOV	Government	ATM	Attempt/Conspiracy
REL	Religious	MIL	Military	BOM	Bombing
TRP	Transport	POL	Police	FRG	Firing
UNK	Unknown			HST	Hostage
				KDP	Kidnap
				SCD	Suicide
				STB	Stabbing
				VRM	Vehicle ramming

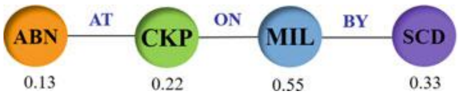


Fig. 2 Sample weighted ontology for terrorist attack episode in Abyan

Terrorist Attack itemsets of Observed Episodes (E_O).
Table 3 displays the attack episodes from all four regions that have occurred already. Their episode support score and related proximity component are also presented along with each episode which maps the proximity of its repetition in the future. For example, the attack episode of {GAO, CHECKPOST, MILITARY, BOMBING} has occurred 4 times just in years 2016 and 2017, making it a frequently expected episode with highest proximity of 47.12%.

Terrorist Attack itemsets of Projected Episodes (E_P).
In addition to the attacks occurred already, the connections of attribute nodes in the ontologies also disclose possible episodes which have not taken place yet. But their proximity component scores indicate high chances of their occurrence in the future. The attribute nodes of {ABYAN, OPEN, MILITARY, FIRING} itemset get connected due to mapping of other observed episodes, but display a proximity of 32.10% of occurrence in the future.

Terrorist Attack itemsets of Drawn Episodes (E_D):
On connecting the nodes of {PUBLIC, CIVILIAN} and {CIVILIAN, SUICIDE} chains, an attack episode of proximity 35.31% becomes visible in Shabwah region {SHABWAH, PUBLIC, CIVILIAN, SUICIDE}. Such nonexistent but possible episodes can be derived by systematically connecting disjoint nodes at each class level and computing their support scores.

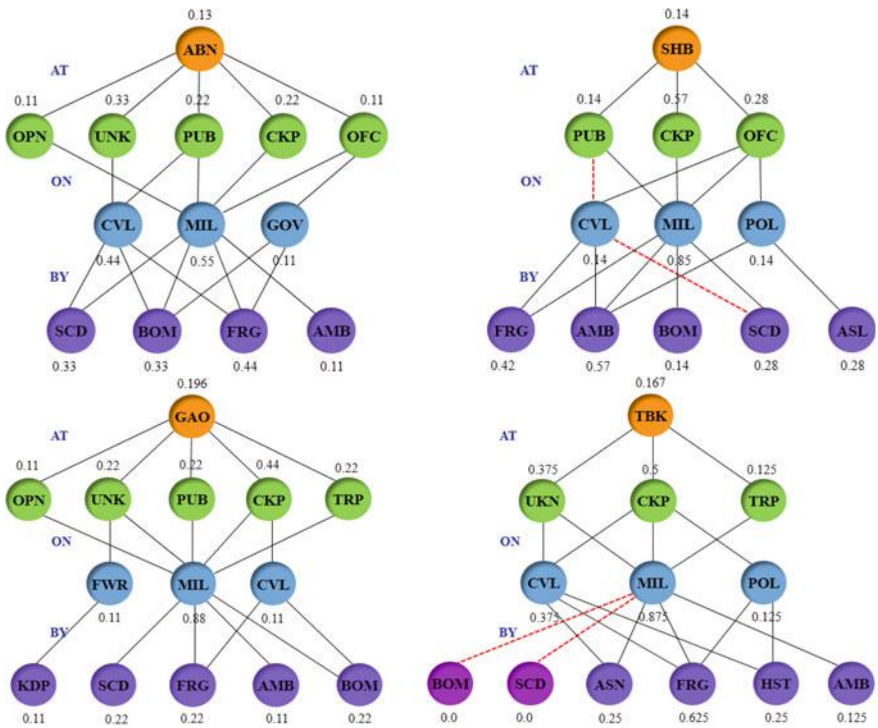


Fig. 3 Weighted ontologies for terrorist attack episodes in Abyan, Shabwah, Gao and Timbuktu

Terrorist Attack itemsets of Appended Episodes (E_A):
The impact of an attack occurred in a distant region is obvious in the region under consideration. Terrorist groups learn and update their methods and tactics from other attacks [13]. For example, suicide attacks and bombing at targets are the major attack methods in other regions of Mali. Undoubtedly, the chances of use of these methods cannot be ignored in Timbuktu region where such attacks have not taken place under the considered timeline. Hence, the addition of new attribute nodes from other datasets of similar domain can uncover upcoming episodes of surprising nature in the future.

5 Results and Discussions

To establish the efficiency of our approach, we now consider the case study of terrorist attacks that took place in New York City of USA during the year 1990–2015 and test the episode rules against the attacks of the year 2016–2017. These occurrences or support values have been utilized to construct the weighted ontology of terrorist episodes in New York City presented in Fig. 4. The prominent terrorist attack episodes

Table 3 Terrorist episodes with episode category, support score and proximity component

S.no	Attack episode itemsets	Episode category	Support score	Proximity component (%)
1	{GAO, CHECKPOST, MILITARY, BOMBING}	E_O	1.736	47.12
2	{TIMBUKTU, CHECKPOST, MILITARY, FIRING}	E_O	2.167	42.04
3	{ABYAN, PUBLIC, MILITARY, BOMBING}	E_O	1.23	35.87
4	{SHABWAH, PUBLIC, MILITARY, SUICIDE}	E_O	1.41	41.06
5	{GAO, TRANSPORT, MILITARY, BOMBING}	E_P	1.516	41.06
6	{GAO, OPEN, MILITARY, BOMBING}	E_P	1.406	38.03
7	{ABYAN, OPEN, MILITARY, FIRING}	E_P	1.23	32.10
8	{SHABWAH, PUBLIC, CIVILIAN, SUICIDE}	E_D	0.7	35.31
9	{TIMBUKTU, CHECKPOST, MILITARY, BOMBING}	E_A	1.542	25.38
10	{TIMBUKTU, CHECKPOST, MILITARY, SUICIDE}	E_A	1.542	25.38

of New York City highlighted in this ontology graph are in conformance to the following categories of episodes (Table 4).

Observed Episode (E_O): From September 17–19, 2016, in a series of bombings on civilian targets, a pressure cooker bomb filled with shrapnel was detonated in the crowded public space of Sixth and Seventh Avenue of West 23rd Street in New York and New Jersey [12]. This attack was an observed repetition of bombing on a US Armed Forces recruiting station in Times Square by an improvised explosive device (IED) on March 6, 2008.

Appended Episode (E_A): On August 17, 2017, an attacker rammed a van on a pedestrian crowd in Barcelona (Spain) causing death of 13 people and injuring 150

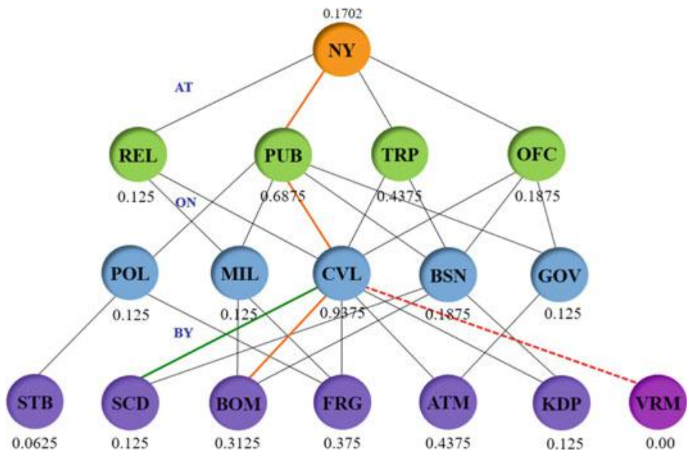


Fig. 4 Weighted ontology of terrorist attack episodes for New York

Table 4 Terrorist attack episodes with corresponding episode category, support score and proximity component for New York

S.no	Attack episode itemsets	Support score	Proximity component (%)	Episode category	Date of attack
1	{NEWYORK, PUBLIC, CIVILIAN, BOMBING}	2.1077	44.01	(E_O)	17–19 SEP 2016
2	{NEWYORK, PUBLIC, CIVILIAN, RAMMING}	1.7952	36.77	(E_A)	31 OCT 2017
3	{NEWYORK, PUBLIC, CIVILIAN, SUICIDE}	1.9202	39.67	(E_D)	11 DEC 2017

others [14]. The selection of weapon and perpetrator involved in this attack makes it a special case for analysis. Soon after this attack in Spain, on October 31, 2017, a pickup truck drove into a crowd of pedestrians and cyclists at West Side Highway and West Street section, starting from Houston Street and finally collided with a school bus at Chambers Street in New York [15].

The attackers in both attacks used a vehicle as a non-obvious weapon of attack. But in spite of the absence of any explosive and involvement of two different terrorist groups, selection of civilian crowd as a common target on street was common. This can be concluded as the learning of attackers from the Barcelona attack, which leads to inclusion of ‘vehicle ramming’ as a new attribute node under ‘attack type’ class.

Drawn Episode ($\mathbf{E_D}$): A Bangladeshi suicide bomber attempted to detonate a pipe bomb and a battery pack attached to himself on December 11, 2017, in the New York City Subway's Times Square–42nd Street/Port Authority Bus Terminal station. This previously nonexisting attack episode was drawn from connection of attribute nodes {CIVILIAN, SUICIDE} itemset of 09/11 attacks on World Trade Center [16] to {PUBLIC, CIVILIAN} itemset. This case study presents a strong evidence of the efficiency of our proposed approach. If executed exhaustively by analyzing all possible combinations of existing and new attribute nodes, this approach can provide a potential forecast solution.

6 Conclusion and Future Work

In this paper, we have demonstrated that by treating itemsets as a combination of constituent attributes, they can be modeled into episodes. We have also presented a new approach of generating weighted ontological graphs for analysis and profiling of episodes in a subject domain. To prove our approach, we executed it on terrorist attack data of Abyan, Shabwah, Gao and Timbuktu regions of Yemen and Mali nations fetched from Global Terrorism Database (GTD). We also tested this approach on terrorist attack data of New York City for the year 2016–2017. The support of an item also contributes to support of the itemset or episode. However, the frequency of support can alter the impact of the episode according to the nature of the problem. Hence, a rare attribute value of low support can also lay a high impact on proximity of episode. It is also shown that episode support score and proximity of occurrence are in direct proportion to each other. Connection of existing but disjoint attribute nodes and addition of new attributes from other episode datasets can generate forecasts about upcoming potential episodes of high impact. In the future, we aim to improve the computational efficiency of the approach as an effective forecasting solution by using mechanisms like directed graphs, possibility logic and uncertainty theory.

References

1. Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: principles and methods. Data Knowl. Eng. **25**(1–2), 161–197 (1998)
2. Gómez-Pérez, A., Corcho, O.: Ontology languages for the semantic web. IEEE Intell. Syst. **17**(1), 54–60 (2002)
3. Pasquier, N., Bastide, Y., Taouil, R., Lakhal, L.: Efficient mining of association rules using closed itemset lattices. Inf. Syst. **24**(1), 25–46 (1999)
4. Zaki, M.J., Hsiao, C.J.: Efficient algorithms for mining closed itemsets and their lattice structure. IEEE Trans. Knowl. Data Eng. **17**(4), 462–478 (2005)
5. Han, J., Pei, J., Kamber, M.: Data mining: concepts and techniques. Elsevier. (2011)
6. Phillips, P., Lee, I.: Mining co-distribution patterns for large crime datasets. Expert Syst. Appl. **39**(14), 11556–11563 (2012)

7. Berzinji, A., Kaati, L., Rezine, A.: Detecting key players in terrorist networks. In: 2012 European Intelligence and Security Informatics Conference. IEEE, pp. 297–302 (2012)
8. Will, U.K., Menon, N., Karampelas, P.: Detecting new trends in terrorism networks. In: International Conference on Advances in Social Networks Analysis and Mining (2010)
9. Inyaem, U., Meesad, P., Haruechaiyasak, C., Tran, D.: Ontology-based terrorism event extraction. In: 2009 First International Conference on Information Science and Engineering. IEEE, pp. 912–915 (2009)
10. Mannila, H., Toivonen, H., Verkamo, A.I.: Discovering frequent episodes in sequences. In: KDD 1995, vol. 95, pp. 210–215 (1995)
11. LaFree, G., Dugan, L.: Introducing the global terrorism database. Terror. Polit. Violence. **19**(2), 181–204 (2007)
12. Tutun, S., Khasawneh, M.T., Zhuang, J.: New framework that uses patterns and relations to understand terrorist behaviors. Expert Syst. Appl. **78**, 358–375 (2017)
13. Santora, M., Goldman A.: Ahmad Khan Rahami was inspired by Bin Laden, charges say. The New York Times, 01 November 2017. https://www.nytimes.com/2016/09/21/nyregion/ahmad-khan-rahami-suspect.html. Accessed 14 Aug 2019
14. Carrera, S., Guild, E., Mitsilegas, V.: Reflections on the terrorist attacks in Barcelona: constructing a principled and trust-based EU approach to countering terrorism. CEPS Policy Insight 2017–2032 (2017)
15. Barron, J.: What We Know and Don't Know About the Attack in Lower Manhattan. The New York Times, 31 October 2017. https://www.nytimes.com/2017/10/31/nyregion/nyc-shooting-truck-attack.html. Accessed 14 Aug 2019
16. Krebs, V.E.: Mapping networks of terrorist cells. Connections. **24**(3), 43–52 (2002)

Regulatory Technology (RegTech) and Money Laundering Prevention: Exploratory Study from Bahrain

Meaad Turki, Allam Hamdan, Jasim Al Ajmi, and Anjum Razzaque

Abstract This study aims to illustrate the impact of adopting RegTech innovations in banks on their money laundering prevention programs. The banking industry changed massively after the financial crisis of 2008. New regulations and enforcements are being imposed on banks causing the compliance cost to increase dramatically. RegTech has been invented by technology firms acting as a potential solution to banks. The study will demonstrate the ability of RegTech to reduce the compliance cost, strengthen money laundering prevention and reduce the reputational risk in banking sectors. This study target sample was banks' employees in Bahrain because of their proper knowledge about anti-money laundering. Data were collected from 100 respondents from the banking sector of Bahrain. Descriptive analysis was used to analyze the data while regression model and Spearman's correlation were used to test the hypothesis. The results of this analysis indicate that RegTech has positive impact on strengthening and enhancing money laundering prevention in banks. The findings of the study will help banks understand the effectiveness of RegTech solutions, raise bankers' awareness about the new technologies and provide insight for regulators about RegTech capabilities in preventing money laundering.

Keywords Regulatory technology (RegTech) · Anti-money laundering · Bahrain

1 Introduction

For more than 40 years, Bahrain has been recognized as the financial center in the Middle East. The financial industry is considered a key component of Bahrain's economy, representing more than 17.5% of GDP. Protecting this sector from money laundering threats has been the objective of the government and the Central Bank of

A. Hamdan (✉) · J. A. Ajmi · A. Razzaque
Ahlia University, Manama, Bahrain
e-mail: allamh3@hotmail.com

M. Turki
Conventional Wholesale Bank, Compliance Assistant, DMLRO, Inspection Officer, Manama, Bahrain

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_32

Bahrain (CBB). Money laundering is the process of hiding illegal sources of money and making them appear legal. It is a process of transforming dirty money to clean money. Money laundering can proceed through three stages: placement, layering and integration. Bahrain is more likely to be used in the layering stage of money laundering rather than in other stages. Its relatively large financial sector with a network of foreign branches makes it particularly vulnerable to this risk (FATF 2006). Fighting money laundering requires government intervention. In Bahrain, Decree Law No. (4) of 2001 "Anti-Money Laundering Law" imposes obligations generally in relation to the prevention of money laundering to all residents in Bahrain. Banks are therefore under the statutory obligations of that law (CBB Rulebook). The ability to counter money laundering effectively remains challenged by a variety of factors; these include introduction of new and emerging threats (e.g., cyber-related financial crimes); gaps in regulatory regimes, including uneven availability of technical assistance for anti-money laundering purposes; and the costs associated with banks compliance with global anti-money laundering guidance [6]. The anti-money laundering compliance has changed massively over the years by applying stricter rules and regulations, in addition to numerous increases in enforcements and fines imposed on banks as a result of breaches. In September 2018, it has been revealed that regulators around the world have imposed $26 billion fines on banks for non-compliance with anti-money laundering, know your customer (KYC) and sanction regulations in the last decade [3]. As time passes, technology is advancing. Criminals and banks are taking the most advantage of it for different purposes. Banks fought money laundering for years using several methods and technologies in order to fulfill their obligation. Recently, we witnessed the merge of two terms finance and technology to create nowadays familiar term, i.e., financial technology "FinTech" which means the combination of both innovation and technology into delivering financial products and services. FinTech has a subset that is recognized as regulatory technology "RegTech"; it depends on technologies to enable meeting the regulatory requirements more effectively and efficiently; hence, it is being a valuable element in preventing money laundering. Banks have shown massive interest to use new technologies as they act proactively instead of wait and see approach in their effort to fight money laundering. Regulators also support advanced technology as an effective mechanism in responding to money laundering evolving issues. Complying with anti-money laundering rules, monitoring customer transactions, customer onboarding and know your customer is a very costly, complicated and time-consuming process. However, through innovation from the RegTech arena, it is highly expected by field specialists that RegTech will provide efficient compliance solutions in terms of monitoring, lowering cost, better analysis and reducing associated risks.

As we are moving forward, it is noticeable how money launderers are creating more innovative and new methods to launder their money by using the latest technologies. Despite the advantages of FinTech involvement in the banking system, it has also areas of vulnerability and residual risks that need more stringent control. Anti-money laundering compliance is a critical function for banks to guard against the risk of being unwittingly used in money laundering as the cost of non-compliance becomes enormously high. Several banks were fined massive penalties for violating anti-money laundering regulations and exposed to reputational risk.

2 Research Methodology

2.1 Population and Sample Size

Quantitative method was used to conduct the study. This study targeted bank's employees because of their awareness and knowledge about money laundering prevention. The banking sector in Bahrain has 7447 employees in 2017 as per CBB manpower survey. Staffs in banking sector are being educated and trained regularly about the importance of combating money laundering. Therefore, they will be able to recognize the impact of regulatory technology on money laundering prevention. The sample size for this study is 100 employees from banks to represent the population and contains most fields in the banking sector, mainly focusing on money laundering prevention specialists.

3 Descriptive Analysis

3.1 Descriptive Analysis of Demographic Variables

This part of the research is intended to analyze the demographic data collected from the questionnaire's first section. The results obtained from 100 respondents are summarized in the below Table 1 by gender, age, experience and position, whereas Table 2 outlines the awareness of RegTech in Bahrain and the enforcement actions that were taken against banks as a result of their violations.

As shown in Table 1 that presents the results of demographic data regarding gender, a total of 62% of the sample size were male who represent respondents' majority, while female was 38% only. Unsurprisingly, the result agrees with CBB's manpower report of year 2017 that indicates 4,910 males are working in banks, representing 66% of the total workforce and the remaining 34% are females. With respect to participants' age, 87% are below the age 40. Results show that there are 50% of bankers aged between 30 and 40 years, 37% under 30 years, while 12% are between 41 and 50 years, and only one participant is above 50 years. Obviously, most employees in banks are young since banks are focusing on recruiting younger generation in order to adapt with the rapid changes in the industry and their ability to interact with emerging technologies. In terms of experience, the majority of the questionnaire respondents (32%) have banking experience ranged between 5 and 10 years, 28% between 11 and 15 years, 26% below 5 years and 14% of bankers are above 15 years of experience. Noticeably, results were almost collected equally from different level

Table 1 Demographic data

	Response category	Frequency (N = 100)	Percentage (%)
Gender	Male	62	62
	Female	38	38
Age	Under 30 years	37	37
	30–40 years	50	50
	41–50 years	12	12
	Above 50 years	1	1
Experience	Below 5 years	26	26
	5–10 years	32	32
	11–15 years	28	28
	Above 15 years	14	14
Position	Front office staff	28	28
	Operations	16	16
	Compliance/AML	36	36
	Audit	5	5
	Others	15	15

Table 2 General data

	Response category	Frequency (N = 100)	Percentage (%)
RegTech Awareness	Yes	47	47
	No	53	53
Enforcement against bank	Yes	25	25
	No	75	75

of banking experiences, and above 42% of the participants have extensive professionalism in banks. This fact strengthens the effectiveness and reliability of the research outcomes and will provide proper reflection of bankers' perspective toward regulatory technologies. In regard to the sample's job roles, the top two banking roles for respondents were compliance/anti-money laundering officer (36%) and front office staff (28%), while operation and others were fairly even at 16% and 15%, respectively, and only 5% were having audit role. Accordingly, more than one-third of the participants are working directly in anti-money laundering field and above one quarter are performing front office activities which include interacting with customers and conducting know your customer requirements. Consequently, the study results will represent compliance specialists' view on RegTech impact on money laundering prevention and also variety of opinions from other roles addressing the same.

The results indicated a close to 50/50 split about the knowledge of the term RegTech. Regardless of this finding, the term is expanding rapidly nowadays in

Bahrain's daily newspapers, regulator's publications and among banks' employees. Furthermore, RegTech just emerged recently during the past two years, and it is well known by anti-money laundering staff particularly. Likewise, an assessment was conducted to display the seriousness of anti-money laundering topic in banks. Exactly, one quarter on the respondents confirmed that actions were brought against their banks due to violations concerning money laundering. Unsurprisingly, the cost of non-compliance increased rapidly globally. In Bahrain, the regulator believes that a successful supervision over the banks activates requires also stringent enforcements in case they fail to comply with requirements. Just recently, CBB amended its methodology for calculating financial penalties in July 2018.

3.2 Descriptive Analysis of Variables

The following Table 3 illustrates the percentages of each response to the statements, mean and standard deviation.

The first independent variable presents the effectiveness of conducting know your customer requirements electronically. Five statements were set for the above variable, and majority of opinions belong to agree with mean ranging from 3.58 to 4.09 and standard deviation between 0.922 and 1.093. This determines that most of respondents established that electronic know your customer has significant effect on money laundering prevention. It is noted that the first statement "deficiencies in filling KYC

Table 3 Electronic know your customer variable analysis

	Frequency %						
Statements	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)	Mean	Std. deviation
1	6	8	29	36	21	3.58	1.093
2	3	4	17	45	31	3.97	0.958
3	1	5	17	38	39	4.09	0.922
4	2	7	17	40	34	3.97	0.989
5	5	6	20	44	25	3.78	1.05

1. There are deficiencies in filling KYC form manually as some mandatory information may left out blank by client
2. Automation that eliminates deficiencies in collecting required information from customer will strengthen the KYC process
3. Obtaining customer data from government sources through automation can strengthen data reliability and verification requirements
4. Updating KYC information electronically using government portal improve KYC process effectiveness
5. Having inaccuracies in the information collected from customer may lead to your bank being used to launder funds and a regulatory penalty

form manually" and fifth statement "Inaccuracies in customer information may lead to money laundering and penalty" had the lowest mean (3.58 and 3.78, respectively) and highest standard deviation (1.093 and 1.05, respectively) which could indicate bankers' disagreement in opinion to these observations. This may be interpreted as some participants are unaware of the effect of having improper customer information. Supported by [8], HSBC acknowledged its failure in conducting proper know your customers on its account holders and agreed to pay huge fine to regulatory authorities. Also [4] stated that not having enough information or having inaccuracies in the information that bank collect may create a domino effect that lead to an institution being used for money laundering and a subsequent penalty. However, 36 and 21% of the respondents were agreed and strongly agreed of the first statement, whereas 44 and 25% were agreed and strongly agreed for the fifth. Majority of the participants strongly agreed (39%) and agreed (38%) the third observation in this section which is strengthening customer data reliability and verification requirement by obtaining information from government sources using automation. It scored the highest mean 4.09 and lowest standard deviation 0.922. With a mean equal to 3.97 and standard deviation 0.989, the respondents also affirmed the effectiveness of updating customer information electronically by exploiting government portal on improving the KYC process as 34% of them were strongly agreed and 40% agreed the remark. Moreover, results revealed 45 and 31% of the respondents agreed and strongly agreed, respectively, that eliminating deficiencies in collecting know your customer information by using automated system will strengthen the process. It is recognized that RegTech has the ability to eliminate the deficiencies in KYC implementation as bank will obtain all required data from an official and reliable source. Reported by [1], sound KYC policies and procedures are critical in protecting the safety and soundness of banks and the integrity of banking system.

As illustrated above in Table 4, five statements were designed to measure respondents' perspectives regarding the independent variable "transaction monitoring." Overall, findings revealed that the standard deviations were below 1 except for the fifth statement. It indicated a high level of reliability within the responses. Furthermore, a minimum mean of 3.80 was scored for all statements which made abundantly clear that participants confirmed the importance of automations and advanced systems on monitoring banking transactions. The first observation was about detecting suspicious transaction accurately by analyzing vast volumes of transactions using an improved system. About 75% of the respondents have answered the first statement affirmatively. It scored a mean equal to 3.9 and 0.927 for standard deviation. Reuters [8] US survey stated that the bar for transactions monitoring was raised by US authorities as they are demanding more stringent and sophisticated technology for managing anti-money laundering risk. The second statement was formatted to explore participants' aspect regarding the ability of advanced checking systems in helping banks comply with economic sanctions. There was greater consensus by 79% of the respondents. The statement had also the highest mean of 4.02 and the second lowest standard deviation 0.876. The results confirmed by [8] US survey which found that 64% of respondents always monitor their customers by automated systems and 39% use an automated system for screening. It is little wonder that 52%

Table 4 Transaction monitoring variable analysis

	Frequency %						
Statements	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)	Mean	Std. deviation
1	3	4	18	50	25	3.9	0.927
2	2	3	16	49	30	4.02	0.876
3	3	2	20	52	23	3.9	0.882
4	2	4	19	53	22	3.89	0.863
5	2	12	18	40	28	3.8	1.044

1. Improved analytics for vast volumes of transactions to identify abnormal patterns can help detect suspicious activity more accurately
2. Advanced system that checks accounts against watch-lists, screen transactions for sanctions, can effectively help banks comply with economic sanctions
3. Automation that enables access to authority's databases to conduct background screening/criminal record will enhance bank's AML risk assessment
4. Improved data analysis that provides smart prediction and enables banks to visualize customer behavior will help banks to act proactively
5. It is nearly impossible to monitor transactions without the support of an automated system

of banks employees agreed and 23% strongly agreed that risk assessment will be enhanced by automation that enabled access to authority's databases for criminal screening. However, 20% were neutral and 5% expressed their disagreement. The third statement obtained a mean of 3.9 and 0.882 for standard deviation. About 75% of the respondents have also exhibited their confirmation to the fourth statement. They agreed by 53% and strongly agreed by 22% that visualizing customer behavior and providing smart prediction will assist banks to act proactively. The statement scored a mean of 3.89 and standard deviation 0.863. It finds that bankers are willingly supportive to the idea of having developed and smart data analysis. The last statement had the highest standard deviation of 1.044 but a reasonable mean of 3.8. It examined the respondents' views regarding the impossibility of monitoring transactions without automation. Despite the agreement of 68% of the bankers, 14% of the participants expressed their disagreement, whereas 18% remained neutral. The reason behind these differences could be related to their long years of experience in the banking industry when they were not depending much on technology, since majority of disagreed are having more than 11 years of experience (Table 5).

In order to measure the last independent variable which is cost and time, five statements were used. All statements were confirmed by the respondents as the minimum mean value was 3.77 and the maximum standard deviation was 0.92 which is below 1, indicating a good consistency, in which majority of respondents agreed to all statements with 51% and above. More than 70% of the participants stated their agreement on new technologies ability of cutting the cost of money laundering prevention in banks. Particularly, it was strongly agreed by 19%, agreed by 53% and about 20% were neutral. Also the mean and standard deviation of the first statement were 3.82

Table 5 Cost-and-time variable analysis

	Frequency %						
Statements	Strongly disagree (1)	Disagree(2)	Neutral (3)	Agree (4)	Strongly agree (5)	Mean	Std. deviation
1	1	7	20	53	19	3.82	0.857
2	2	1	18	53	26	4	0.816
3	4	3	23	52	18	3.77	0.92
4	1	1	11	60	27	4.11	0.709
5	0	8	21	51	20	3.83	0.842

Statements

1. New technologies (advanced software) can help banks cut the total cost of money laundering prevention
2. Automated system for detecting suspicious activities that reduces false-positive alerts will help AML specialist executing alerts in short time
3. Automation submission of suspicious transaction reports enabled authorities to receive reports in real time
4. Information collected in real time by KYC automation reduces both cost and time
5. The rapid changes in regulatory environments and fines imposed on banks cause the cost of money laundering prevention to increase

and 0.857, respectively. These results led to strongly believe that technology helps reducing the cost. The acceptance level of the second statement was high at a mean of 4, and standard deviation of 0.816 that reducing false-positive alerts by an automated system will shorten bankers' time. Findings showed that 53% of respondents agreed and 26% strongly agreed, which express bankers need to RegTech solutions that aims to enhance productivity by cutting the time and effort consumed on alerts. Similarly, majority of the respondents 52% agreed the third statement which was exploring bankers' opinion regarding submitting suspicious report to authorities in real time by using automation. It was also strongly agreed by 18% of the participants; however, another 23% were neutral. With a scored mean of 3.77 and standard deviation of 0.92, the results support the statement that technology speeds the submission process and leads to efficiency. The fourth inquiry regarding testing both cost and time by collecting KYC in real time through automation, quite a number of participants 60% expressed their agreement. Moreover, 27% of bankers strongly agreed the fourth statement, it is important to note that the statement had the highest mean of 4.11 and the lowest standard deviation of 0.709, which reflects respondents support to automate know your customer requirement in order to receive customers' information in real time. The final statement was formed to address the cost increase of money laundering prevention due to the rapid changes in the regulatory system. Majority of participants 51% agreed, while 20% strongly agreed. A mean of 3.83 and standard deviation of 0.842 confirmed that money laundering prevention cost is increasing. In UK [5] reported that 63% of their survey respondents confirmed an

Table 6 Money laundering prevention variable analysis

	Frequency %						
Statements	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)	Mean	Std. deviation
1	3	4	16	46	31	3.98	0.953
2	1	1	14	57	27	4.08	0.734
3	2	10	29	40	19	3.64	0.969
4	3	1	22	46	28	3.95	0.903
5	2	2	17	48	31	4.04	0.864

Statements

1. Know your customer is one of the best defenses a bank can maintain to guard against the dangers of money laundering
2. Monitoring system that detects suspicious transaction in timely manner is a major tool to combat money laundering
3. Penalties and enforcement imposed on banks are sign of weakness in anti-money laundering program
4. Advanced technology is an essential component for banks against the new emerging money laundering threats
5. Reliable money laundering prevention program guards your bank against regulatory and reputational risk

increase in their compliance cost. This raise was driven by an increase in the regulation volumes which caused banks to recruit more staff and invest in technology, along with the cost of enforcement actions.

In summary, the outcomes of this section supported RegTech solutions as an effective tool for reducing both cost and time of money laundering prevention program as a minimum of 70% of the participants agreed on how sufficient RegTech is for future banking system.

Table 6 illustrated the participants' responses toward the dependent variable "money laundering prevention" in banks. In general, the outcomes showed a mean above 3.64 and standard deviation below 1 which indicate respondents' consistent judgment over all these section statements. Participants expressed their opinion by 46% agreeing and 31% strongly agreeing with a mean of 3.98 and standard deviation of 0.953 in regard to the best defense in banks against money laundering prevention is KYC, which shows strong relationship between money laundering prevention and know your customer. Supported by [2], this result matched with experts' opinions that sound know your customer procedure is the best way to prevent money laundering. The second statement measured importance of monitoring systems that detects transaction in timely manner. Almost all bankers stated their agreement with 57%, strongly agreed 27%, while only two bankers expressed their disagreement. The statement had also the highest mean 4.08 and lowest standard deviation 0.734 confirming the positive impact of advanced monitoring system on money laundering prevention. This finding was in line with [7] paper that emphasized on the importance of a well-designed monitoring system as a vital element of an effective money laundering prevention which will help banks detect suspicious transaction and report them to regulatory authorities. The imposition of enforcement actions on banks by regulatory authorities is surely a sign of their failures in complying with regulation. Thus, the third statement questioned the respondents if the penalties imposed on banks are sign of weakness in their prevention program, which resulted in a mean of 3.64 and standard deviation 0.969. The outcomes found that 40% of participants agreed and 19% strongly agreed, whereas 29% were neutral and 10% expressed their disagreement, hence concluding that once enforcement is imposed on a bank, it is an indication that deficiencies were exist in their anti-money laundering program. The fourth aspect examined the impact of advanced technology on new money laundering threats, which had a mean of 3.95 and a standard deviation of 0.903. This was because 46% of respondents agreed and 28% strongly agreed. The results showed technology as the most significant tool that is capable to fight the new methods of money laundering.

Similarly, the last statement has the majority vote for agreed and strongly agreed with 48% and 31%, respectively, which shows that sufficient money laundering program protects banks from reputational and regulatory risk. The mean was 4.04, and the standard deviation was 0.864. The statement was about reliable money laundering prevention which can guard bank against regulatory and reputation risk. These findings summarize that banks can strengthen money laundering prevention by an effective know your customer procedures, efficient monitoring system for detecting and reporting suspicious transactions and adopting advanced technologies. Moreover, a stringent prevention program will protect banks against regulatory and reputation risk.

4 Conclusions

The main objective of the study is to determine the impact of regulatory technology on money laundering prevention in banks and demonstrate the relationship between adopting RegTech and money laundering prevention's cost and time. The study aims to measure the impact of RegTech on know your customer, transaction monitoring, cost and time. The descriptive results provide some insights into perceptions of the respondents—a majority agreeing or strongly agreeing with all the statements with respect to all of the independent variables—eKYC, transaction monitoring, cost and time—being linked to effective compliance and AML practices. However, some nuances are discernible. With respect to KYC, greater importance is attached to linking up with government sources of information as well as to automation; less importance is attached to gaps attributable to manual data collection. While, with respect to transaction monitoring, participants confirmed the importance of such features as improved analytics and advanced checking on monitoring banking transactions, some modicum of skepticism was raised with respect to over-automating. With respect to minimizing cost and time, participants, with greater consistency relative to other independent variables, placed weight on employment of automation and deployment of new technologies as drivers of efficiencies. Of all statements, the greatest assent and consistency was registered to the cost-and-time statement that "Information collected in real time by KYC automation reduces both cost and time." For the most part, bankers consider that their respective FIs have achieved a substantive degree of success in money laundering prevention effectiveness with the caveat that not all banks have achieved containment of enforcement costs and, indeed, ¼ of bankers admit that their FIs have been on the receiving end of punitive regulatory measure for non-compliance.

References

1. Basel Committee on Banking Supervision.: Implications of fintech developments for banks and banks supervisors. (2018). https://www.bis.org/bcbs/publ/d431.pdf
2. CAMS: Study Guide for the CAMS Certification Examination (5th Ed.). Association of Certified Anti-Money Laundering Specialists (2012)
3. Fenergo: Global financial institutions fined $26 billion for AML, sanctions and KYC Non-compliance. (2018). https://www.fenergo.com/press-releases/global-financial-institutions-fined-$26-billion-for-aml-kyc.html
4. Geister, D.: Concepts in customer Due diligence: meeting the challenge of regulatory expectations. (2008). https://risk.lexisnexis.com/
5. LexisNexis: Future financial crime risks. (2017)
6. Miller, R.S., Rosen, L.W.: Anti-money laundering: an overview for Congress Congressional Research Service. (2017). https://fas.org/sgp/crs/misc/R44776.pdf
7. Protiviti: Views on AML transaction monitoring systems. (2013)
8. Reuters, T.: US anti-money laundering insights report. (2017). Retrieved from http://iffoadatabase.trustafrica.org/iff/2017-aml-insights-survey-report.pdf

A Cognitive Knowledge Base for Learning Disabilities Using Concept Analysis

M. Srivani, T. Mala, and S. Abirami

Abstract In recent days, the amount of unstructured data has been widely increased, and it is a challenging task to derive meaningful insights and knowledge from the data. A Cognitive Knowledge Base (CKB) is constructed to derive the hidden semantic knowledge from the unstructured text data. A Knowledge Base (KB) is a collective structure of knowledge developed by ontology or rule-based techniques. A CKB is an Object-Attribute-Relational (OAR) model constituted by a weighted hierarchical concept network. CKB relies on the denotational mathematical structure of formal concepts like concepts, objects, attributes, and knowledge. This paper deals with the construction of a CKB using Formal Concept Analysis (FCA). Construction of CKB using FCA consists of four steps, namely (i) topic modeling guided concept extraction, (ii) determination of objects and attributes, (iii) determination of relevant objects, relevant attributes, hyper-attributes, and (iv) construction of relational hierarchy. The proposed framework derives meaningful insights by extracting main concepts and representing the weighted hierarchical concept network as a relational hierarchy. This paper demonstrates a CKB developed for dyslexic children.

Keywords Cognitive knowledge base · Formal concept analysis · Concept extraction · Relational hierarchy · Semantic knowledge

1 Introduction

Nowadays, the growth of data is increasing, and it is becoming complicated to derive meaningful and useful insights from the data. The levels of transformation are data to information, information to knowledge, and finally, knowledge to wisdom [1]. A KB is a processed database which consists of a repository of information. Ontologies and rule-based techniques developed traditional knowledge bases. The issues of traditional KB are modeling small-scale knowledge, inconsistent, incom-

M. Srivani (✉) · T. Mala · S. Abirami
Anna University, Chennai, India
e-mail: sri.srivani94@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_33

plete, and redundant. So, this paper proposes a CKB based on Machine Learning (ML) techniques. A CKB is an intelligent database which consists of a hierarchical concept network [2] depicting concepts, objects, attributes, relevant objects, relevant attributes, and hyper-attributes. A CKB also comprises semantic relational structures like synonyms, antonyms, partial synonyms, homonyms, hypernyms, and hyponyms. A formal concept is a probabilistic structure with three tuples, namely object, attribute, and relation (hyper-attribute). An object is an abstract concept, and an attribute is a sub-object. A relation is a hyper-attribute which indicates the connection between an object to object, object to attribute, and attribute to attribute. FCA is a mathematical process of organizing concepts and extracting hierarchical concept network [3]. The significant contributions of this paper are

- Extraction of meaningful insights from unstructured data.
- Semantic knowledge extraction from relational hierarchy.
- Construction of CKB using concept analysis.

The rest of the paper is standardized as follows: Sect. 2 describes the work concerning the construction of CKB. Section 3 explains the system architecture, and Sect. 4 describes the methodology of CKB construction. Section 5 presents the discussion and results, and Sect. 6 concludes the work by suggesting possible future works.

2 Background and Related Work

The main challenges in constructing a CKB are large unstructured data, human understandability, precise knowledge elicitation, and representation.

2.1 Construction of CKB

A CKB is built using concept algebra-based algorithm of CKB generation (ACKBG) [2], an unsupervised and autonomous ML algorithm for generating a hierarchy of formal concepts. The main issue of this work is large-scale CKB generation. A CKB m-learning framework for modeling and enhancing the knowledge level is built which encourages collaborative learning [4]. A CKB for adaptive learning environment [5] with adaptive feedback integrates three elements of knowledge such as pedagogical, domain, and learner using concept algebra. The CKB architecture comprises a logical, physical model, linguistic KB, and knowledge manipulation engine. A novel CKB is designed for cognitive robots [6] using ML techniques and concept algebra. The CKB processes are knowledge acquisition, manipulation, fusion, and retrieval. A system of formal concepts is autonomously generated and Cognitive Knowledge Learning (CKL) [7] is implemented using Algorithm of Cognitive Concept Elicitation (ACCE). Concept algebra is a denotational mathematical structure for manipulating the formal concepts in knowledge representation [8]. Formal concepts are extracted, and semantic hierarchy is built autonomously [9] using Concept Semantic Hierarchy (CSH) algorithm. Formal model of weighted Hierarchical Concept Network (HCN) is built [10] and semantic equivalence is derived using Concept Equivalence Analysis (CEA) algorithm. Semantic hierarchy is derived using Relational Semantic Classification (RSC) algorithm.

2.2 Knowledge Representational Model

A fundamental cognitive unit of knowledge is a formal concept. A set of formal cognitive models [11] is elicited as cognitive knowledge systems. A semantic knowledge computational and representational model is designed [12] using real-time process algebra (RTPA). The model involves knowledge acquisition and encoding, knowledge representation, knowledge inference, and validation. The main issue lies in the missing values, noisy data, uncertainty of the system, and creation of representation rules for multi-categorization. A Formal Knowledge Representation System (FKRS) is implemented [13] for autonomous concept extraction using concept algebra and RTPA. Concepts serve as input, and knowledge evolves as output by the interaction of the two sub-systems. The main issue is that the RTPA is abstract and makes the system more complex. The various cognitive architectures [14] such as SOAR, ACT-R, CLARION, and Vector-LIDA portray the knowledge level, design, and the process of representation. The main issue is knowledge size and homogeneity. Many of the cognitive systems lack conceptual KB.

3 System Architecture

The process of CKB construction, as shown in Fig. 1, starts with preprocessing of unstructured data such as textbooks and journal articles. Preprocessing steps include stop words removal, removing punctuation, removing numbers, converting to lowercase, tokenization, and normalization. Secondly, extract the formal concepts using topic modeling-based concept extraction.

Thirdly, the objects and attributes are determined. Objects are the elements of each concept in all documents. The attributes are the most frequent words of each concept in all documents. Weights assigned to the objects and attributes have a range of 0–1. Then, determine the relevant objects, attributes, and hyper-attributes for each concept. Calculate the average information weight, total average information weight and the granular weight for each object and attribute. After applying the threshold values and eliminating, the retained items are the relevant objects and relevant attributes. Next, determine the hyper-attributes (relation) by performing the Cartesian product of formal concepts. At the final stage, assign a threshold, and extract the relational hierarchy for each concept. Finally, visualize the semantic knowledge from the relational hierarchy.

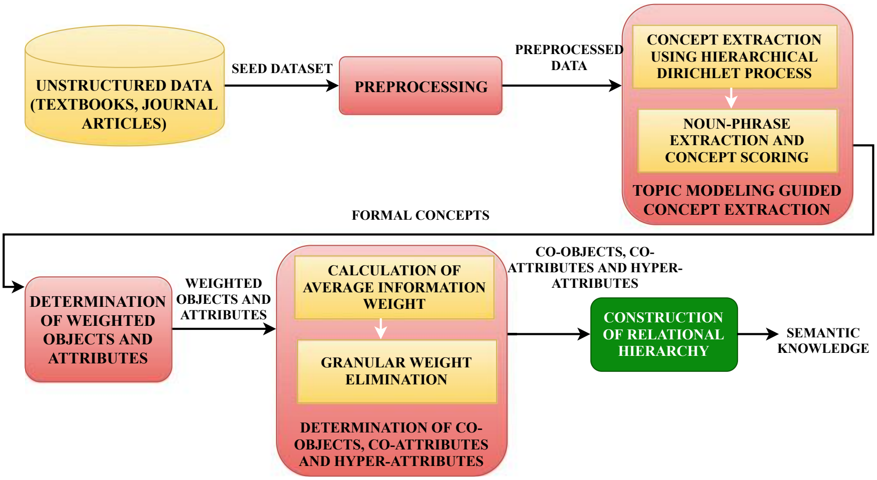


Fig. 1 Process of cognitive knowledge base construction

4 Methodology

4.1 Topic Modeling-Based Concept Extraction

The primary process of this step is to extract semantic topics with concepts from the corpus of unstructured data. This step consists of two sub-steps, namely.

Concept Extraction using Hierarchical Dirichlet Process Topic modeling is a process to uncover the semantic structures of text. For topic modeling, Hierarchical Dirichlet Process (HDP) algorithm is used [15]. HDP is an extension of Latent Dirichlet Allocation (LDA) which is used to infer the number of topics from the corpus. HDP abstracts and contrives huge repositories of text data. In HDP, the number of topics with topical words is generated based on the data.

Extraction of Noun Phrases and Concept Scoring Function Noun phrases are a collection of words that functions as a single noun. Extract the noun phrases for each topic using the Stanford POS tagger. The Stanford POS tagger is a probabilistic tagger which tags up a word with a specific POS. After the noun phrase extraction, perform noun phrase chunking, calculate the scores for each topic, and extract top formal concepts. Equation 1 indicates the formula for concept scoring function. For each topic of words,

$$CSF = \sum_{documents} Topic\ Probability \times TF\ of\ Noun\ Phrase \tag{1}$$

where CSF is concept scoring function, and TF is term frequency. The topic of words with the top score indicates formal concepts. The basic element of knowledge in CKB is a formal concept represented by OAR model. So, perform the determination of weighted objects, attributes, and hyper-attributes.

4.2 Determination of Objects and Attributes

Objects are the elements of the concepts. In the formal context, the collection of attributes depicts the objects. Search the concept from all documents of the corpus, and extract the elements as objects. Attributes are the specific instances of the concepts. Similarly, extract the most frequent words as attributes. Assign weights to the objects and attributes in the range of 0 to 1 based on their term frequencies. In this step, the binary relations between objects and attributes for each formal concept are realized.

4.3 Determination of Relevant Objects, Relevant Attributes, and Hyper-Attributes

Learning Relevant Objects and Relevant Attributes The learning step deals with calculating the average information weight, granular weight, and eliminating the less relevant objects and attributes. Algorithm 1 describes the process of learning relevant objects and relevant attributes. Repeat the same procedure for attributes, and determine the relevant attributes.

Algorithm 1 Learning Relevant Objects and Attributes

- Input: Objects and Attributes
- Output: Relevant Objects and Relevant Attributes

for For each object (x_i) and attribute (y_i) **do**
  Calculate I_i = P(x_i / y_i) × $\log_2$ P(x_i / y_i), where x_i - $Object_i$, y_i - $Attribute_i$, I_i - Average Information Weight
  Calculate P(x_i / y_i) = P($x_i \cap y_i$) / P(y_i) where P(x_i / y_i) is the conditional probability of object x_i given attribute y_i
  $\cap$ - intersection operation
  Calculate $I = \sum_{i=1}^{m} I_i$, where m - Total Number of Objects, I - Total Average Information Weight
  Calculate W_i = I_i / I, where W_i - Granular Weight
  Select the threshold θ_o, where θ_o - Elimination Threshold
  if If $I_i < W_i$ **then**
    eliminate the corresponding object and repeat the process for all objects
  end if
  After elimination, retain the relevant objects
end for

Learning Hyper-attributes Hyper-attributes express the Cartesian product relation (set of all possible ordered pairs) between relevant objects and relevant attributes, as indicated in Eq. 2.

$$Hyper - attributes = Relevantobjects \times Relevantattributes \tag{2}$$

Then, use Algorithm 1 to determine the relevant hyper-attributes.

4.4 Construction of Relational Hierarchy

A relational hierarchy of concepts interprets the semantic knowledge for a concept, as shown in Fig. 2. Synonyms indicate relevant and equivalent concepts. Partial synonyms reveal the partially equivalent concepts. Hyponyms reflect specific instances of the concept that is the first hyper-attribute, relevant object, and relevant attribute. Homonyms demonstrate the derived or sub-concepts. Antonyms depict non-equivalent and irrelevant concepts. Hypernyms are super concepts. Algorithm 2 shown below constructs relational hierarchy and extracts semantic knowledge.

Algorithm 2 Semantic Knowledge

– Input: Hyper-attributes and Relevant Attributes
– Output: Relational Hierarchy

for For each concept (C_i) **do**
    Calculate C_i = P($A_i \cap A_j$) / P($A_i \cup A_j$), where A_i - Hyper-attribute and A_j - Relevant Attribute
    Hyponym = A_{i1} and A_{j1} (Specific Instance)
    if If $C_i > 0.4$ **then**
        then Synonym (Equivalent Concept)
    end if
    if If $C_i \geq 0$ && $C_i < 0.1$ **then**
        then Antonym (Opposite Concept)
    end if
    if If $C_i \geq 0.3$ && $C_i < 0.4$ **then**
        then Homonym (Sub-Concept)
    end if
    if If $C_i \geq 0.2$ && $C_i < 0.3$ **then**
        then Partial Synonym (Partially Equivalent Concept)
    end if
    if If $C_i \geq 0.1$ && $C_i < 0.2$ **then**
        then Hypernym (Super-concept)
    end if
end for

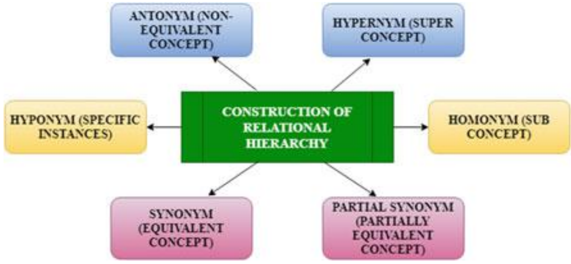


Fig. 2 Construction of relational hierarchy

5 Discussion and Results

Dyslexia is a language processing and specific learning disability (SLD) characterized by difficulty in reading, memorizing, spelling, thinking, and understanding. In this paper, a CKB for learning difficulties is constructed using FCA, and the hidden semantic knowledge is interpreted from the unstructured data. The proposed framework discusses the generation of semantic knowledge by the conceptual analysis of text data. The dataset comprises ten journal articles (teaching strategies) and 20 dyslexia books. The implementation is carried out in Eclipse Integrated Development Environment (IDE) Spring Source Tool Suite–Java platform, Angular 8, a popular JavaScript framework and Visual Studio Code. Nearly, 661 topics with concepts (topical words) are extracted using HDP algorithm by Gibbs sampling. Figure 3 portrays the top three topics with concepts.

Noun phrases are extracted using wsj 0–18 bidirectional distsim tagger with Minfeaturethreshold = 2 and Sigmasquared = 0.5. Noun phrases are chunked using en-parser-chunking. Figure 4 displays top three chunked noun phrases.

Nearly, 554 top formal concepts are generated using concept scoring function. Table 1 displays top ten formal concepts.

The graph shown in Fig. 5 displays the weighted objects and attributes determined for the concept "Services." X-axis indicates the indices starting from 0 to 10 for objects and attributes. Y-axis indicates the probabilities for objects and attributes. From the graph, it is evident that the heavy weighted object and attribute is "Information (0.898)" and "Dyslexia (0.8994)." Similarly, the least weighted object and attribute

Topic:0 disability(0.8055) software(0.0845) technology(0.4628) information(0.5191) assistive(0.1898) adjustments(0.5205) services(0.0249) recording(0.7598) recognition(0.9286) education(0.1797)

Topic:1 skills(0.1024) phonemes(0.9234) adults(0.1412) homework(0.4382) school(0.2257) teacher (0.4104) children(0.3424) teachers(0.9070) parents(0.8751) classroom(0.8394)

Topic:2 learning(0.5932) styles(0.9615) learners(0.3539) relation(0.2648) responsibility(0.1889) environment(0.2354) child (0.3394) physical(0.6865) structure(0.8137) environmental(0.7745)

Fig. 3 Top three topics with concept probabilities

POS Output : disability/NN software/NN technology/NN information/NN assistive/JJ adjustments/NNS services/NNS recording/VBG recognition/NN education/NN
List of Noun Parse : [adjustments, education, software, disability, recognition, information, technology, services]
Concepts: adjustments(0.6580) education(0.9065) software(0.6850) disability(0.5583) recognition(0.3246) information(0.8390) technology(0.7455) services (0.8384)

POS Output :
offered/VBD identify/VB strategies/NNS appropriate/JJ systems/NNS activities/NNS literacy/NN failure/NN opportunities/NNS
List of Noun Parse : [literacy, strategies, systems, activities, failure, opportunities]
Concepts: literacy(0.2244) strategies(0.6180) systems(0.2047) activities(0.0663) failure(0.4369) opportunities(0.6451)

POS Output :
skills/NNS phonemes/NNS adults/NNS homework/NN school/NN teacher/NN children/NNS teachers/NNS parents/NNS classroom/NN
List of Noun Parse : [skills, teacher, homework, school, children, teachers, adults, classroom, phonemes, parents]
Concepts: skills(0.3347) teacher(0.6605) homework(0.3861) school(0.5322) children(0.0531) teachers(0.4149) adults(0.0624) classroom(0.2521) phonemes (0.7891) parents(0.2827)

Fig. 4 Top three chunked noun phrases along with concept probabilities

Table 1 Top ten formal concepts

S.No.	Formal concepts
1	Dyslexia (0.9765)
2	Learning (0.9691)
3	Difficulties (0.9409)
4	Education (0.9065)
5	Spelling (0.8711)
6	Comprehension (0.8427)
7	Reading (0.8390)
8	Services (0.8384)
9	Phonemes (0.7891)
10	Student (0.7585)

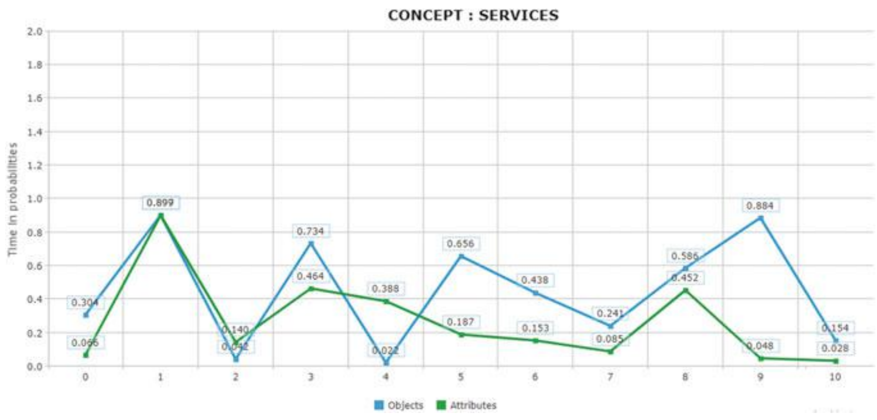


Fig. 5 Line graph for weighted objects and attributes of concept “Services”

Table 2 Weighted objects and attributes of concept "Services"

Index	Weighted objects	Weighted attributes
0	Address (0.3035)	Introduction (0.0663)
1	Information (0.8974)	Dyslexia (0.8994)
2	Phonics (0.0415)	Symptoms (0.1403)
3	Emergency (0.7335)	Social (0.4640)
4	Overcome (0.0216)	Emotional (0.3880)
5	Special (0.6556)	Connection (0.1871)
6	Eligibility (0.4380)	Classroom (0.1533)
7	Education (0.2406)	Strategies (0.0848)
8	Understand (0.5864)	Multisensory (0.4515)
9	Critical (0.8842)	Structured (0.0481)
10	Southern (0.1537)	Language (0.0284)

Table 3 Relevant objects and relevant attributes of concept "Services"

Index	Relevant objects	Relevant attributes
0	Address (0.3035)	Dyslexia (0.8994)
1	Information (0.8974)	Emotional (0.3880)
2	Emergency (0.7335)	Multisensory (0.4515)
3	Special (0.6556)	Screening (0.8752)
4	Eligibility (0.4380)	
5	Education (0.2406)	
6	Critical (0.8842)	
7	Psychological (0.5858)	
8	Children (0.9918)	

is "Overcome (0.0216)" and "Important (0.0284)." Table 2 displays the indices along with the weighted objects and attributes concerning the line graph.

Table 3 displays the weighted relevant objects and relevant attributes determined for the concept "Services." Threshold is 0.2, so the objects and attributes less than 0.2 are eliminated. New relevant objects "Psychological (0.5858)" and "Children (0.9918)" are added. Table 4 indicates the various relational hierarchies concerning the semantic line graph. The most equivalent concept to "Services" is "Dyslexia (0.46)." Irrelevant concept is "Emotional (0.04)" and "Multisensory (0.04)". Sub-concept is "Dyslexia (0.38)" and "Screening (0.37)." Hyper-attributes are "Dyslexia (0.2694)," "Emotional (0.3197)," and "Multisensory (0.0484)."

Table 4 Relational hierarchy of concept "Services"

Synonyms	Antonyms	Homonyms	Partial synonyms	Hyponyms	Hypernyms
Dyslexia (0.46)	Emotional (0.04)	Dyslexia (0.38)	Dyslexia (0.21)	Dyslexia (0.21)	Emotional (0.16)
Screening (0.45)	Multisensory (0.04)	Screening (0.37)	Emotional (0.16)	Dyslexia (0.24)	Multisensory (0.17)

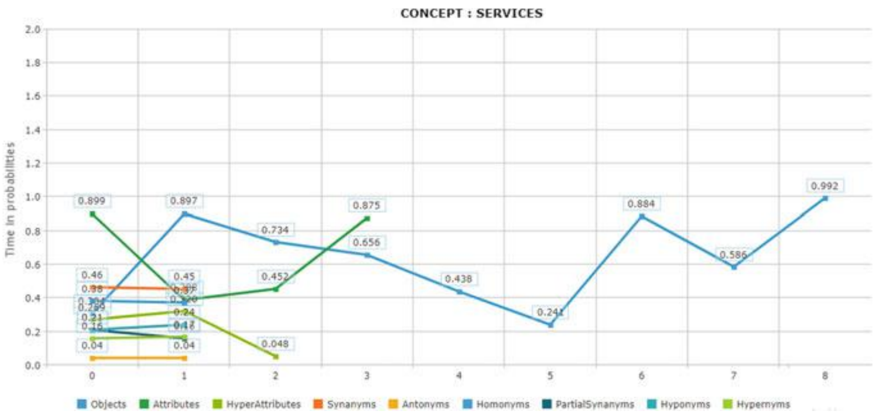


Fig. 6 Semantic line graph for relational hierarchy of concept "Services"

Table 5 Performance analysis

Dataset	Precision	Recall	Accuracy
Journal articles	0.89	0.95	0.87
Dyslexia books	0.94	0.94	0.89

Figure 6 displays semantic relational hierarchical graph for concept "Services." The total number of refined concepts from the journal articles is 86. The total number of refined concepts from dyslexia books is 575. Potential metrics such as accuracy, precision, and recall are calculated to depict the performance of the system, as shown in Table 5.

6 Conclusion

The proposed framework deals with the process of construction of CKB using FCA. The CKB reveals the semantic knowledge from the unstructured text data using the concept analysis techniques. CKB deals with formal semantics and topical structures. CKB is an OAR and denotational mathematical model for discovering knowledge from unstructured text data. CKB helps in the process of generating new knowledge from the existing knowledge. Some limitations of this work include the lack of semantic analysis techniques for knowledge representation. In the future, CKB can be improved by applying the process of semantic analysis and algebra. The CKB can further be enhanced by composing the operations of concept and semantic operators. Semantic clustering of formal concepts technique can be used to enhance the performance of the system.

Acknowledgements The authors gratefully acknowledge DST, New Delhi, for providing financial support to carry out this research work under DST-INSPIRE Fellowship scheme. One of the authors, Ms. M. Srivani, is thankful for DST for the award of DST-INSPIRE fellowship.

References

1. Wang, Y.: Formal cognitive models of data, information, knowledge, and intelligence. WSEAS Trans. Comput. **14**(3), 770–781 (2015)
2. Wang, Y., Zatarain, O.A., Valipour, M.: Building cognitive knowledge bases sharable by humans and cognitive robots. In: IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 3189–3194. Canada (2017). https://doi.org/10.1109/SMC.2017.8123119
3. Zhao, M., Zhang, S., Li, W., Chen, G.: Matching biomedical ontologies based on formal concept analysis. J. Biomed. Semant. **9**(11), 1–27 (2018). https://doi.org/10.1186/s13326-018-0178-9
4. Al-Hunaiyyan, A., Bimba, A. T., Idris, N., Al-Sharhan, S.: A cognitive knowledge-based framework for social and metacognitive support in mobile learning. Interdiscip. J. Inf., Knowl. Manag. 12, 75–98 (2017). https://doi.org/10.28945/3670
5. Bimba, A.T., Idris, N., Mahmud, R.B., Al-Hunaiyyan, A.: A cognitive knowledge-based framework for adaptive feedback. In: International Conference on Computational Intelligence in Information System, pp. 245–255. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48517-1_22
6. Wang, Y.: On a novel cognitive knowledge base (CKB) for cognitive robots and machine learning. Int. J. Softw. Sci. Comput. Intell. (IJSSCI) **6**(2), 41–62 (2014). https://doi.org/10.4018/ijssci.2014040103
7. Wang, Y., Zatarain, O.A.: A novel machine learning algorithm for cognitive concept elicitation by cognitive robots. Int. J. Cogn. Inform. Nat. Intell. (IJCINI) **11**(3), 31–46 (2017)
8. Wang, Y.: Concept algebra: a denotational mathematics for formal knowledge representation and cognitive robot learning. J. Adv. Math. Appl. **4**(1), 61–86 (2015). https://doi.org/10.1166/jama.2015.1074
9. Valipour, M., Wang, Y.: Building semantic hierarchies of formal concepts by deep cognitive machine learning. In: IEEE 16th International Conference on Cognitive Informatics Cognitive Computing (ICCI* CC), pp. 51–58 (2017). https://doi.org/10.1109/ICCI-CC.2017.8109729
10. Wang, Y., Valipour, M., Zatarain, O.A.: Quantitative semantic analysis and comprehension by cognitive machine learning. Int. J. Cogn. Inform. Nat. Intell. (IJCINI) **10**(3), 13–28 (2016). https://doi.org/10.4018/IJCINI.2016070102
11. Wang, Y.: On cognitive foundations and mathematical theories of knowledge science. In: Artificial Intelligence: Concepts, Methodologies, Tools, and Applications, pp. 889–914. IGI Global (2017). https://doi.org/10.4018/IJCINI.2016040101
12. ElBedwehy, M.N., Ghoneim, M.E., Hassanien, A.E., Azar, A.T.: A computational knowledge representation model for cognitive computers. Neural Comput. Appl. **25**(7–8), 1517–1534 (2014). https://doi.org/10.1007/s00521-014-1614-0

13. Tian, Y., Wang, Y., Gavrilova, M.L., Ruhe, G.: A formal knowledge representation system (FKRS) for the intelligent knowledge base of a cognitive learning engine. Int. J. Softw. Sci. Comput. Intell. (IJSSCI) **3**(4), 1–17 (2011). https://doi.org/10.4018/jssci.2011100101
14. Lieto, A., Lebiere, C., Oltramari, A.: The knowledge level in cognitive architectures: current limitations and possible developments. Cogn. Syst. Res. **48**, 39–55 (2018). https://doi.org/10.1016/j.cogsys.2017.05.001
15. Andrews, M.: A Gibbs Sampler for a Hierarchical Dirichlet Process Mixture Model. PsyArXiv, pp. 1–9 (2019). https://doi.org/10.31234/osf.io/ebzt8

Native Monkey Detection Using Deep Convolution Neural Network

Pankaj Kumar and Mitalee Shingala

Abstract We present development and analysis of deep learning architectures for real life and real-time native Indian monkey (langur) detection. Our monkey detection work is motivated with the idea of developing an application which will assist farmers in protecting their crops from attack of native wild animals. Protecting destruction of crops from wild animals is very important for agrarian societies. We engineered pre-trained VGG19 convolution neural network (CNN) to detect live monkeys in images streamed in an image sequence generated from a USB webcam. We used transfer learning approach to custom train, YOLO, a much faster CNN architecture to detect monkeys. For training and analysis, we used the Google Open Images V4 dataset. We give details of how two significantly different architectures were customized for native monkey detection. The performance of the detectors developed is analysed in terms of time taken for detection and accuracy of detection. Custom trained architecture achieved high true positive rate of 97.19% and true negative rate of 95.32%, which are significantly higher and better than the pre-trained architecture. The developed system was successfully tried out on a 4GB RAM and Intel quadcore i3 laptop at Indroda National park.

Keywords Object detector · Convolution neural network · Deep learning · Real time

P. Kumar (✉) · M. Shingala
DAIICT, Gandhinagar 382007, Gujarat, India
e-mail: pankaj_k@daiict.ac.in
URL: https://www.daiict.ac.in/profile/pankaj-kumar/

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_34

1 Introduction

There are many incidents where farmers' crops were destroyed by animals like cow, deer, monkey, horse, pig, elephant and nilgai [1]. Usually, these animals enter into farms for food and destroy the crops by their frivolous activities. From surveys conducted on farmers, related to the risks they face in agriculture, unanimously the farmers agreed that wild animals are threat to their crops. For about 57% of the farmers, animals are the major factor in destruction of their crops [1].

Detecting animals of different species in camera images has many other useful applications too. Use of camera traps in natural habitats have transformed wild life ecology and conservation in past two decades [2]. Camera traps have become an important tool for many ecologists for wild animal population sizes and distribution studies [3]. There has been several studies of detecting and monitoring animals in the wild using camera traps. Extracting information from camera trap images has been traditionally done by human experts in the field or community of volunteers. Jeff Clue et al. [4] trained a deep CNN to identify count and describe the behaviour of 48 species in the 3.2 million images of Snapshot Serengeti dataset. They compared the performance of deep neural network architectures for detecting animals, and they reported best performance for VGG net [5] architecture achieving >96.8% accuracy. Their main motivation was to save the costly human endeavour required to detect, classify and count wild animals in a large image dataset. They also demonstrated that the modern machine learning methods using CNNs works almost as good as humans. Human-volunteer species detection and count labels are estimated to be 96.6% and 90.0% accurate, respectively, versus label provided by experts [6]. And these scores were easily matched by their machine learning methods. Inspired by this success in wild animal monitoring, we developed an application which can detect animals in real-life image sequences generated from consumer-grade cameras. Such a detection will trigger notification to the farmers to help them take decisions of preventive action and/or trigger some non-destructive preventive action which will scare animals away using acoustic and/or light repellents.

During this work, we realized that to train neural networks for object detection purpose, available datasets are very limited. We do not have dataset completely dedicated for training of detectors for native animals like langur, nilgai, wild boars etc. Using our techniques, automatic data collection is possible, and later, this data can be processed to use it in training neural networks for varied purposes. For present work, we are concentrated on monkey detection but we can do it for other animals also. We retrained YOLO V3 [7] on monkey images in Google Open Images V4 dataset [8] for online detection of monkeys.

Henceforth, the paper is organized as follows: in Sect. 2, we discuss the reasons for the choice of CNN architecture for animal detection. Then in Sect. 3, we give the details of our modification and engineering of these architectures for training live monkey detector. In experiments and results Sect. 4, we give the details of the data and its restructuring for our work. Finally, we conclude in Sect. 5 with summary of the work done and some future directions for the work.

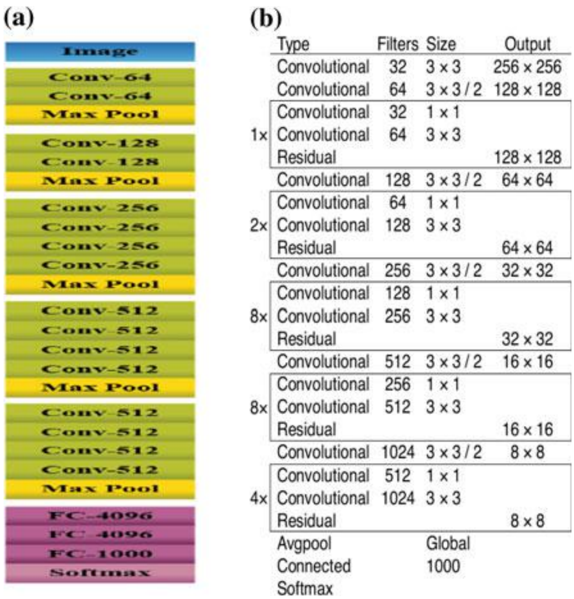
2 Choice of CNN Architecture

We experiment with two CNN architectures. One is designed for object classification and other for object detection. First we briefly elaborate the details of the CNN architecture for object classification.

2.1 VGG19 CNN

There is a pre-trained VGG19 CNN [5] model trained on ImageNet dataset, and it has various species of monkey class in it. To start with, we developed and experimented with a detector without explicitly re-training VGG19 CNN. We used pre-trained VGG19 CNN with 1000 classes of which 21 classes were of monkey species of our interest.

VGG19 uses stacked 3×3 kernel-sized filters instead of using large kernel-sized filter to extract complex features. It contains 16 convolutional layers followed by three fully connected layers and then followed by a soft-max classifier, as shown in Fig. 1a. It takes 224×224 RGB image as an input, and it gives a class score map with the number of channels equal to the number of different classes at output.



	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1x	Convolutional	32	1 × 1	
1x	Convolutional	64	3 × 3	
1x	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2x	Convolutional	64	1 × 1	
2x	Convolutional	128	3 × 3	
2x	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8x	Convolutional	128	1 × 1	
8x	Convolutional	256	3 × 3	
8x	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8x	Convolutional	256	1 × 1	
8x	Convolutional	512	3 × 3	
8x	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4x	Convolutional	512	1 × 1	
4x	Convolutional	1024	3 × 3	
4x	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig. 1 **a** Different layers of the VGG19 CNN as described in the paper [5]. We use this pre-trained network for our monkey detector development and analysis. **b** Different layers of the Darknet-53 architecture as described in [7]. We custom train it for our monkey detection

2.2 YOLOv3 CNN

You Only Look Once (YOLO) is a real time and faster technique compared to other techniques, so in the present work, we used latest version of YOLO algorithm, i.e. YOLOv3 [7, 9, 10].

Figure 1b shows the architecture of YOLOv3 [7]. It has 53 convolutional layers, so it is named as Darknet-53. Darknet is an open-source neural network framework written in C and CUDA [11]. To improve performance, YOLOv3 uses shortcut connections inspired from residual network [12]. YOLOv3 uses a linear activation function for the final layer, and all other layers use the following leaky rectified linear activation:

$$\phi(x) = \begin{cases} x, & \text{if } x > 0 \\ 0.1x, & \text{otherwise} \end{cases} \tag{1}$$

For each cell in the feature map, the detection layer predicts three bounding boxes at each scale, objectness score for those boxes and C class probabilities. These predictions are encoded as an $N \times N \times (3 * (4 + 1 + C))$ matrix. Here, four parameters are x-coordinate and y-coordinate of the centre of the bounding box; width and height of the bounding box. Objectness score defines the presence of an object; if it is 1, then object is present, otherwise object is not present. YOLOv3 predicts an objectness score for each bounding box using logistic regression.

3 Implementation of Different Techniques

Using the above CNNs, we developed following three different techniques for real-time monkey detection in images captured from a webcam.

3.1 Sequential Search with VGG19

To detect the presence of monkey in certain sensitive area, camera will be placed at a distance from where we will be able to capture the image with monkey if they enter the premises. In the captured images, there can be many objects like tree, car or house along with monkey. In such images, monkey will not occupy a large image area except if a monkey is very close to the camera. So we cannot classify such images directly, instead we can use a sliding window and pass the cropped images for classification.

In the Sequential Search with VGG19 technique, we crop the whole image into several 224×224 windows, and each cropped image is given sequentially as input to VGG19 for classification. VGG19 will give class score map at output, from that

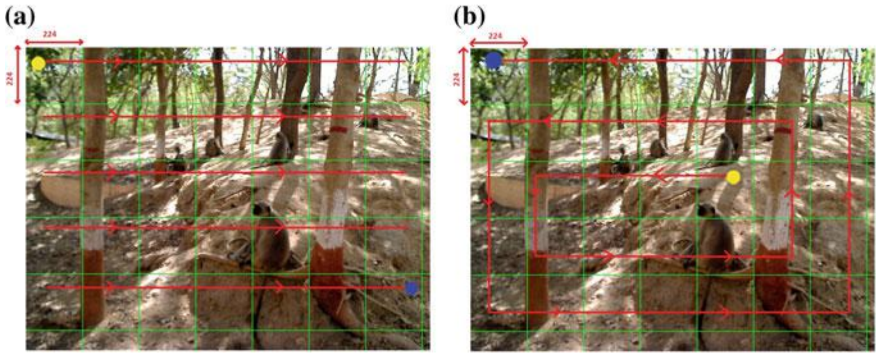


Fig. 2 **a** Illustration of how Sequential Search of bounding box works on an image frame. **b** Illustration of how Spiral Search of bounding box works on an image frame

we will retrieve the class with highest probability. If that class belongs to one of the species of monkey with probability greater than 0.1, then detection of monkey is confirmed.

To understand Sequential Search, let us take an example of image shown in Fig. 2a. It is cropped into $N = 12$, 224×224 windows depending upon the frame size. Yellow circle shows our start crop, from there we will start giving cropped images to VGG19 in sequence shown by red arrow lines, till we reach end crop, shown by blue circle. Before reaching to end crop, if a crop is classified as monkey, we will stop the process and we will say that monkey is detected in the image.

3.2 Spiral Search with VGG19

In the Spiral Search with VGG19 technique, everything is same as Sequential Search with VGG19 technique, except the fact that we give cropped image one after another in a spiral fashion as depicted in Fig. 2b. We designed this technique because in case monkey is present at frame centre, then the processing time for monkey detection will be faster.

3.3 Custom YOLOv3

In Custom YOLOv3 technique, we have used YOLOv3 CNN architecture to train on monkey dataset. Each image in dataset will have a data annotation file which gives information about location of monkeys in image. Format of data annotation file is in Sect. 3.3.

Instead of training the network from scratch, we have used transfer learning to initialize weights from pre-trained model which contains convolutional weights trained on ImageNet. Also, we have set few configuration parameters to train YOLOv3; its details are given in Sect. 3.3.

YOLOv3 gives bounding box description, objectness value and probability associated with each class for each detection. If object is present in the detection, we will find the maximum class probability, and if it is greater than 0.1, then we say object of that class is detected. In our case, we have only one class of monkey, so if class probability is greater than 0.1, then we say monkey is detected.

Data Annotation To train YOLOv3, each image should have its annotation file which gives information about location of monkeys in image. This information is the bounding box description for each monkey present in image. Each annotation file has n number of lines to describe n bounding box, where n is the number of monkeys in an image. The description of each bounding box is in the following format:

$$\begin{aligned} < object - class - id >< center - x >< center - y > \\ < width >< height > \end{aligned} \tag{2}$$

In Eq. 2, the first field object-class-id is an integer representing the class of the object. Class-id starts from 0, so our monkey class has class-id 0. In our case, since we have only one class of monkey, object-class-id is always set to 0. The second and third entries, centre-x and centre-y, are, respectively, the x- and y-coordinates of the centre of the bounding box, normalized (divided) by the image width and height, respectively. The fourth and fifth entries, width and height, are, respectively, the width and height of the bounding box, normalized (divided) by the image width and height, respectively. This last four entries are all floating point values between 0 and 1.

YOLOv3 Configuration Parameters We have set some training parameters to train YOLOv3; they are listed below:

- **Batch Size**: A small subset of images is used in one iteration to update the weights, and this subset is called the batch size. We have set its value to 64, which means 64 images are used in one iteration to update the parameters of the neural network.
- **Subdivisions**: To process a fraction of the batch size at one time on GPU, we use subdivisions. We have set its value to 16, which means GPU will process $batchsize/subdivision$ number of images at any time, but the full iteration would be complete only after all the 64 (as set above) images are processed.
- **Width, Height and Channels**: These configuration parameters specify that input image size should be first resized to $Width \times Height$ before training. We have set width and height to 448 and channels to 3. channels = 3 indicates that we would be processing 3-channel RGB input images.
- **Momentum and Decay**: Momentum is used to penalize large weight changes between iterations. A neural network has millions of weights, and therefore, they can easily overfit any training data. To avoid overfitting, we need to penalize large

value for weights. The parameter decay controls this penalty term. Here we have set momentum to 0.9 and decay to 0.0005.
- **Learning Rate**: The parameter learning rate controls how aggressively we should learn based on the current batch of data. Here we have set learning_rate to 0.001.
- **Data Augmentation**: To increase training data, we can use data augmentation by transforming the colours of the entire image using saturation, exposure and hue. Colour spaces play an important role in computer vision [13]. Here RGB to HSV transform is used for data augmentation. We have used the default values of saturation = 1.5, exposure = 1.5 and hue = 0.1 for training.
- **Number of Iterations**: Number of iterations specify how many iterations should the training process be run for. We have trained YOLOv3 for 10,000 iterations on Nvidia Titan Xp GPU (12GB RAM and 3840 cores) for 2 days.

4 Experiments and Results

In Gujarat, we have tested our techniques on field and off field. In on-field experiment, we have saved all the captured images, with or without monkey in it.

4.1 Dataset

For training YOLOv3 CNN, we have used publicly available monkey dataset from Open Images V4 dataset released by Google [14, 15]. The monkey dataset consists of more than 2000 images with monkey location annotations. We divided the monkey dataset into two parts, 90% for training and 10% for testing.

Open Images V4 dataset is used for training, validation and testing for all the three techniques. As all the images in test dataset are images with monkeys in it, we added equal number of images to test dataset, with no monkey in it, to find accuracy. In our experiment, 10% of Open Images V4 monkey dataset is 214 images, with monkeys present in the image. So we added extra 214 images, with no monkey present in the image, to test dataset. These 214 images without monkey are taken from the images we collected on field and few images of natural scenes from Internet.

4.2 Hardware Configuration

For all the experiment, we have used laptop with 4GB RAM, 64-bit Operating System and Intel(R) Core(TM) i3-3110M CPU@2.40 GHz. For on-field experiment, we also used two USB cameras (Microsoft and Logitech with different resolutions) connected to the laptop.

4.3 On-Field Experiments

To test whether our techniques work on field or not, we went to Indroda Park Gandhinagar as monkeys are available in that area. In this setup, we were capturing images real time through our USB cameras and then we were applying our techniques on it. Resolution of images captured by first USB camera was 1280×960 pixels and by second USB camera was 1600×1200 pixels.

We observed that Sequential Search with VGG19 implementation was not able run successfully in real time for our field trials. During the trials, the detection application built on the VGG19 causes the computer to hang on several instances. This can be attributed to the low-end configuration of the laptop. Maybe it is due to low RAM or memory overflow as the model weights for VGG19 is a large file.

Then, we tested Custom YOLOv3, and it was able to detect monkeys in real time. We saved images in which we detected monkey as well as all the images that were being captured. The number of images captured by first camera was 1040 and by second camera was 471, so overall we captured 1511 images.

4.4 Off-Field Experiments

We tested our techniques on our test dataset of monkey and calculated the accuracy, true positive rate (TPR) (sensitivity), true negative rate (TNR) (specificity) and average time taken by each technique. Equations used for calculating accuracy, TPR and TNR are as follows:

$$\text{Accuracy} = \frac{\text{True Positive (TP)} + \text{True Negative (TN)}}{\text{TP} + \text{TN} + \text{False Positive} + \text{False Negative}} \tag{3}$$

$$\text{TPR} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{4}$$

$$\text{TNR} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \tag{5}$$

Accuracy achieved in detecting the presence of a monkey by Sequential Search with VGG19, Spiral Search with VGG19 and Custom YOLOv3 technique is 80.84%, 80.84% and 96.26%, respectively, and average time taken to process an image 8.41s, 8.14s and 2.51s, respectively.

True positive rate achieved by Sequential Search with VGG19, Spiral Search with VGG19 and Custom YOLOv3 techniques is 81.77%, 81.77% and 97.19%, respectively, and true negative rate achieved is 79.90%, 79.90% and 95.32%, respectively.

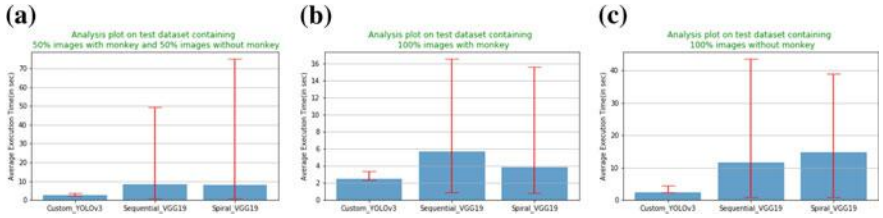


Fig. 3 Plots of the statistics of time taken in processing each image by the three proposed approaches: **a** analysis plot on test dataset containing 50% images with monkey and 50% images without monkey. **b** Test dataset containing 100% images with monkey. **c** Test dataset containing 100% images without monkey. Custom YOLO is fastest of all the three

Highest accuracy is achieved by Custom YOLOv3, and also it takes least average time to detect the presence of monkey. Accuracy of Sequential Search with VGG19 and Spiral Search with VGG19 is equal because they will have same image crops, and only the order in which crops will get processed is different. Because of different order, we can see difference in their average times to detect the presence of monkey.

It can be observed from the Fig. 3 that variation in execution time of Sequential Search with VGG19 and Spiral Search with VGG19 is very high. We thought this high variation is due to presence of 50% of images with monkey and 50% of images without monkey. Images with monkey will take comparatively less execution time than images without monkey because images with monkey may not need to process all the image crops if the monkey is detected in earlier crop. To verify it, we did two other analyses on images with 100% monkey and with 100% no monkey test dataset.

Analysis plot on 100% images with monkey test dataset is shown in Fig. 3. Analysis plot on 100% images with no monkey test dataset is shown in Fig. 3. From Fig. 3, we still observed high variation in execution time of Sequential Search with VGG19 and Spiral Search with VGG19 compared to Custom YOLOv3. We realized that this is due to different input image sizes in Sequential Search with VGG19 and Spiral Search with VGG19. In Custom YOLOv3, input image size is first resized to width (416) and height (416) mentioned in configuration file 3.3, so each image will take almost same execution time. But in other two techniques, execution time depends on the input image size and our test dataset consists of images with different sizes.

4.5 Discussion and Analysis

From on-field and off-field experiment results, it is clear that Custom YOLOv3 performs better in real time than Sequential Search with VGG19 and Spiral Search with VGG19. Both techniques which use VGG19 are slower because it is trained for 1000 classes, whereas YOLOv3 is trained for only 1 class.

We have trained Custom YOLOv3 on images in which around one-fifth area was covered by monkey [15]. We observed that when monkey size is very small in an

Fig. 4 Output of Custom YOLOv3 detection on original and cropped image

image, Custom YOLOv3 could not detect monkey in it, but if we crop the same image such that monkey covers comparatively more area, then the same technique was detecting monkey in it. An example of it is shown in Fig. 4. Such problem is not observed in other two techniques.

To test Sequential Search with VGG19, at early stage we set up a camera looking out of our laboratory window, hooked to a computer for processing. But the experiment was not successful due to the infrequent presence of monkeys. Absence of monkeys near laboratory building is due to the use of radio-frequency and harmful to humans too. We observed from the images which were detected for presence of monkey were mostly false positives. It could also be due to random illumination changes, movements of tree branches and its shadow affects, causing VGG19 to make classification errors.

5 Conclusion and Future Work

In the present work, we designed different techniques to detect the presence of monkey in real time. When we performed experiments on this technique, we found that accuracy of Custom YOLOv3 is 96.26% which is highest among all techniques and average execution time to process an image is 2.51s which is least among all techniques. From this result, it can be concluded that we can use Custom YOLOv3 for real-time monkey detection.

Using the Custom YOLOv3 technique, we have collected and labelled a dataset of native monkeys present in Gujarat area. The dataset will be made publically available and can be used for training the neural networks for even better monkey detection.

We can also train VGG19 for monkey classification using the generated dataset. Average execution time will definitely decrease if number of class is 1 compared to 1000. We can generalize the proposed techniques for other animal detection like nilgai, etc.

References

1. Parikh, A.: Risks in Agriculture Farmers' Perspective. SpringerBriefs in Economics (2017)
2. Allan, K.U.K., O'Connell, F., Nichols, J.D.: Camera Traps in Animal Ecology: Methods and Analyses. Springer Science & Business Media (2010)
3. Silveiraab, L., JÃącomoab, A.T.A., Diniz-Filhoac, J.A.F.: Camera trap, line transect census and track surveys: a comparative evaluation. Biol. Conserv. **114**, 351–355 (2003). https://doi.org/10.1016/S0006-3207(03)00063-6
4. Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C., Clune, J.: Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc. Nat. Acad. Sci. **115**(25), E5716–E5725 (2018). https://www.pnas.org/content/115/25/E5716
5. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations (2015)
6. Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., Packer, C.: Snapshot serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Sci. Data **150026** (2015). https://doi.org/10.1038/sdata.2015.26
7. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. *CoRR* (2018). arXiv:1804.02767
8. Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J.R.R., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Duerig, T., Ferrari, V.: The open images dataset V4: unified image classification, object detection, and visual relationship detection at scale. CoRR (2018). arXiv:1811.00982
9. Redmon, J., Divvala, S.K., Girshick, R.B., Farhadi, A.: You only look once: unified, real-time object detection. CoRR (2015). arXiv:1506.02640
10. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. CoRR (2016). arXiv:1612.08242
11. Redmon, J.: Darknet: open source neural networks in c (2013–2016). http://pjreddie.com/darknet/
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
13. Kumar, P., Sengupta, K., Lee, A.: A comparative study of different color spaces for foreground and shadow detection for traffic monitoring system. In: Proceedings of the IEEE 5th International Conference on Intelligent Transportation Systems, pp. 100–105. IEEE (2002)
14. Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Duerig, T., Ferrari, V.: The open images dataset v4: unified image classification, object detection, and visual relationship detection at scale (2018). arXiv:1811.00982
15. Krasin, I., Duerig, T., Alldrin, N., Ferrari, V., Abu-El-Haija, S., Kuznetsova, A., Rom, H., Uijlings, J., Popov, S., Kamali, S., Malloci, M., Pont-Tuset, J., Veit, A., Belongie, S., Gomes, V., Gupta, A., Sun, C., Chechik, G., Cai, D., Feng, Z., Narayanan, D., Murphy, K.: Openimages: A public dataset for large-scale multi-label and multi-class image classification (2017). Dataset available from https://storage.googleapis.com/openimages/web/index.html

Evaluation and Summarization of Student Feedback Using Sentiment Analysis

Neeraj Sharma and Vaibhav Jain

Abstract Educational data mining facilitates educational institutions to discover useful patterns and apply them to improve the overall quality of education. Analysing student feedback may help institutions to enhance student's learning capabilities in the classroom. We propose a student feedback analysis system that helps in identifying sentiments from student reviews, and it further helps in generating the summary of feedback. It is implemented using sentiment analysis and text summarization techniques. Based on our evaluation, the lexicon-based approach did better than traditional machine learning-based techniques. Finally, we were able to generate a precise summary of student feedback.

Keywords Educational data mining · Sentiment analysis · Machine learning · Text summarization

1 Introduction

India possesses the second largest educational system in the world. Government educational agencies like UGC and AICTE are placing a lot of emphasis on institutions to maintain Student Response Systems and Student Feedback Systems and insist them to take necessary measures based on the feedback from the students to improve the quality of education. It is crucial to understand the patterns generated by student feedback data to effectively improve the performance of the institution. There is also a requirement to automate the student feedback system in order to handle a large amount of data and analyse them effectively.

Educational data mining refers to practices, mechanisms and researches invented for automatically extracting meaning from large repositories of data generated by or

N. Sharma · V. Jain (✉)
Institute of Engineering and Technology, Devi Ahilya Vishwavidhyalaya, Indore, India
e-mail: vjain@ietdavv.edu.in

N. Sharma
e-mail: 17mse106@ietdavv.edu.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_35

related to student's learning activities in educational settings. There are many techniques like machine learning and sentiment analysis which can be used to accomplish the task of educational data mining.

Sentiment analysis techniques can be applied on the student feedback to extract the opinions and subjectivity from the student feedback.

We have proposed a feedback analysis system based on sentiment analysis and text summarization techniques which generates a feedback summary (performance of teacher for the subject taught). We have applied various machine learning techniques like Naive Bayes [14], support vector machines [15] and a lexicon-based approach for classifying student feedback. The classified student feedbacks are later used to generate a precise summary. In order to ensure the correctness of the feedback summary, we chose the technique yielding the maximum classification accuracy, and then with the help of a text summarization technique, we get the more accurate feedback summary.

2 Background

Sentiment analysis is a discipline of research in the natural language processing domain that deals with analysing the opinions, sentiments and emotions present in a text body by the computational treatment of subjectivity in the content.

2.1 Machine Learning-Based Sentiment Analysis

There has been a considerable proportion of researches in the past that rely on the machine learning-based mechanisms for discovering the features related to statements in the given body of text. The Naive Bayes classifier works based on the fundamentals of the Bayesian probability and an assumption that the feature probabilities are independent of each other.

A support vector machine is a classifier defined by a separating hyperplane. Given labelled training data as input, the algorithm outputs an optimal hyperplane which categorizes new examples.

2.2 Lexicon-Based Sentiment Analysis

A sentiment lexicon is a list of words which are generally labelled as positive or negative based on their meanings and orientations. For many applications, it would be beneficial to be able to determine the strength of the sentiments expressed in the text rather than just the binary polarity (positive/negative).

Researchers and analysts need to be able to observe the degree of change in the emotion of the text in order to understand whether the author (one who wrote the sentence) is happy, sad or angry.

2.3 Extractive Summarization

Extractive summarization approaches generate summaries by selecting a subgroup of the sentences in the original text. These summaries encompass the most essential sentences of the input. Input can either be a single document or multiple documents. For understanding how summarization system works, we describe three separate tasks that all the summarizers perform (Fig. 1).

2.3.1 Intermediate Representation

Every summarization system constructs some intermediate representation of the text it aims to summarize and discovers salient content on the basis of this representation. There are two approaches based on the representation: *topic representation* and *indicator representation*. Topic representation approaches transform the text into an intermediate representation and construes the topics discussed in the text. Indicator representation techniques express every sentence as list of features having some importance such as sentence length and position in the document.

Sentence Score When the intermediate representation is generated, an importance score is accredited to each sentence. In topic representation methods, the score of a sentence represents how well the sentence explains some of the most important topics of the text. In the indicator representation methods, the score is calculated by aggregating the evidence from different indicators.

Summary Sentences Selection Eventually, the summarizer system selects the top n most important sentences to produce a summary. Some methodologies make use of greedy algorithms to select the important sentences while others may consider the selection of sentences as an optimization problem where a subset of sentences is chosen considering the constraints that it should maximize the overall importance and coherency and minimize the redundancy.

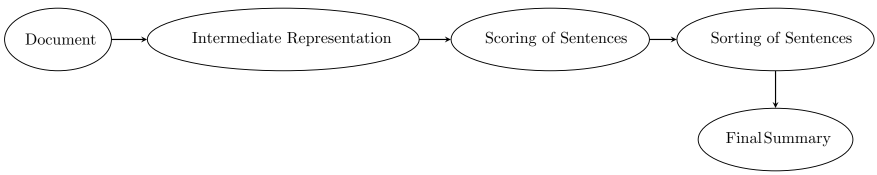


Fig. 1 Steps involved in extractive text summarization

3 Related Work

3.1 Sentiment Analysis and Text Summarization on Educational Data

Here is a brief overview of related work. Chaplot et al. [1] have provided an algorithm that predicts and pinpoints the time when the student is going to drop out. They have used click stream logs and forum posts data from the Coursera MOOC. Their work allows the course instructor team to take necessary steps to prevent or reduce student attrition during the course. However, the amount of student feedback taken into account for carrying out the experiment is limited to only one course, and hence, the sample space is limited.

Wen et al. [2] have studied the correlation between the student sentiment ratio measured based on the daily forum post and the number of students dropping out each day. They have used the MOOC post course surveys to collect student opinions. Sentiment analysis is carried out in two phases in the post retrieval phase and opinion estimation phase, and in the post retrieval phase only the messages containing the topic keywords are used. From the results, a keyword list for the topics is constructed. Opinion estimation for all the distinct posts is related to any topic, i.e. the topic sentiment ratio on a particular day.

In another work, Meishan Hu et al. [4] proposed two approaches to incorporate comments into document summarization. The first approach scores document sentences based on keywords derived from comments, while the second approach scores document sentences and comments altogether. The latter approach showed significant improvement over the techniques using only the keywords. However, their work does not address the similarity between two sentences in a given piece of text which can prove to be an important aspect for text summarization.

Maite Toboado et al. [6] presented a lexicon-based approach for detecting the emotions reflected in a body of text. They have proposed a system known as Semantic Orientation CALculator. The system uses dictionary of words labelled along with their semantic orientation.

Efstratios Kentopoulos et al. [7] proposed the deployment of an efficient technique on the basis of ontology-based approach for sentiment analysis of the posts on Twitter.

After reviewing the above work, we found that the majority of them used machine learning-based sentiment analysis for mining educational data, but none of them addressed the summarization of educational data reviews.

4 Proposed System

Our proposed solution consists of developing a student feedback evaluation system and a feedback summarizer which use sentiment analysis and text summarization techniques. The proposed architecture is shown in Fig. 2 The architecture consists

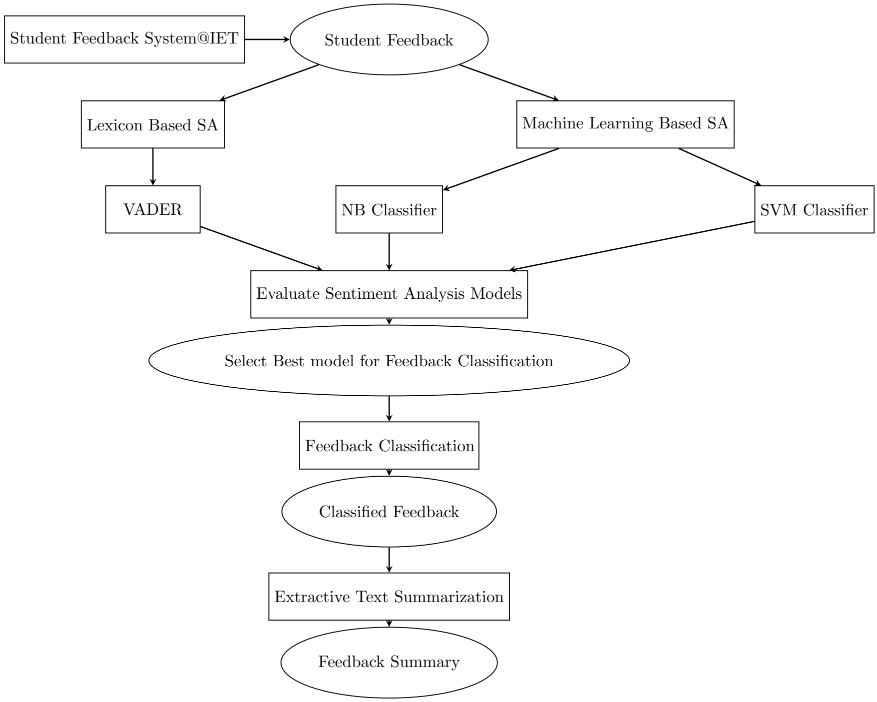


Fig. 2 Proposed system architecture

of various techniques applied at each step of the work in order to achieve the goal of generating a feedback summary based on the student feedback data. The input to our work is the student feedback which is taken using an existing online feedback system in IET-DAVV [9]. This system collects student's feedback, and relevant data can be extracted from this system for further analysis.

The various steps involved in the architecture are as follows:

- Collection of feedback through existing online feedback collection system.
- Then, we apply various lexicon-based and machine learning-based approaches sentiment analysis techniques on the student feedback data.
- Evaluating all the sentiment analysis models.
- Applying the optimal model for sentiment analysis obtained from the former step to the feedback.
- Using the results of feedback classification for feedback summary generation through extractive summarization techniques.

Table 1 Sample feedback format

Section A	Very Poor	Poor	Good	Average	Excellent
1. Teacher's subject knowledge	0	0	0	0	0
2. Compliments and coverage of course	0	0	0	0	0
3. Compliments theory with practical examples	0	0	0	0	0
4. Teacher's computer/IT skills	0	0	0	0	0
Section B 1. Result of test declared within two weeks of it being conducted	0 Yes	0 No	0 No Comments		
2. Adequate number of assignments/cases given	0 Yes	0 No	0 No Comments		
Section C					
What are the strengths of the teacher	————————————				
What are the areas of weakness in teacher	————————————				
Any other suggestions (regarding curriculum, subject and faculty)	————————————				

4.1 Existing Feedback Application

As shown in Table 1, the online feedback form contains a mix of free response and quantitative (also called Likert scale) questions. The Likert scale questions are easy to evaluate as the answers to these questions are scores only. The free responses in the feedback form contain three questions asking about the strengths, weaknesses and suggestions about the respective faculty for whom the feedback is being taken. We get the data in this format, for applying sentiment analysis we have focused on the subjective comments from section C.

4.2 Sentiment Analysis

We used two sentiment analysis approaches for evaluating student feedback:
Machine Learning-Based Approach The machine learning-based approach of sentiment analysis typically involves three steps—text preprocessing, vectorization and feedback classification.
Text Preprocessing Preprocessing is the process of cleaning and organizing the data and preparing the text for classification.
Vectorization Vectorization is the process of dividing the data and then converting it into numerical form. The feedbacks are extracted from the input file and stored in a feature matrix, and then, it is count vectorized using a vectorizer, i.e. it is converted to a document-term matrix.
Feedback Classification The machine learning approaches treat sentiment analysis to be a classification problem, wherein based on the training dataset

(consisting of observations whose categories are known) the problem is to identify which class does new feedback belong to.

4.2.1 Lexicon-Based Approach

The lexicon-based approach for sentiment analysis comprises three steps that are tokenization, dictionary-based scoring and classification.

Tokenization Tokenization is the process of splitting the text into tokens. These tokens could be paragraphs, sentences or individual words. Tokenization needs to be done because amongst different lexicon-based approaches, some of the methods perform sentence level sentiment scoring, while others perform word level sentiment scoring.

Dictionary-Based Scoring Each token (word/sentence based on the approach used) is assigned a score based on the scores of the lexicons in the dictionary,and then, the normalized sentiment score for the sentence is determined.

Classification Once the sentiment scores of the sentences are normalized, the sentences whose polarity is found to be less than are classified as negative, those with the polarity equivalent to zero are classified as neutral and sentences with the polarity greater than zero are classified as positive.

Method Lexical approaches aim to map words to sentiment by constructing a dictionary or a lexicon of sentiments. This dictionary can be used to judge the sentiment of the phrases and sentences. Sentiments can be categorical such as positive, negative, neutral or they can be numerical like a range of intensities or scores. Valence Aware Dictionary for sEntiment Reasoning (VADER) is the lexical sentiment intensity analyser used while classifying the sentiments of the feedback using the lexicon-based approaches. Emotional intensity is very arbitrary since it depends on whom you ask. To counter this, VADER enlists not just one but a number of human raters from the Amazon Mechanical Trunk and averages their rating for each word. The greatest utility provided by VADER sentiment analysis is that not only words, but also emoticons like :-), acronyms like LOL and slangs like *meh* also get mapped onto intensity values. For the feedback supplied as input emotional intensities or sentiment scores are measured on a scale from -4 to $+4$ where -4 is the most negative while, $+4$ is the most positive. The midpoint 0 represents a neutral sentiment. The sentiment score of each word as listed in the VADER dictionary is summed up to obtain the overall sentiment score of the sentence. The normalized sentiment score for the entire sentence is determined by using the following normalization formula:

$$X/\sqrt{X^2+\alpha} \tag{1}$$

where X = sum of the sentiment values of all the constituent words in the sentence. α = Normalization parameter set to 15.

Table 2 Example of adjusted sentiment scores for words

Domain words	Original score	Adjusted score
Punctual, knowledge, depth, case, industrial and complete	0.0	+4
Irregular and inappropriate	0.0	−4

Table 3 Example of sentence score after adjustment of sentiment score

Source feedback	Sentence score	Adjusted sentence score
He has in-depth knowledge of the subject and is punctual	0.0	0.9

Adjusting Sentiment Scores for Educational Domain Table 2 shows the sentiment scores of some of the lexicons were adjusted so as to make them fit for the educational domain. Table 3 shows the adjusted sentiment score obtained after adjusting the sentiment score of domain-specific words. This helped us in generating the correct summary of the feedback.

4.3 Feedback Summarization

Extractive text summarization techniques have been used in our work, to accomplish text summarization. The technique makes use of the TextRank [17] algorithm to assign scores to different sentences in the given text.

5 Implementation and Results

We used Google Colab [18] platform for the experiment purpose. The platform itself used 12GB NVIDIA Tesla K80 GPU. We have used Python3 for developing our feedback evaluation system. A brief list of libraries used are as follows: Pandas [10], Nltk [12], Scikit-learn [11] and Gensim [13].

5.1 Implementation

Student feedbacks are collected using existing online student feedback system as per the format specified in Table 1. The collected feedbacks are subjected to sentiment analysis. Sentiment analysis has been performed using machine learning algorithms for classification such as multinomial Naive Bayes, support vector machine(linear) and support vector machine(radial basis function). We also used lexicon-based approach VADER for feedback classification. Based upon the comparative study, the best method is selected to perform feedback classification, and the classified feedbacks are considered for feedback summary generation using TextRank algorithm.

5.2 Results

Classification performance depends greatly on the characteristics of the data to be classified. There is no single classification technique that works best on all the given data. Various empirical tests have been performed to compare classifier performance and to find characteristics of data that determine classifier performance. The different parameters used for the evaluation of the classification accuracy are accuracy, precision, recall and F1-score (Table 4).

Above are the results of the experiments performed on the IET and Coursera MOOC student review dataset [16]. It can be seen that among all the three machine learning algorithms used for sentiment classification, the multinomial Naive Bayes model predicts the best accuracy. Also, as the dataset size increases, the accuracy of the model improves. Multinomial Naive Bayes gives the highest accuracy when trained using 70% dataset. A saturation in the values of the accuracies can be noticed when dataset size increases. For MOOC dataset, linear SVM gives the highest accuracy when trained using 70% dataset (Table 5).

Results of different performance metrics are taken on three different ratios of train/test data partition. Constant accuracy is witnessed in case of multinomial Naive Bayes, whereas the values of precision, recall and F-score increase with increase in size of training data (Table 6).

Performance metrics are obtained during the experimentation using the lexicon-based approach. We have got a descent increment in the values of all the metrics as compared to all the three methods used previously. There has been a significant growth in the value of accuracy from 82.4% using the multinomial Naive Bayes algorithm to 93% using the lexicon-based approach.

Table 4 Results of ML-based classification on datasets

Classification techniques	Training = 50% Testing = 50%		Training = 60% Testing = 40%		Training = 70% Testing = 30%	
	DS1	DS2	DS1	DS2	DS1	DS2
MNB	82.4	91.14	80.3	91.02	83.07	90.98
SVM-L	77.3	91.14	79.5	91.20	80.3	91.29
SVM-RBF	78.9	91.02	77.25	91.14	77.23	91.22

DS1 =>IET-DAVV dataset DS2 =>Coursera MOOC dataset

Table 5 ML-based model evaluation on IET-DAVV dataset

Data distribution									
Classification techniques	Training = 50% Testing = 50%			Training = 60% Testing = 40%			Training = 70% Testing = 30%		
	Precision (%)	Recall (%)	F-Score (%)	Precision (%)	Recall (%)	F-Score (%)	Precision (%)	Recall (%)	F-Score (%)
Multinomial Naïve Bayes	84	82	79	83	80	76	84	83	80
SVM-Linear	60	77	67	84	80	72	84	80	73
SVM-RBF	62	79	70	60	77	67	60	77	67

Table 6 Results of VADER model on the IET-DAVV student feedback dataset

Accuracy	Precision (%)	Recall (%)	F-Score (%)
93	100	92	95

Source Feedback
Good at clearing the concepts interactive with students deep knowledge of subject practical knowledge. best teacher for c++. makes us understand things very nicely. subject knowledge is excellent little bit angry. good conceptual clarity. his knowledge about programming is ex-cellent and he deliver tiny to tiny concepts to students. sir is the best to teach computer programming they know very minute and important things regarding subject.
Generated Summary
His knowledge about programming is excellent and he deliver tiny to tiny concepts to students.

The above box shows an example of the feedback summary generated for a given student feedback.

6 Conclusion

We aimed at automating feedback evaluation system for existing feedback application and generating a summary of feedback collected. We used sentiment analysis for feedback classification. The lexicon-based sentiment analysis technique, VADER showed an accuracy of 93% and an F1-score of 95%, which is higher as compared to the ML-based methods. In addition, the precision, recall and F-score are also higher. In the learning-based approach, we have considered different supervised machine learning techniques, Naïve Bayes (multinomial), SVM-Linear and SVM-RBF. We then considered three different schemes for data distribution, i.e. 1:2, 3:2 and 7:3. The generated feedback summary obtained is able to a precise overview of feedback evaluated.

References

1. Chaplot, D.S., Eunhee R., Jihie, K.: Predicting student attrition in MOOCs using sentiment analysis and neural networks. In: AIED Workshops, vol. 53 (2015)
2. Wen, M., Yang, D., Rose, C.: Sentiment analysis in MOOC discussion forums: what does it tell us? In: Educational Data Mining (2014)
3. Hutto, Clayton J., and Eric Gilbert. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." Eighth international AAAI conference on weblogs and social media. 2014

4. Hu, M., Sun, A., Lim, E.P.: Comments-oriented document summarization: understanding documents with readers' feedback. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM (2008)
5. Turney, P.D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics (2002)
6. Taboada, M., et al.: Lexicon-based methods for sentiment analysis. Comput. Linguist. **37**(2), 267–307 (2011)
7. Kontopoulos, E., et al.: Ontology-based sentiment analysis of twitter posts. Expert Syst. Appl. **40**(10) 4065–4074 (2013)
8. Mihalcea, R.: Graph-based ranking algorithms for sentence extraction, applied to text summarization. In: Proceedings of the ACL Interactive Poster and Demonstration Sessions (2004)
9. Institute of Engineering and Technology. Devi Ahilya Vishwavidhyalaya, Indore www.ietdavv.edu.in
10. Pandas Python library. https://pandas.pydata.org/
11. Scikit-learn is a machine learning software library freely available for python programming languages. https://scikit-learn.org/
12. NLTK is a library to deal with natural language text for english written in Python programming language. https://www.nltk.org
13. Python library for topic modelling, document indexing and similarity retrieval with large corpora. https://pypi.org/project/gensim/
14. Naive Bayes for Machine Learning. https://machinelearningmastery.com/naive-bayes-for-machine-learning/
15. Support Vector Machines for Machine Learning. https://machinelearningmastery.com/support-vector-machines-for-machine-learning/
16. 100k Coursera's Course Reviews. https://www.kaggle.com/septa97/100k-courseras-course-reviews-dataset
17. Mihalcea, R., Tarau, P.: Textrank: bringing order into text. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (2004)
18. Colaboratory–Google. https://colab.research.google.com

Predicting Competitive Weightlifting Performance Using Regression and Tree-Based Algorithms

Uma Mahajan, Anup Krishnan, Vineet Malhotra, Deep Sharma, and Sharad Gore

Abstract Athletes are rigorously trained and practice for gradually improving their fitness, sports-specific skills, and motor qualities to achieve success in various competitions. Periodic testing of physiological and physical fitness components is done to assess the progress of athlete in the training. Athletes are also selected for the upcoming competitions using these measurement results. However, there is a lack of an objective method for evaluating and selecting athletes for competitions. Machine learning predictive models are developed for important feature selection and prediction of weightlifting performances at the national or international level and how much total weight an athlete could lift in the future competitions. Predictive modeling requires historical data; hence, five-year fitness testing and actual competition performances were used for the prediction. Various fitness components are often highly correlated, and multicollinearity is a major issue in developing regression models for predicting performances. Hence, regularized ridge, lasso, and elastic net regression and tree-based random forest algorithms were developed to predict the performances along with the conventional statistical regression models. Boruta algorithm was applied to identify the most important predictors of the weightlifting performances. Back and handgrip strength, lower body average power, upper body explosive power and strength, age, and height were the most important predictors of the weightlifting performance. Thus, random forest with confirmed most important variables is the best model to predict the total weight-lifted and the national/international performances in selecting the weightlifters objectively for the upcoming competitions.

Keywords Random forest · Regularized regression · Physiological and physical fitness feature selection

U. Mahajan (✉) · S. Gore
JJT University, Rajasthan, India
e-mail: umasmahajan@gmail.com

U. Mahajan · A. Krishnan
Army Sports Institute, Pune, India

V. Malhotra · D. Sharma
Armed Forces Medical College, Pune, India

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_36

1 Introduction

Weightlifting is known as a strength sport, and first competitive performances among male weightlifters were documented long back in Egypt, ancient Greece, and China. In modern sports competitions era, weightlifters lift the maximum weight in two types of lifts, namely snatch, and clean and jerk. At least one successful performance in three attempts is recorded as the overall total weight-lifted in the competition within a bodyweight category of the weightlifters [1]. Weightlifting performance depends on height, weight, power, muscle strength, techniques, and kinematics [2–12]. Rigorous training and practice are the keys to successful performances in the competitions. Periodic testing of athletes is done to evaluate their progress in training by monitoring various physiological and physical fitness components [13]. These fitness components are assessed during different training phases and practice for the alteration of training as per the need of physiological and physical demands of the individual athlete. During pre-competition phases, the assessment is carried out to select athletes for the near future competitions at state- or national-, or international-level competitions. Thus, the objective is to identify the significant fitness components associated with the weightlifting performance and predict the fitness and performance of Indian weightlifters for the upcoming competitions.

Regression models are widely used for predicting consequences, and least square method was developed by Carl Gauss in 1821 [14]. Regression models are useful to identify the important physiological and physical fitness components to predict sports performance. Regression model estimates the outcome/dependent variable using independent variables or predictors. The choice of regression model depends on the type of outcome variable. Linear regression is developed when the outcome variable is quantitative, i.e., continuous or discrete [15]. The logistic regression is developed using logit function when the outcome variable is qualitative, i.e., binary categorical [16]. The univariate regression model is developed for each independent variable, and only significant independent variables are included in the multiple regression modeling. Multicollinearity is a major issue in multiple regression modeling. Independent variables are called multicollinear if they are moderate to highly correlate with each other. The conventional statistical approach is to calculate the variance inflation factor (VIF) for each predictor variable and remove the variable with the highest VIF one by one from the model [17].

In the machine learning approach, Ridge, lasso (least absolute shrinkage and selection operator), and elastic net regressions are useful to deal with the issue of multicollinearity in the data [18]. Ridge regression shrinks the parameters asymptotically to zero and handles multicollinearity. Lasso regression is also similar to the ridge, but it shrinks the non-important parameters to zero and hence also useful in feature selection. Elastic net regression is a mixture of ridge and lasso regression, and it handles multicollinearity and also shrinks the less important parameters to zero. Multiple linear regression minimizes the sum of squares of residuals or error (SSE). Ridge, lasso, and elastic net regressions are also known as regularized or penalized regression since they produce additional error along with the regular sum of squares of residuals. Ridge and lasso regression shrinks the less important variables and hence add penalty terms L2 and L1 norm, respectively. The error functions in linear and regularized regression are as follows

Multiple linear regression:

$$\mathrm{SSE} = \sum_{i=0}^{n} (y_{i\text{actual}} - \hat{y}_{i\text{predicted}})^2 \tag{1}$$

Ridge regression:

$$\mathrm{SSE} = \sum_{i=0}^{n} (y_{i\text{actual}} - \hat{y}_{i\text{predicted}})^2 + \lambda \sum_{i=0}^{n} \beta_i^2 \tag{2}$$

Lasso regression:

$$\mathrm{SSE} = \sum_{i=0}^{n} (y_{i\text{actual}} - \hat{y}_{i\text{predicted}})^2 + \lambda \sum_{i=0}^{n} |\beta| \tag{3}$$

Elastic net regression:

$$\mathrm{SSE} = \sum_{i=0}^{n} (y_{i\text{actual}} - \hat{y}_{i\text{predicted}})^2 + \lambda(1-\alpha) \sum_{i=0}^{n} \beta_i^2 + \alpha \sum_{i=0}^{n} |\beta| \tag{4}$$

In above SSE, lambda (λ) is the strength of the penalty also known as the hyperparameter since it is not measured directly. Elastic net reduces to ridge regression when α is zero and reduces to lasso when α is 1. α varies from 0 to 1, while λ varies from 0 to ∞. λ needs to be tuned to get the optimal penalized regression model. Lasso and elastic net regression shrink the less important parameters to zero, hence useful for variables or feature selection. Lower the error in actual and predicted values, the model fit is better.

Apart from regularized regression, the random forest which is a tree-based algorithm is useful for both classification when the outcome variable is categorical and regression when the outcome variable is quantitative. It split the variable into further branches or leaf nodes using optimal criteria for splitting [19]. The random forest, as well as regularized regression, has an inbuilt mechanism for selecting important variables. However, random forest and ridge regression could not give the parsimonious subset of variables, i.e., they could not reduce the parameters to zero [20]. Boruta algorithm, which is also a random forest-based algorithm, is useful to overcome the issue of the non-parsimonious solution and select variables as confirmed, rejected, and tentatively important [21].

Considering some of these limitations of conventional statistical methods, we used machine learning algorithms to identify important physiological and physical fitness predictors of the weightlifting performance, to predict the performance by the level of competition and also to predict the maximum weight the athlete could lift in the future competition.

2 Methods

2.1 Physiological and Physical Testing

Male weightlifters were trained in competitive weightlifting between the year 2013 and 2017 in an international-level sports training institute in India. During this period, weightlifters were voluntarily participated and tested for various physiological and physical fitness components using a standard battery of test. Weightlifters those who were not doing training due to injury were excluded from the periodic testing. The details of the test battery are as follows

(a) Anthropometric measurements

- Age (years) was recorded on the day of testing
- Standard methods were used to measure the height (cm) and body weight (kg) of the weightlifters.

(b) Human performance laboratory tests

- Wingate lower body anaerobic test was used to measure relative peak power (W kg^{-1}) and average power (W kg^{-1}).
- Astrand cycle ergometer test was used to assess the aerobic capacity by measuring sub-maximal oxygen consumption VO_2 (mL min^{-1} kg^{-1}).
- Bioelectrical impedance analysis methods were used to carry out body composition analysis and estimate the body fat (%).

(c) Field tests

- Standing start 30-m sprint (s) was carried out to measure the acceleration speed.
- Standing broad jump (cm) was used to measure lower body explosive power and strength.
- Medicine ball throw 3 kg (cm) was used to measure the upper body explosive power and strength through distance covered during the test.
- Back squat 1 repetition maximum (RM) and 5 RM (kg) were used to measure the maximum strength of the lower body.
- Number of sit-ups completed in 30 s was used to measure the core abdominal strength.

- Hyperextensions 30 s (number) was carried out to measure lower back strength.
- Back pull-ups (number) was measured to assess back muscle strength and endurance.
- Handgrip (kg) was measured to assess the arm strength of right and left hand.

2.2 *Weightlifting Performances*

Weightlifters participated in various national and international competitions between 2013 and 2017. Weightlifting performance was recorded using the following two outcome variables

(a) Total weight-lifted (kg) during the competition was documented as a total of weight-lifted in snatch, and clean and jerk lifts.
(b) Competition level performance was recorded as the type of medal won in the competition and the level of competition. Weightlifters who participated in national competitions and won any medal bronze/silver/gold were categorized as national performers and the rest of the weightlifters who participated in the international competitions and won any medal were categorized as international performers.

2.3 *Predictive Model Development*

Three datasets, namely human performance laboratory data, field testing data, and the competition data, were merged using a unique identification number of weightlifters. Descriptive analysis was done to explore outliers and missing values. The last observation carried forward (LOCF) method was used to impute missing values within an individual weightlifter [22, 23]. Weightlifters with heavy body weight >96 kg were excluded from the analysis. Total of 286 cases, 17 independent variables, and 2 outcome variables were used for the analysis.

The Kolmogorov–Smirnov test was used to test the normality of outcome as well as independent variables. Median and inter-quartile range (IQR) were calculated for independent variables. Spearman's rank correlation was used to test the correlation between the independent and outcome variables. Correlogram was plotted using Spearman's rank correlation. The median test was used to compare the fitness components by competition level, i.e., national/international performer. Statistical significance was set at p-value ≤ 0.05. R 3.6.1 and R studio were used for the data analysis and model development.

Seventy percent of cases were selected randomly with the replacement for training the models and 30% for validating the model performance. Tenfold cross-validation with five repetition resampling methods were applied for improving the performance of the model in training dataset. The multiple linear and logistic regression models were developed for total weight-lifted and categorical outcome variable weightlifting performances by competition level using training dataset, respectively. Regularized regression, namely ridge, lasso, and elastic net regression models, were developed for both regression and classification. In predictive models with the continuous outcome variable, root mean square error (RMSE) of each model is compared while training and validating the models. For models with the categorical outcome variable, the accuracy of classification using confusion matrix and Kappa statistics were used to compare and validate the model performance [24]. Higher accuracy and Kappa statistics value indicates a better fit of the model.

Boruta algorithm was developed to identify the most important variables or predictors of the weightlifting performances using both continuous and categorical outcome variables. Random forest regression and classification models were developed using all variables, confirmed important variables only and non-rejected, i.e., confirmed and tentatively important variables only. In all 14 predictive models were developed, 7 each for feature selection and prediction of continuous and categorical weightlifting performance variables.

3 Results

3.1 Profile of Weightlifters

Out of total 286 cases, 171 (59.8%) were national performances and 115 (40.2%) were international performances. The median age of male weightlifters was 17.6 years (IQR = 15.52, 20.87), median height was 163.0 cm (IQR = 159.0, 169.0), and median weight was 67.3 kg (IQR = 60.0, 75.2). Summary statistics for other independent variables are shown in Table 1.

In the median analysis, it was observed that the international performer weightlifters were aged 16.6 years old, i.e., significantly younger than the national performers aged 18.9 years ($p = 0.001$). In international performers, standing broad jump was 271 cm ($p = 0.012$), peak power was 12.59 W kg^{-1} ($p = 0.004$), and the average power was 8.99 W kg^{-1} ($p < 0.0001$) which were significantly higher than the national performer weightlifters (Table 1).

3.2 Multicollinearity Diagnostics

In Spearman's test, the rank correlation coefficients along with significance are shown in Fig. 1. Squares with blue and red color represent significant positive and negative

Table 1 Anthropometric, physiological, and physical fitness profile of weightlifters

Fitness variables	All ($n = 286$)	National performances ($n = 171$)	International performances ($n = 115$)	p value
	Median (IQR)	Median	Median	
Age (years)	17.67 (15.52, 20.87)	18.86	16.58	**0.001**
Height (cm)	163 (159, 169)	163.50	162.00	0.665
Weight (kg)	67.9 (60.0, 75.2)	69.00	65.30	0.228
Standing start 30 m sprint (s)	4.3 (4.20, 4.43)	4.30	4.30	0.100
Standing broad jump (cm)	267 (253, 278)	264	271	**0.012**
Medicine ball throw 3 kg (cm)	1323 (1207, 1459)	1334	1307	0.228
Back squat 1 RM (kg)	180 (160, 190)	175	180	0.537
Back squat 5 RM (kg)	155 (135, 170)	150	160	0.562
Sit-ups 30 s (No.)	34 (31, 36)	34	34	0.836
Hyperextensions 30 s (No.)	31 (29, 32)	31	31	0.182
Back pull-ups (No.)	15 (12, 18)	15	15	0.367
Grip strength right hand (kg)	52.7 (45.2, 64.0)	55.20	49.90	**0.005**
Grip strength left hand (kg)	50.7 (42.9, 62.0)	53.40	47.90	**0.001**
VO_2 (mL min^{-1} kg^{-1})	44.86 (38.08, 50.69)	44.45	46.27	0.469
Peak power (W kg^{-1})	12.16 (11.12, 13.01)	11.86	12.59	**0.004**
Average power (W kg^{-1})	8.75 (8.17, 9.17)	8.45	8.99	**<0.0001**
Body fat (%)	11.8 (10.1, 14.0)	12.00	11.50	0.101

IQR inter-quartile range
Bold indicates P-value <= 0.05 is considered as statistically significant

correlations, respectively, while squares without color represent nonsignificant correlations. Darker shades represent strong significance, while lighter shades represent weak significance.

It was observed that for age, height, weight, and fat, medicine ball throw 3 kg, handgrip strength, and back squat 1 RM and 5 RM correlation coefficients

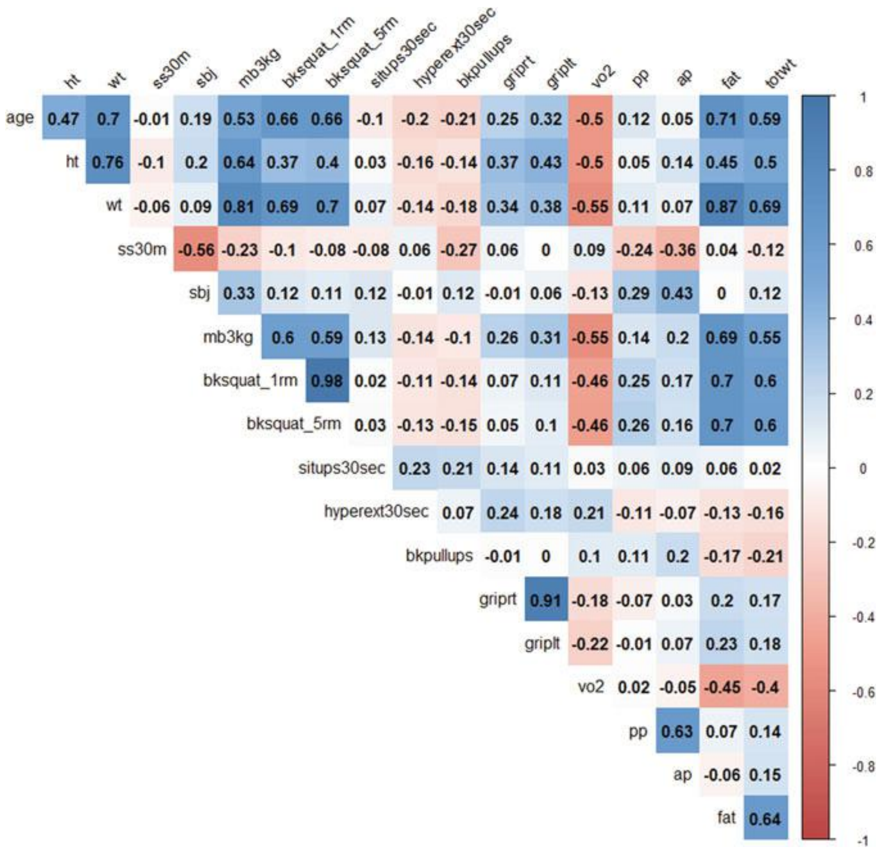


Fig. 1 Correlogram of fitness components and the total weight-lifted. age = Age, ht = Height, wt = Weight, ss30 m = Standing start 30 m sprint, sbj = Standing Broad Jump, mb3kg = Medicine ball throw 3 kg, bksquat_1rm = Back squat 1 RM, bksquat_5rm = Back squat 5 RM, situps30sec = Sit-ups30 s, hyperext30sec = Hyperextensions 30 s, bkpull-ups = Back pull-ups, griprt = Grip strength right hand, griplt = Grip strength left hand, vo2 = VO_2, pp = Peak power, ap = Average power, fat = Body fat, totwt = Total weight-lifted

were 0.500–0.690, and these independent variables were significantly positively correlated ($p \leq 0.05$) with the total weight-lifted in the competitions. The sub-maximal/maximum oxygen VO_2, back pull-ups, and hyperextension 30 s were weakly negatively significantly correlated with the total weight-lifted performance ($p > 0.05$).

High multicollinearity was observed since the predictor variables were significantly highly correlated with each other. In multiple linear regression model, severe multicollinearity was observed in back squat 1 RM (VIF = 32.17), for 5 RM (VIF = 30.21), weight (VIF = 22.06), and fat, right/left grip (VIF > 9.5).

3.3 Variable Importance in Predicting Total Weight-Lifted

All independent variables were nonsignificant predictors of the total weight-lifted in the multiple linear regression ($p > 0.05$) while in variable importance plot, sit-ups 30 s and hyperextension 30 s were unimportant predictors of the total weight-lifted. In ridge and elastic net regression, medicine ball throw 3 kg was unimportant, while sit-ups 30 s, VO_2, standing broad jump, and handgrip strengths were shrunk to near zero which shows that they were unimportant predictors. In lasso regression, only five variables were important predictors, namely fat, age, weight, back squat 1 RM, and back pull-ups (Fig. 2a–d).

In Boruta algorithm, three variables, namely left-hand grip strength, standing broad jump and peak power, were tentatively important and only two variables sit-ups 30 s and hyperextension 30 s were rejected, i.e., unimportant, and rest of 15 variables were confirmed important for the prediction of total weight-lifted (Fig. 2e).

Among 15 variables, average power, medicine ball throw 3 kg, back squat 1 and 5 RM, body fat along with anthropometric measurements were 100% important predictor variables. VO_2, back pull-ups, and right-hand grip strength were about 90% important, and standing start 30-m sprint was 71% important predictors in predicting total weight-lifted in the competitions (Table 2).

3.4 Variable Importance in Predicting Performance by Competition Level

Age and average power were significant predictors of the performance by competition level in the multiple logistic regression ($p < 0.05$), while in variable importance plot, sit-ups 30 s and back squat 1 RM were unimportant predictors. In ridge, lasso, and elastic net regression, VO_2, height, weight, medicine ball throw 3 kg, standing broad jump, handgrip strength, back squat 1/5 RM, and sit-ups 30 s were unimportant predictors (Fig. 3a–d).

In Boruta algorithm, 3 variables, fat, weight, and back squat 5 RM, were tentatively important and 4 variables, sit-ups 30 s, hyperextension 30 s, standing start 30 m, and VO_2, were unimportant, and rest of ten variables were confirmed important for the prediction of performance by competition level (Fig. 3e).

Among 10 variables, handgrip strength, average power, peak power, and age were 100% important, and height was 98% important variables. Medicine ball throw 3 kg, back squat 1 RM, and back pull-ups were about 80% important, and standing broad jump was 65% important predictors for performances at national and international level (Table 2).

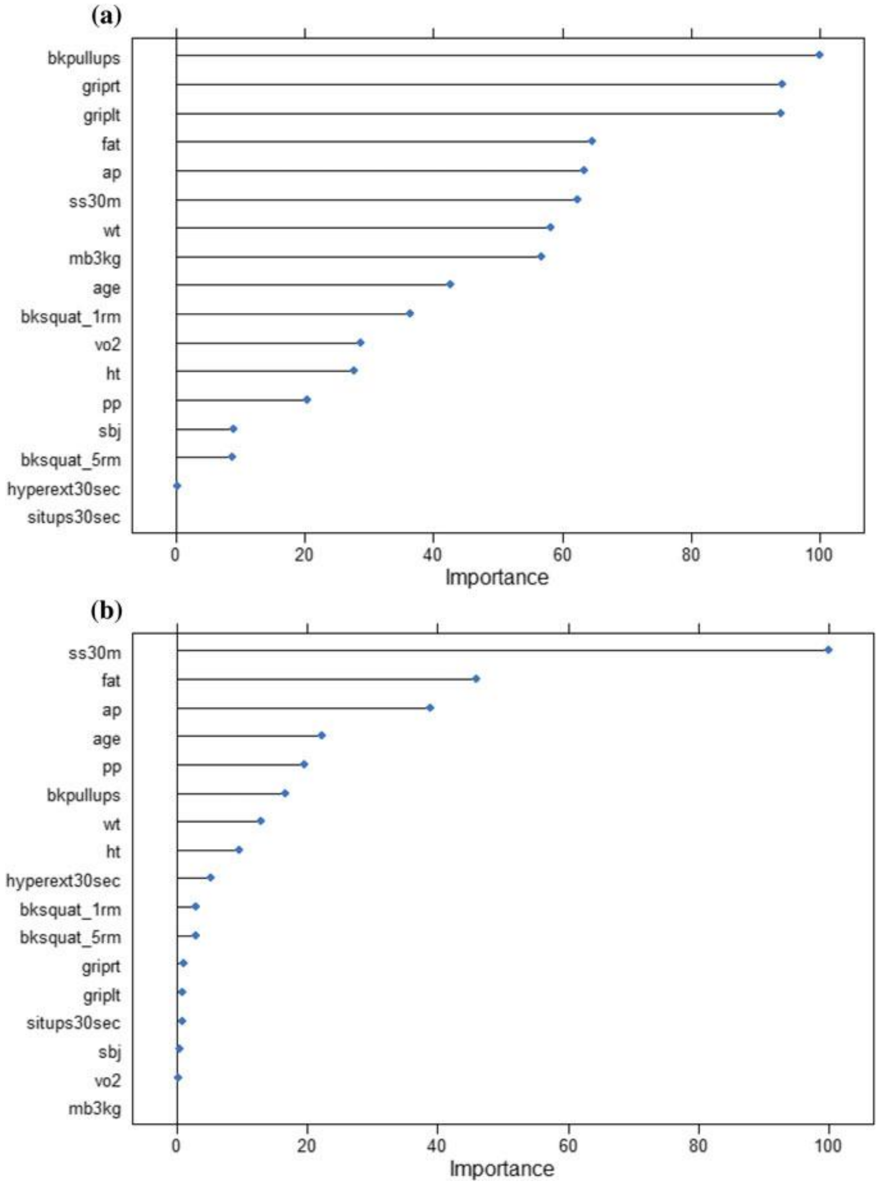


Fig. 2 **a** Variable importance in total weight predicted by multiple linear regression. **b** Variable importance in total weight predicted by ridge regression. **c** Variable importance in total weight predicted by lasso regression. **d** Variable importance in total weight predicted by elastic net regression. **e** Variable importance in total weight predicted by Boruta algorithm

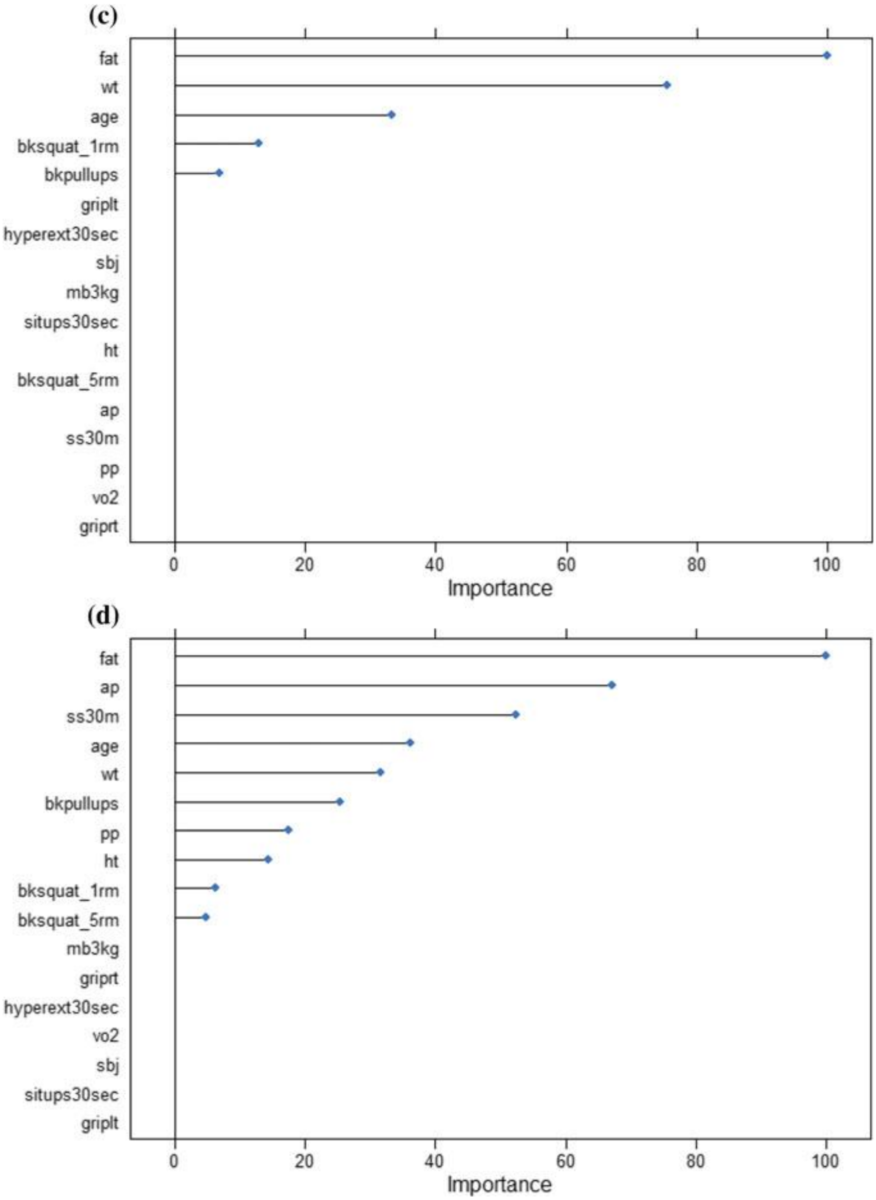


Fig. 2 (continued)

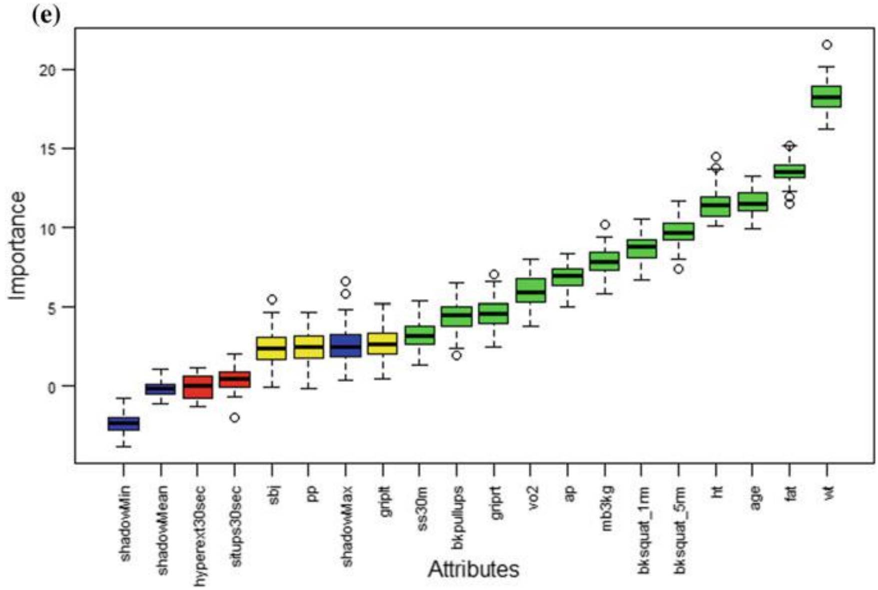


Fig. 2 (continued)

3.5 Model Performance in Predicting Weightlifting Performances

In continuous variable total weight-lifted predictive modeling, RMSE was highest 44.693 and 36.230 for multiple regression models in training and testing datasets, respectively. RMSE was lowest for random forest with 12 confirmed important variables; it was 43.793 and 34.279, respectively, for training and testing datasets (Table 3).

In categorical variable predictive modeling, i.e., national or international performer prediction, the accuracy of prediction was lowest in multiple logistic regression models, which was 60.601% and 69.770% in training and testing datasets, respectively. Accuracy of random forest with all variables was highest 75.135% in training dataset, and it was 74.000% with 10 confirmed variables in the training dataset. Accuracy with 10 variables random forest was highest 75.580% in the testing dataset. Similar results were also observed in the testing dataset, and Kappa statistics were 0.450 and 0.441, respectively, for the random forest with all and 10 confirmed variables in testing dataset (Table 4).

The confusion matrix of random forest with all 17 variables and confirmed 10 variables is shown in Table 4. It was observed that the sensitivity with 10 variables model was 85.71 and 82.14% in the model with all variables. Also, the specificity was 56.67% for both 10 variables and all variables' model (Table 5).

Table 2 Variable importance by regression and tree-based Boruta algorithm

Fitness variables	Variable importance in predicting weightlifting performances					
	Total weight-lifted			Competition level		
	Mean	%	Decision	Mean	%	Decision
Age (years)	11.61	100.00	Confd	13.79	100.00	Confd
Height (cm)	11.44	100.00	Confd	5.57	97.98	Confd
Weight (kg)	18.31	100.00	Confd	2.88	56.57	Tentv
Standing start 30 m sprint (s)	3.20	71.72	Confd	1.79	20.20	Reject
Standing Broad Jump (cm)	2.41	43.43	Tentv	3.27	65.66	Confd
Medicine ball throw 3 kg (cm)	7.82	100.00	Confd	3.80	79.80	Confd
Back squat 1 RM (kg)	8.68	100.00	Confd	4.09	80.81	Confd
Back squat 5 RM (kg)	9.69	100.00	Confd	3.06	59.60	Tentv
Sit-ups 30 s (No.)	0.43	1.01	Reject	2.08	32.32	Reject
Hyperextensions 30 s (No.)	−0.01	1.01	Reject	0.71	0.00	Reject
Back pull-ups (No.)	4.41	88.89	Confd	4.13	82.83	Confd
Grip strength right hand (kg)	4.60	92.93	Confd	7.56	100.00	Confd
Grip strength left hand (kg)	2.68	54.46	Tentv	8.57	100.00	Confd
VO_2 (mL min^{-1} kg^{-1})	5.95	97.98	Confd	1.67	14.14	Reject
Peak power (W kg^{-1})	2.47	47.48	Tentv	8.60	100.00	Confd
Average power (W kg^{-1})	6.83	100.00	Confd	16.03	100.00	Confd
Body fat (%)	13.55	100.00	Confd	3.09	59.60	Tentv

% importance in percentage, *Confd* confirmed, *Reject* rejected, *Tentv* tentative

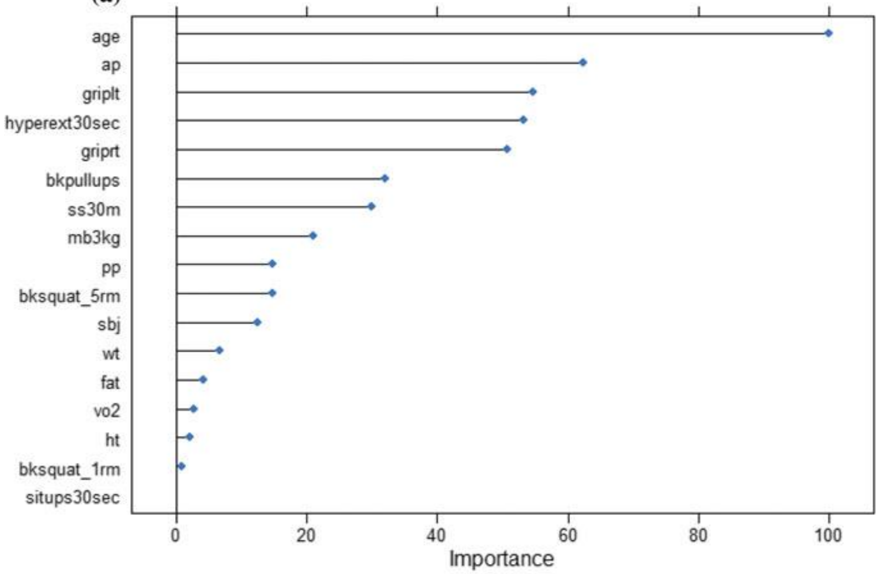
4 Discussion

Multiple linear and logistic regression models' accuracy was poor due to high multicollinearity in the independent variables. Boruta algorithm is better in the identification of variable importance as compared to the regularized regression. It identified lower body average power, upper body explosive power, and strength, handgrip strength, back strength along with age, and height as the most important predictors for the weightlifting performances, namely total weight-lifted and the national/international performers.

RMSE and the accuracy of random forest with all variables were marginally higher as compared to the random forest with confirmed variables. Hence, it is proposed to use the confirmed variables as the predictors of the weightlifting performances to quickly assess the weightlifters in real time. Random forest models are better models as compared to the regularized regression models.

Weightlifters who were national performers and predicted as national performers could be selected for the upcoming national competitions. Similarly, international performers who are also predicted as international performers could be selected for the international competition. Rest misclassified weightlifters could be evaluated on

(a)



(b)

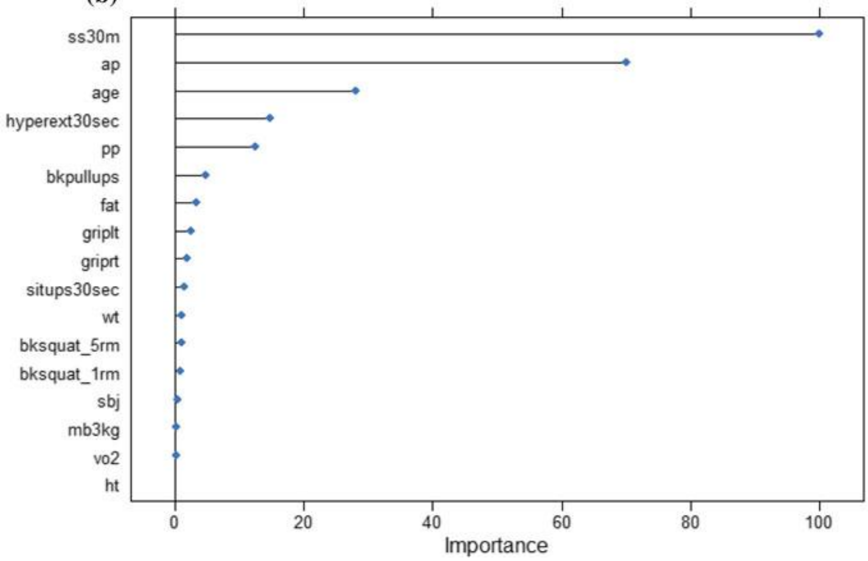


Fig. 3 **a** Variable importance in competition level (national/international performer) predicted by multiple logistic regression. **b** Variable importance in competition level (national/international performer) predicted by ridge regression. **c** Variable importance in competition level (national/international performer) predicted by lasso regression. **d** Variable importance in competition level (national/international performer) predicted by elastic net regression. **e** Variable importance in competition level (national/international performer) predicted by Boruta algorithm

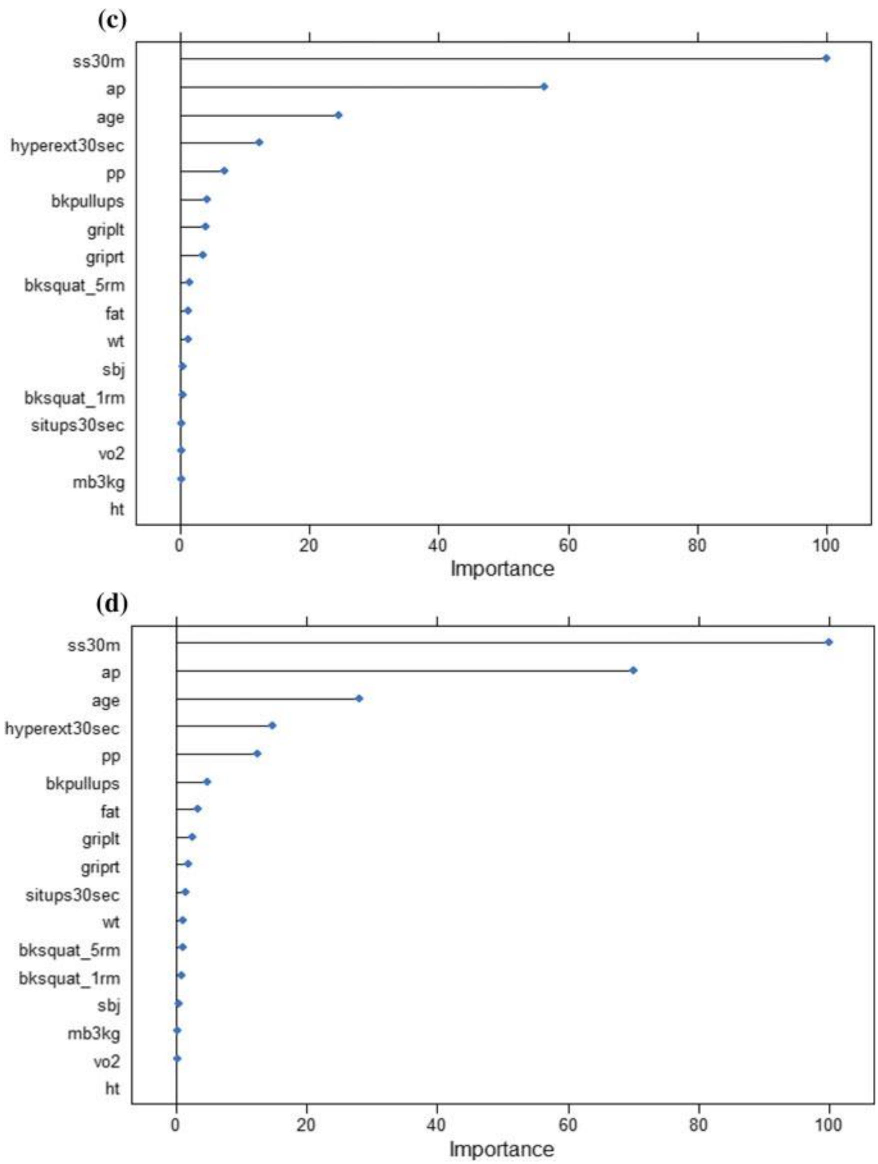


Fig. 3 (continued)

each important fitness component, and the training may be planned considering their physiological and physical fitness needs.

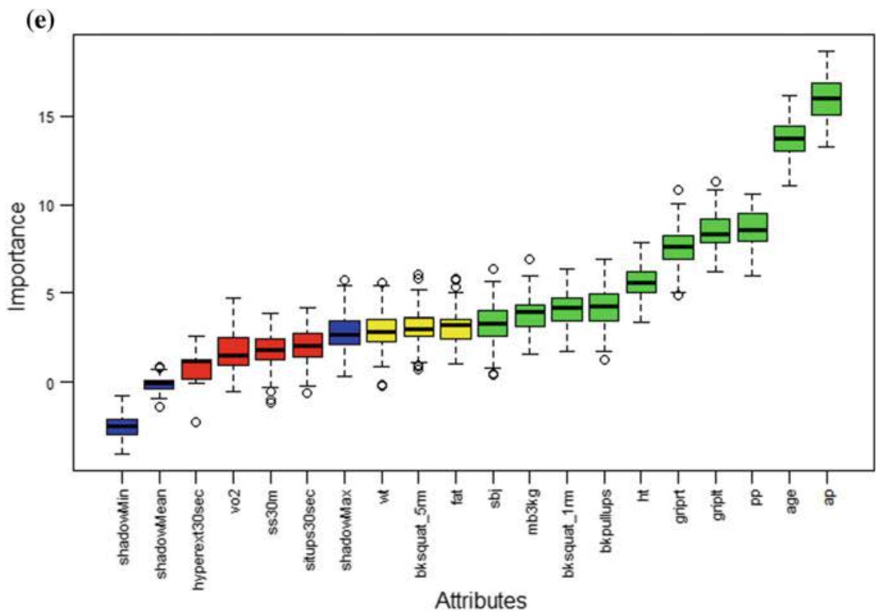


Fig. 3 (continued)

Table 3 Model performance in predicting total weight-lifted

Algorithms	Total weight-lifted	
	Training data	Testing data
	RMSE	RMSE
Multiple linear regression	44.693	36.230
Ridge regression	43.469	34.852
Lasso regression	43.388	36.013
Elastic net regression	43.296	35.280
Random forest with all variables	42.730	34.464
Random forest with non-rejected variables	43.820	33.631
Random forest with confirmed variables	43.793	34.279

5 Conclusion

The upper body power, lower body average power, height, back, and handgrip strength are the most important predictors of the weightlifting performances. The tree-based machine learning algorithm with confirmed important variables could be used for the prediction of total weight-lifted and the performance by competition

Table 4 Model performance in predicting competition level

Algorithms	Competition level			
	Training data		Testing data	
	Accuracy (%)	Kappa statistics	Accuracy (%)	Kappa statistics
Multiple logistic regression	62.601	0.224	69.770	0.355
Ridge regression	63.521	0.239	72.090	0.395
Lasso regression	62.811	0.229	69.770	0.355
Elastic net regression	63.521	0.239	72.090	0.395
Random forest with all variables	75.135	0.474	75.580	0.450
Random forest with non-rejected variables	75.000	0.476	72.090	0.376
Random forest with confirmed variables	74.000	0.455	75.580	0.441

Table 5 Confusion matrix of final random forest models in testing dataset

Prediction models		Actual performance		Total
		National performers	International performers	
Random forest with all 17 variables				
Predicted performance	National performers	46	13	59
	International performers	10	17	27
Total		56	30	86
Random forest with 10 confirmed variables				
Predicted performance	National performers	48	13	61
	International performers	8	17	25
Total		56	30	86

level. Athletes could be selected objectively at the respective competition level using predicted performances.

Acknowledgements The authors thank Brig PS Cheema, VSM, Ex Commandant, Army Sports Institute, India, Dr. Geetanjali Bhide, Sports Nutritionist, Army Rowing Node, India, all coaches, athletes, and staff at Army Sports Institute, India.

References

1. Olympic weightlifting—Wikipedia. https://en.wikipedia.org/wiki/Olympic_weightlifting
2. Mcbride, J.M., Triplett-Mcbride, T., Davie, A., Newton, R.U.: A comparison of strength and power characteristics between power lifters, Olympic lifters, and sprinters (1999)
3. Huebner, M., Perperoglou, A.: Performance development from youth to senior and age of peak performance in Olympic weightlifting. Front. Physiol. **10**, (2019). https://doi.org/10.3389/fphys.2019.01121
4. Stone, M.H., Sands, W.A., Pierce, K.C., Carlock, J., Cardinale, M., Newton, R.U.: Relationship of maximum strength to weightlifting performance. Med. Sci. Sports Exerc. **37**, 1037–1043 (2005). https://doi.org/10.1249/01.mss.0000171621.45134.10
5. Garhammer, J.: A review of power output studies of Olympic and powerlifting: methodology, performance prediction, and evaluation tests (1993). https://doi.org/10.1519/00124278-199305000-00002
6. Solberg, P.A., Hopkins, W.G., Paulsen, G., Haugen, T.A.: Peak Age and performance progression in world-class weightlifting and powerlifting athletes. Int. J. Sports Physiol. Perform. 1–7 (2019). https://doi.org/10.1123/ijspp.2019-0093
7. Storey, A., Smith, H.K.: Unique aspects of competitive weightlifting: performance, training and physiology (2012). https://doi.org/10.2165/11633000-000000000-00000
8. Vizcaya, F.J., Viana, O., Olmo, M.F.Del, Acero, R.M.: Could the deep squat jump predict weightlifting performance? J. Strength Cond. Res. **23**, 729–734 (2009). https://doi.org/10.1519/JSC.0b013e3181a04dc3
9. Fry, A.C., Ciroslan, D., Fry, M.D., Leroux, C., Schilling, B.K., Chiu, L.Z.: Anthropometric and performance variables discriminating elite American Junior men weightlifters. J. Strength Cond. Res. **20**, 861–866 (2006)
10. Lucero, R.A., Fry, A.C., LeRoux, C.D., Hermes, M.J.: Relationships between barbell squat strength and weightlifting performance. Int. J. Sports Sci. Coach. **14**, 562–568 (2019). https://doi.org/10.1177/1747954119837688
11. Ho, L.K.W., Lorenzen, C., Wilson, C.J., Saunders, J.E., Williams, M.D.: Reviewing current knowledge in snatch performance and technique. J. Strength Cond. Res. **28**, 574–586 (2014). https://doi.org/10.1519/JSC.0b013e31829c0bf8
12. Suchomel, T.J., Nimphius, S., Stone, M.H.: The importance of muscular strength in athletic performance. Sports Med (2016). https://doi.org/10.1007/s40279-016-0486-0
13. Bangsbo, J., Mohr, M., Poulsen, A., Perez-Gomez, J., Krustrup, P.: Training and testing the elite athletes. J Exerc Sci Fit. **4**, 1–14 (2006)
14. Gauss, C.F.: Theoria Combinationis Observationum Erroribus Minimis Obnoxiae. Vol 1. Henricus Dieterich (1823)
15. Rao, T.H.: Linear models, pp. 3–18. Springer, New York (1995). https://doi.org/10.1007/978-1-4899-0024-1_2
16. Press, S.J., Wilson, S.: Choosing between logistic regression and discriminant analysis. J. Am. Stat. Assoc. **73**, 699–705 (1978). https://doi.org/10.1080/01621459.1978.10480080
17. Midi, H., Sarkar, S.K., Rana, S.: Collinearity diagnostics of binary logistic regression model. J. Interdiscip. Math. **13**, 253–267 (2010). https://doi.org/10.1080/09720502.2010.10700699
18. Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. **33**, 1–22 (2010)
19. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001). https://doi.org/10.1023/A:1010933404324
20. Granitto, P.M., Furlanello, C., Biasioli, F., Gasperi, F.: Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. Chemom. Intell. Lab. Syst. **83**, 83–90 (2006). https://doi.org/10.1016/j.chemolab.2006.01.007
21. Kursa, M.B., Rudnicki, W.R.: Feature selection with the Boruta package (2010)
22. Gadbury, G.L., Coffey, C.S., Allison, D.B.: Modern statistical methods for handling missing repeated measurements in obesity trial data: Beyond LOCF. Obes Rev (2003). https://doi.org/10.1046/j.1467-789X.2003.00109.x

23. Lane, P.: Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. Pharm. Stat. **7**, 93–106 (2008). https://doi.org/10.1002/pst.267
24. Kuhn, M.: Building predictive models in R using the caret package. J. Stat. Softw. (2008)

Predicting the Primary Dominant Personality Trait of Perceived Leaders by Mapping Linguistic Cues from Social Media Data onto the Big Five Model

P. S. Dandannavar, S. R. Mangalwede, and P. M. Kulkarni

Abstract In today's world of a virtually interconnected society, the number of social media users (Facebook, Twitter, Instagram, etc.) is ever-increasing. Social media has become very popular and one of the main channels of communication with users share different types of information—updates, posts, pictures, etc. Twitter—a microblogging site—has become increasingly popular platform over the last few years with 330 million monthly active users, who on average generate 6000 tweets every second. The continued and growing use of social media has resulted in the generation of very large volumes of social media data, which can be used for research purposes. One potential application is automatic personality prediction—which aims to analyze social data and predict the users' personality traits. This is possible because people inadvertently leave behind linguistic cues to personality in their social data. These cues, if properly mined, provide a short cut to personality detection of users. Using digital records to predict personality offers an alternative solution to overcome the drawbacks of manual methods, which while being accurate are time-consuming and expensive. In this work, an attempt is made to predict the primary dominant trait of leaders using social media data.

Keywords Personality · Leadership · Sentiment analysis · Big five · Personality trait prediction · Machine learning

P. S. Dandannavar (✉) · S. R. Mangalwede
Department of CSE, KLS Gogte Institute of Technology, Karnataka, India
e-mail: padmad@git.edu

S. R. Mangalwede
e-mail: mangalwede@git.edu

P. M. Kulkarni
Department of MBA, KLS Gogte Institute of Technology, Karnataka, India
e-mail: pmkulkarni@git.edu

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_37

1 Introduction

Proliferation of digital platforms has led to a very closely interconnected society. This has also led to social media and social networking sites to become increasingly popular. Social media has become an unavoidable medium/avenue for social interactions and has changed the way people interact. Almost every second person—except the most networking averse—uses one or the other social networking sites. Twitter and Facebook, among others, are very popularly used social networking and microblogging sites. They have become a very integral part of the daily lives of millions of users. Motivation for using social networks differs from one user to another. While some focus on actively disseminating/broadcasting information about themselves, more passive users seek to simply consume information produced by others. The reasons for use of microblogging sites also vary—for communication with family/friends/acquaintances, as content sharing venues, as recommendation services or as sources for real-time news—the major reason being as a medium of communication. Of all the online activities, the most popular is using social media and as reported by Statista [31]—"in 2018, an estimated 2.65 billion people were using social media worldwide, a number projected to increase to almost 3.1 billion in 2021." Penetration of social network worldwide is constantly increasing. As per the rankings based on the number of active accounts as of July 2019, social networking market leader Facebook emerged as the first one to cross 1 billion registered accounts. Worldwide, there are over 2.41 billion monthly active users (MAU) as of June 2019, compared to 2.38 billion MAUs for Q1 2019 (Source: Facebook 7/24/19). The use and users of Facebook have seen an increase year over year. As per estimates, 20 min is the average time spent every Facebook visit; 510,000 comments posted, 293,000 statuses updated, and 136,000 photographs are uploaded every minute on Facebook. Instagram, a photo-sharing app has 1 billion monthly active accounts. With about a four percent increase over the last year, there are 330 million monthly active users (MAUs) of Twitter. As many as 6,000 Tweets are sent every second—resulting in 500 million Tweets being sent every day.

The implication of this unprecedented use of social media is that—there is not only a large pool of engaged users who are active, but also a large volume of content being shared and generated at a continuous rate every second. This content is an extension of an individual's self and reflects (i) the online behavior and (ii) the actual personality, of the users. Data from social media is a very rich source of personality data. It is likely going to show hints of their creator's characters. There is also an increased body of evidence to indicate the presence of links between online behavior and personality. It is possible to identify traits of an individual by analyzing his/her content that can subsequently be used to describe an individual's personality. Over the years, researchers have explored the possibility of using data from social media to infer/predict personality traits. Research in this direction—of predicting personality traits using social media—saw a sudden surge in the mid-2000s due to increasing amount of personal information easily available and publicly accessible on social sites. Various approaches have been used to predict personality traits using online social data and results have indicated that personality can be predicted with a high accuracy. **Automatic prediction of personality (APP)** is the process of automatically inferring the personality traits of a user from data such as text, speech, etc.

Twitter has emerged as the most popular microblogging site with millions of active users. Through short informal text called "tweets," users share different types of information (such as posts, updates, and photographs) with the world using natural language and language use has correlations to personality. Choice of words in textual content is driven by not only the meaning but by psychological factors as well, e.g., attitude, emotions, personality traits, etc. User-generated content is a rich source of information containing cues to personality. Several different types of cues/features have been employed for APP,such as natural language term, linguistic patterns, lexical, syntactic, vocabulary-based features, function words, and individual parts of speech. Works of several researchers have shown that the usage of individual words can reveal the personality of the user. It is thus possible to analyze several such textual cues to predict personality of users.

With its origin in the Latin word persona, the word "personality" means "mask." Every individual is unique because of his/her personality—which is a combination of traits and patterns. A person's character impacts his/her conduct and drives him/her to feel, think, and act in a particular way. An individual's personality can be assessed using one from among the several approaches. The Big Five (BF) personality model, Myers–Briggs Type Indicator (MBTI), and Keirsey Temperament Sorter (KTS) are the three commonly used approaches for assessing personality types in the field of psychology. Over the last 50 years, the BF also known as five-factor model (FFM) has become a standard in psychology and is the most widely used personality assessment model. It describes the human personality in terms of five basic traits called the OCEAN traits—**openness** (the extent to which a person is open to experiencing different activities), **conscientiousness** (tendency to act in an organized or thoughtful way), **extraversion** (tendency to seek stimulation in the company of others), **agreeableness** (tendency to be compassionate and cooperative toward others), and **neuroticism** (a person's emotions are sensitive to the person's environment)

A user's online behavior (reflected through the language, more specifically the words in the text messages, status updates, etc., that they share) can be connected to her/his personality in general and with his/her Big Five traits in particular. This work focusses on the primary dominant trait of twitter users using linguistic cues from social media data and the Big Five model.

2 Personality Prediction Based on Social Media Data: A Brief Review

Several studies that have been undertaken in the recent past have revealed that personality traits can automatically be obtained from the text written by the user. Use of textual data from different online environments for automatic personality prediction has been the focus of several researchers. The field of APP using social media content (Facebook, Twitter, etc.) has been researched by various authors [1–30]. The first study to explore the relationship between profile information and PTs was the work carried out by Golbeck et al. [1]. With the core research question of "whether social media profiles can predict PTs?" the authors presented a method—which used the information from Facebook (FB) profiles of users—to accurately predict their personality. Few studies that were undertaken before this focused in general on identifying how personality relates to social networking. Authors in [2] addressed the task of automatic user classification in Twitter based on features drawn from the user information. Machine learning techniques to predict user personality from FB profiles were first used in [3]. The authors used the BF experiment dataset together with numeric prediction techniques to predict the OCEAN traits using 111 attributes. The predictions so obtained were then used to rank individuals in terms of these traits. To understand the relationship between microblogging and PT, the authors in [4] measured the BF traits of 142 participants. Their tweets were also measured over a period of one month. Their work provided an empirical evidence suggesting that it is possible to predict a user's personality from tweets. The authors [5] proposed a methodology—based on multi-task regression—to predict the BF personality traits of Sina Microblog user's using their microblogging behavior. Fabio Pianesi in his work [6] discussed two approaches to APP. Another approach to identifying personality from text was proposed by authors in [7]. They combined common sense knowledge, sentiment polarity, and affective labels for the proposed approach.

Another work that attempted to predict personality traits of Sina Microblog users was that of Dong Nie et al. [8]. The prediction model was established by using the semi-supervised learning regression algorithm. Authors in [9] introduced "PERSOMA"—a personality prediction approach based on BF model which worked with a group of tweets instead of individual tweets. It did consider information about users' profile. PERSOMA predicted accurately in predicting personality traits of tweeters. The work of authors in [10] provided evidence that social media language can be harnessed and be used to predict a user's personality. Work in [11] is a survey of technologies that can be used to deal with human personality. The authors also highlighted the issues that are still open in the field of personality computing and also identified the potential application areas for the same. Kanupriya and Amanpreet [12, 13] reviewed existing literature in the field of user's personality prediction and listed the state of research that exists currently and also provided an overview of the several techniques used for the same and proposed a personality prediction framework based on logistic regression classifier with a minimized error function using stochastic gradient descent. While most automatic personality detection studies focused on

BF model as a framework, work of authors in [14] used the three-factor personality (PEN) model as a framework for automatic personality detection, specifically on the psychoticism trait. Main focus of this research work was an attempt to identify the behavior of criminals by using language cues. A meta-analysis was conducted by authors in [16] to determine the predictive power of digital footprints over BF traits from data that was collected from different social media. The impact that different types of digital footprints had on the prediction accuracy of was also studied. Another work that tried to explore the "dark side" of personality is reported in [17]. Using the age, gender, personality and native language of authors, the authors in [18] applied machine learning to text categorization for authorship profiling. Using frequency of words in text as the feature set, it was shown that linguistic features when used with machine learning methods can be used to determine different aspects of an anonymous author effectively. Author's [19] contribution to the domain under consideration was studying the relation between emotions expressed in status updates of Facebook users and the personality of users. The results showed that users with open personality more frequently expressed their emotions than neurotic users who were found to be more reserved to share their feelings.

3 Study Model

3.1 Resources Used

MyPersonality Dataset
This study and its results are based on a dataset from the myPersonality project [20], which was created in 2007 by David Stillwell and was originally a dataset of more than 3 million Facebook users. It contained a variety of their personal details and was available for download from 2007 to 2012. In 2018, the authors decided to stop sharing the dataset. A smaller dataset (10 k samples) which can still be found for academic research is used in this work. The dataset contains a score (1–5; decimal values being permitted; with 1 being the lowest and 5 the maximum) for each of the OCEAN traits for each user. It is to be noted that the same user can exhibit more than one personality trait at once and hence a score for each trait. This dataset is used for training and testing the model.

NRC (National Research Council Canada) Emotion Lexicon
"The NRC Emotion Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The annotations were manually done by crowd sourcing" [22]. This lexicon is available for research purposes and can be directly downloaded.

3.2 Data Collection Methodology

For the purpose of this study, 26 public figures (well-known celebrities) were randomly selected and the latest 3000 tweets for each of them were used for predicting their BF personality traits. The celebrities chosen were from different fields, viz. entertainment, politics, literature, business, and sports. Researchers are allowed to freely download tweets from Twitter by using a Twitter API.

3.3 Study Model Process

The system consists of four modules: preprocessing, transformation, classification, prediction and visualization. Tweets that are downloaded tweets are not in a form that is suitable for prediction and hence need cleaning, which includes conversion to lower case, removal of special symbols, stop words removal, tokenization etc. This is performed by the preprocessing module. The transformation module includes two parts. It converts the training dataset (myPersonality dataset) and the data that it receives from the previous module into a binary classification problem. It uses the NRC word list to convert the myPersonality dataset into labeled feature vector which acts as the training set. Using the same word list, it converts the preprocessed user data into feature vector. The classification and prediction module which follows uses a trained classifier for classification. There are classifiers for each OCEAN personality trait. The training of data uses machine learning techniques. This training process is aimed to predict traits' category, i.e., predict the primary trait of the user. It uses regression method to get the score predictions for each trait. The model with the lowest fivefold MSE will be selected as regression model. This step is to increase the accuracy of the system. Finally, the visualization module provides a graphical interpretation of the scores.

4 Results

We employed three supervised machine learning algorithms, viz. multinomial Naive Bayes (MNB), support vector machine (SVM), and random forest tree (RFT). With the datasets for training and testing remaining the same, the numeric score prediction of the OCEAN traits for a subset of the celebrities is depicted in Table 1.

Table 1 OCEAN trait scores of celebrities

Celebrity	Method	O	C	E	A	N
Narendra Modi	MNB	3.34	3.76	4.14	3.15	3.37
	SVM	3.34	4.10	4.16	3.04	3.39
	RFT	**3.39**	**3.83**	**3.92**	**3.29**	**3.25**
Shah Rukh Khan	MNB	3.34	3.73	4.13	3.13	3.38
	SVM	3.34	4.10	4.16	3.03	3.39
	RFT	**3.37**	**3.87**	**3.95**	**3.32**	**3.28**
Virat Kohli	MNB	3.33	3.72	4.11	3.13	3.38
	SVM	3.33	4.10	4.15	3.04	3.39
	RFT	**3.40**	**3.77**	**3.91**	**3.28**	**3.26**
Chetan Bhagat	MNB	3.34	3.88	4.15	3.23	3.36
	SVM	3.34	4.10	4.16	3.06	3.39
	RFT	**3.39**	**3.83**	**3.96**	**3.30**	**3.26**

4.1 Reliability Measure of the Predicted Scores of the Study Model

This work employed three measurement procedures (methods) to predict scores of OCEAN traits for the same sample—the results of which are depicted in the Table 1. The reliability of the scores predicted by these methods is measured using the Cronbach's alpha test which gives us a simple way to measure whether or not a score is reliable [23]. The Cronbach's alpha reliability coefficient was computed for the three methods administered to the same sample and the Cronbach's alpha coefficient values for each trait as predicted by each of the three methods is depicted in Fig. 1.

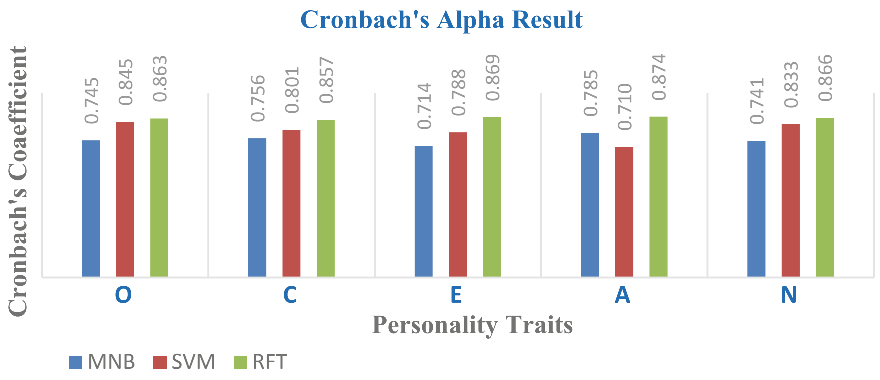


Fig. 1 Graph of the Cronbach's alpha test for each trait

Theoretically, Cronbach's alpha results give a value between 0 and 1. The minimum acceptable value for Cronbach's alpha is 0.70. The general rule of thumb is that a Cronbach's alpha of 0 0.70 and above is good, 0.80 and above is better, and 0.90 and above is best [23]. The Cronbach's alpha values for each trait as predicted by all the methods is well above the minimum acceptable value indicating that the results of the MNB, SVM, and RFT are reliable. Of the three methods, **RFT has higher Cronbach's alpha coefficient values**.

4.2 Standard Evaluation Metric Values

Evaluation using different metrics like confusion matrix, cross-validation, AUC-ROC curve, etc., is an important step of building an effective machine learning model. We used the confusion matrix and associated metric measurements, viz. accuracy, precision, recall and f1-score. Table 2 depicts the values of these measurement for the three methods employed. **RFT scores over the other methods on each metric measurement for each of the traits**.

Based on the result of the Cronbach's alpha test and the values of the metrics based on the confusion matrix, **we consider the results of RFT method**.

4.3 Study Model Validation

Judge, Bono in their work [21] provided a meta-analysis of the trait perspective in leadership research. The results of their work suggested that the most important trait of leaders and effective leadership is ***extraversion***. Open and conscientious individuals are more likely to emerge as more effective leaders. Thus, after extraversion, ***conscientiousness*** and ***openness to experience*** were the strongest traits of leadership. Agreeable individuals tend to be compliant and passive and are less likely to emerge as leaders. ***Agreeableness*** was the least applicable of the Big Five characteristics. **Neurotic** individuals are less likely to be perceived as leaders.

Based on the above work, it is evident that the **primary dominant trait** of effective leaders is **extraversion** followed by **conscientiousness**, **openness to experience**, **agreeableness, and** lastly **neuroticism**. Results of the RFT scores are consistent with the conclusions in [21]. And hence, the model stands validated, as the sample of celebrities chosen is perceived as leaders in their respective fields (Table 3).

Table 2 Metric values for each trait predicted by different methods

	Accuracy			Precision			Recall			f1-score		
	MNB	SVM	RFT	MNB	SVM	RFT	MNB	SVM	RFT	MNB	SVM	RFT
O	74.24	74.19	**78.79**	0.72	0.55	**0.81**	0.74	0.74	**0.79**	0.63	0.63	**0.73**
C	53.38	55.15	**63.94**	0.51	0.3	**0.65**	0.53	0.55	**0.64**	0.48	0.93	**0.62**
E	56.06	56.16	**66.82**	0.51	0.32	**0.7**	0.56	0.56	**0.66**	0.41	0.4	**0.61**
A	53.99	52.98	**63.69**	0.54	0.52	**0.68**	0.54	0.53	**0.66**	0.5	0.39	**0.6**
N	62.53	63.08	**70**	0.54	0.4	**0.75**	0.63	0.63	**0.7**	0.5	0.49	**0.64**

Table 3 RFT scores for personality traits

Celebrity	O [3]	C [2]	E [1]	A [4]	N [5]
Narendra Modi	3.39	3.83	**3.92**	3.29	3.25
Shah Rukh Khan	3.37	3.87	**3.95**	3.32	3.28
Virat Kohli	3.40	3.77	**3.91**	3.28	3.26
Javed Akhtar	3.40	3.78	**3.90**	3.34	3.20
Chetan Bhagat	3.39	3.83	**3.96**	3.30	3.26

5 Discussion

Leadership is related to personality. The trait theory of leadership suggests that leaders are differentiated from other individuals by certain traits. Empirical evidence exists supports this perspective. There have been two meta-analyses of personality traits and leadership. The first meta-analysis was by Lord, DeVader, and Alliger (1986) and the only other meta-analysis was by Judge et al. [21]. The results of both these examinations have shown that there exists a strong correlation between personality traits and leadership. Also, several studies [24–29] have examined the relationship between personality and leadership using different personality trait models and self-rating. The results obtained supported the use of the FFM in exploring the role of personality among leaders.

The general observations are:

- Friendliness (i.e., agreeableness), achievement (i.e., conscientiousness), dominance, assertiveness, and sociability (i.e., extraversion), and self-confidence (i.e., openness) are traits that are positively related to leadership.
- Agreeableness, conscientiousness, and extraversion are positively related leadership effectiveness; and this positive relationship continues over time.
- Leadership effectiveness can be characterized as having individuals score high on traits of openness, conscientiousness, extraversion and assertiveness, and low score on neuroticism.
- Extraversion was mostly strongly related to leadership.
- As individuals score higher on extraversion, they are more likely to report endorsing leadership characteristics.
- Highly agreeable individuals may be less likely to lead.

It can thus be concluded that the **primary dominant trait** of effective leaders is **extraversion** followed by **conscientiousness**, **openness to experience**, **agreeableness and** lastly **neuroticism**.

6 Conclusion

A social media user shares information about himself through posts, updates, pictures, etc., with the world. Such information contains "behavioral residue" and linguistic cues to the users' personality. By simply analyzing the text, the personality of a user can be ascertained, with varying degrees of accuracy. An individual can have more than one personality trait or it may also be possible that a person may not have anyone dominant trait. The present work focused on predicting the primary dominant trait of celebrity twitter users (leaders drawn from cross various domains) by mapping linguistic cues onto the Big Five model, using three different machine learning algorithms. Characteristics that are generally related to leadership are: Intelligence (mental and emotional), personality (conscientiousness, **extraversion**, openness to experience, self-esteem), and integrity. **Extraversion is known to be the most necessary personality trait for any leader to have** [30]. All the three methods predicted "**extraversion**" as the primary dominant trait of each celebrity with varying degrees of accuracy (MNB—56.06, SVM—56.16, and RFT—66.82). This syncs with the conclusions of various studies carried to examine the relationship between leadership and personality traits. It can thus be concluded that extraverted individuals generally tend to lead and that personality traits of an individual can be successfully predicted using social media data.

References

1. Golbeck, J., Robles, C., Turner, K.: Predicting personality with social media. CHI 2011, May 7–12, Vancouver, BC, Canada, ACM (2011). ISBN 978-1-4503-0268-5/11/05
2. Pennacchiotti, M., Popescu, A.M.: A machine learning approach to twitter user classification. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (2011)
3. Wald, R., Khoshgoftaar, T., Summer, C.: Machine prediction of personality from Facebook profiles. In: IEEE IRI (2012). ISBN 978-4673-2284-3/2012
4. Qiu, H., Lin, H., Ramsay, J., Yang, F.: You are what you tweet: personality expression and perception on Twitter. J. Res. Personality **46**, 710–718 (2012)
5. Bai, S., Hao, B., Lim, A., Yuan, S., Gao, R., Zhu, T.: Predicting big five traits of microblog users. In: 2013 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (2013). https://doi.org/10.1109/wi-iat.2013.70
6. Pianesi, Fabio: Searching for personality. IEEE Signal Process. Mag. (2013). https://doi.org/10.1109/MSP.2012.2219671
7. Poria, S., Gelbukh, A., Agarwal, B., Cambria, E., Howard, N.: Common sense knowledge based personality recognition from text. In: MICAI 2013, Part II, LNAI 8266, pp. 484–496. Springer-Verlag Berlin Heidelberg (2013)
8. Nie, D., Guan, Z., Hao, B., Bai, S., Zhu, T.: Predicting personality on social media with semi-supervised learning. In: 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology. IEEE (2014). https://doi.org/10.1109/wi-iat.2014.93. ISBN 978-1-4799-4143-8
9. Lima, A.C.E.S., de Castro, L.N.: A multi-label, semi-supervised classification approach applied to personality prediction in social media. https://doi.org/10.1016/j.neunet.2014.05.020

10. Park, G., Schwartz, H.A., Kosinki, M., Stillwell, D.J.: Automatic personality assessment through Social Media Language. J. Personality Soc. Psychol. (2014). https://doi.org/10.1037/pspp0000020
11. Vinciarelli, A., Mohammadi, G.: A survey of personality computing. IEEE Trans. Affect. Comput. **5**(3) (2014)
12. Sharma, K., Kaur, A.: A review of the existing state of Personality predictions of Twitter users with Machine Learning Algorithms. IOSR J. Comput. Eng. (IOSR-JCE) **17**(4), 35–38. e-ISSN: 2278-0661, p-ISSN: 2278-8727
13. Sharma, K., Kaur, A.: Personality prediction of Twitter users with logistic regression classifier learned using stochastic gradient descent. IOSR J. Comput. Eng. (IOSR-JCE) 17(4), 39–47. e-ISSN: 2278-0661, p-ISSN: 2278-8727
14. Sagadevan, S., Malim, N.H.A.H., Husin, M.H.: Sentiment valences or automatic personality detection of online social network users using three factor model. In: The Third Information Systems International Conference (ISICO2015) (2015)
15. Kedar, S.V., Bormane, D.S.: Automatic personality assessment: a systematic review. In: 2015 International Conference on Information Processing (ICIP), Vishwarkarma Institute of Technology. IEEE (2015). ISBN 978-1-4673-7758-4/15
16. Azucar, D, Marengo, D., Settanni, M.: Predicting the Big 5 personality traits from digital footprints on social media: a meta-analysis. J. Personality Individ. Differ. **124**, 150–159 (2018)
17. Akhtar, R., Winsborough, D., Ort, U., Johnson, A.: Detecting the dark side of personality using social media status updates. J. Personality Individ. Differ. **132**, 90–97 (2018). https://doi.org/10.1010/j.paid.2018.05.026/
18. Argamon, S., Koppel, M., Pennebaker, J.W., Schler, J.: Automatic profiling the author of an anonymous text. Commun. ACM **52**, 119–123 (2009). https://doi.org/10.1145/1461928.1461959
19. Farnadi, G., Sitaraman, G., Rohani, M., Kosinki, M., Stillwell, D., Moens, M.-F., Davalos, S., De Cock, M.: How are you doing? Emotions and personality in Facebook. In: EMPIRE2014: 2nd Workshop on Emotions and Personality in Personalized Services, pp. 45–56 (2014)
20. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. In: Proceedings of the National Academy of Sciences of the United States of America, p. 110 (2013). https://doi.org/10.1073/pnas.1218772110
21. Judge, T.A., Bono, J.E., Ilies, R., Gerhardt, M.W.: Personality and leadership: a qualitative and quantitative review. J. Appl. Psychol. **87**(4), 765–780. https://doi.org/10.1037//0021-9010.87.4.765
22. http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm
23. https://www.statisticssolutions.com/cronbachs-alpha/
24. McCormack, L., Mellor, D.: The role of personality in leadership: an application of the five-factor model in the Australian military. School of Psychology, Deakin University, Burwood 3215
25. Correlations between the Big Five Personality Traits and Leadership. ID 2382763, Skidmore College. www.skidmore.edu/~hfoley/Handouts/Lab1APA6.doc
26. O'Neil, D.P.: Predicting leader effectiveness: personality traits and character strengths. Department of Psychology and Neuroscience Duke University
27. Özbağ, G.K.: The role of personality in leadership: five factor personality traits and ethical leadership. In: 12th International Strategic Management Conference, ISMC 2016, 28–30 October 2016, Antalya, Turkey. https://doi.org/10.1016/j.sbspro.2016.11.019. ISSN 1877-0428
28. Baptiste, B.: The relationship between the big five personality traits and authentic leadership. Walden University (https://scholarworks.waldenu.edu/dissertations)
29. Colbert, A.E., Judge, T.A., Choi, D., Wang, G.: Assessing the trait theory of leadership using self and observer ratings of personality: the mediating role of contributions to group success. The Leadership Quarterly (www.elsevier.com/locate/leaqua)
30. https://sites.psu.edu/leadership/2019/05/16/leadership-with-the-big-five-personality-traits/
31. https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users

Analysis of Users Behavior on Micro-blogging Site Using a Topic

Deepanshu Bharadwaj, Aprna Tripathi, Aman Agrawal, and Mohd. Aamir Khan

Abstract Peoples are generally influenced what everyone else is saying. A person observes a thing according to their nature, whether they are positive or negative kind of person. When these persons tweet other persons influenced by their thoughts. Social media is nowadays a huge platform to spread a news. In this paper, we proposed a method to identify the user's behavior on the micro-blogging site Twitter. Tweets are extracted according to the topic followed by the users behavior is analyzed according to their previous tweets. This method can be used in many ways to stop spamming on social media.

Keywords Tweet analysis · Tweet user behavior analysis · User behavior

1 Introduction

Considerations of individuals differ as indicated by attitude of that user, social condition, and circumstances. Considerations of individuals are caught via web-based networking media organizing applications in a type of content. These writings are likewise relied upon the mental and well-being conduct of a person. It can assist us with predicting the future states of mind since user temperament assumes a critical job in basic leadership. This analysis can decrease costs, exertion, time expected to actualize studies and expectation on a huge scale open. Twitter messages have numerous one of a kind characteristics. Since the length of Twitter message is 140 characters, it, for the most part, seems like a short message with various kinds of shortened forms, short documentation of words which is every now and again utilized in tweets. These contractions and words make tweets more diligently to get it.

D. Bharadwaj (✉) · A. Agrawal
GL Bajaj Group of Institutions, Mathura 281406, India
e-mail: erdeepanshubharadwaj@outlook.com
URL: http://www.glbajajgroup.org

A. Tripathi · Mohd. A. Khan
GLA University, Mathura, India

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_38

Chakra et al. [2] and Sa et al. [8], Twitter utilization is significant, both for understanding human correspondence designs and enhancing the exhibition. Online news has the ability to impact a large number of people groups, and however, there are no approaches to set up what it is valid and what it is fake. We may figure by common sense says it is actual or not yet at some point we have to check the conduct of the user and their past news to anticipate the user conduct and news. Our intention is to get the connection to a tweet from the user and to figure its validity and conduct [1, 12]. It additionally processes different statistics, for example, the general feeling (assessment) of the tweet, thinking about the emoticon, and hashtags utilized in the Twitter.

Twitter started encountering an enormous development in 2007. Everything began at the South By Southwest Intelligent Gathering, where the participants were effectively utilizing Twitter to associate with one another. As per statistics, in excess of sixty thousand tweets were sent each day all through the entire occasion. This was quickened by the nearness of the Twitter group at the meeting. The equivalent happened to a few other tech meetings in that year. The vast majority of the progressions that occurred on Twitter were motivated by the user as opposed to the organization itself. This is on the grounds that the user invented their methods for utilizing the platform proficiently. At first, there was no characterized strategy that could be utilized to answer to different clients or give them a whoop. To handle this issue, the users thought of the @ symbol which is set just before another username. Twitter included it in its usefulness, and up to now, the symbol has turned into a genuine strategy for recognizing another Twitter user. Actually, a few other social media have likewise received the utilization of this logo. Users additionally concocted the hashtags likewise. They needed to distinguish tweets dependent on the points that were being talked about, and today, they have developed to turn into a characteristic piece of Twitter. Another user propelled advancement is the retweets. Shah et al. [9] Twitter clients searched for a way that could empower them to send a tweet that was made by another person without appearing though they are stealing it. Twitter adopted this functionality in 2010, henceforth offering users to repost messages with no feeling of blame. As indicated by the most recent insights, Twitter has 200 million dynamic users every month [6].

If we talking about Twitter, it serves more than 2 billion search queries a day, and 93% of people who follow small- and medium-sized businesses on Twitter plan to purchase from the brands they follow. Sixty-nine percent have already purchased from an SMB because of something they saw on the network. At the starting of 2018, Twitter has 336 million monthly active users. That makes it the third most prominent social networking (pages/visit average site pages saw per visit) is 8.09, where 46% people check the network at least once in a day. Let us take the example of important celebrities that have millions of followers, Barack Obama is the most followed Twitter user (108,206,663 followers and following of 612,300), followed by Katy Perry (107,942,796 followers and following of 221) and Justin Bieber (106,424,989 followers and following of 296,468) [10].

Let us see some interesting statistics about Twitter: firstly, according to the gender demographics. In Great Britain, 59% of men are Twitter users, while 41% are women.

In Japan, 52% women are using Twitter who are of age 20, but after the age of 30, this number decreases to 36%. Generally, 53% of male users use the platform to receive news, compared to 47% of female users. Secondly, according to location demographics. Around 80% of users live outside the USA. 72 million users live in USA, and most of the users live in the urban areas of USA. After USA, Japan has consider highest Twitter users with 50.9 million users followed by the UK (18.6 million users) and Saudi Arabia (13.8 million users). It is estimated that China has 10 million users even though the network is blocked in the country. Thirdly, according to the age demographics. In USA, 40% of users are of age between 18 to 29 years. Few research found that 27% of those aged 30–49, 15% of those aged 50–64, and 8% of those aged 65 uses Twitter. Around 43% of 30–49 years old users use the platform to receive news. Lastly, if we consider income and education demographics then about 30% of Americans who earn $75000 or more use Twitter, around 18% of high school or less and 32% of graduates use Twitter, respectively [5, 7].

2 Proposed Method

Fake tweet is an oxymoron. Fake information on social media is now viewed as one of the greatest threats to democracy, journalism, and freedom of expression. It has weakened public trust in governments and its potential impact on the contentious "Brexit" referendum and so on. In state-of-the-art method detection mechanism, various methodologies came into existence to avoid the spreading of irrelevant information. In the proposed system, users and their responses about the topic are analyzed. The main aim of this work is to identify the sensitivity of the tweets and user's behavior according to that topic. The proposed work is divided into two modules. Verma et al. [11], the first module contains information about the topic which is to be analyzed. In this module, all the tweets are extracted and sentiments of the tweets are analyzed. In the second module, the user score is calculated according to their recent tweets.

The whole process of the proposed automatic user behavior recognition system is shown in Fig. 1. Extraction of Tweets, Score Calculation and Users Reputation.

2.1 Classification of Tweet

First phase, i.e., Extraction of Tweets, is used for the extraction of tweets. For tweet extraction, Twitter API is used. Tweets can be extracted on the basis of hashtags, users, and timestamp. In the proposed approach, tweets are extracted on the basis of hashtags provided in the tweets by the user. We have extracted top five hundred tweets.

In the second phase, i.e., Score Calculation, after extraction of tweets, information about the tweets is extracted. This information is divided into two modules. The first

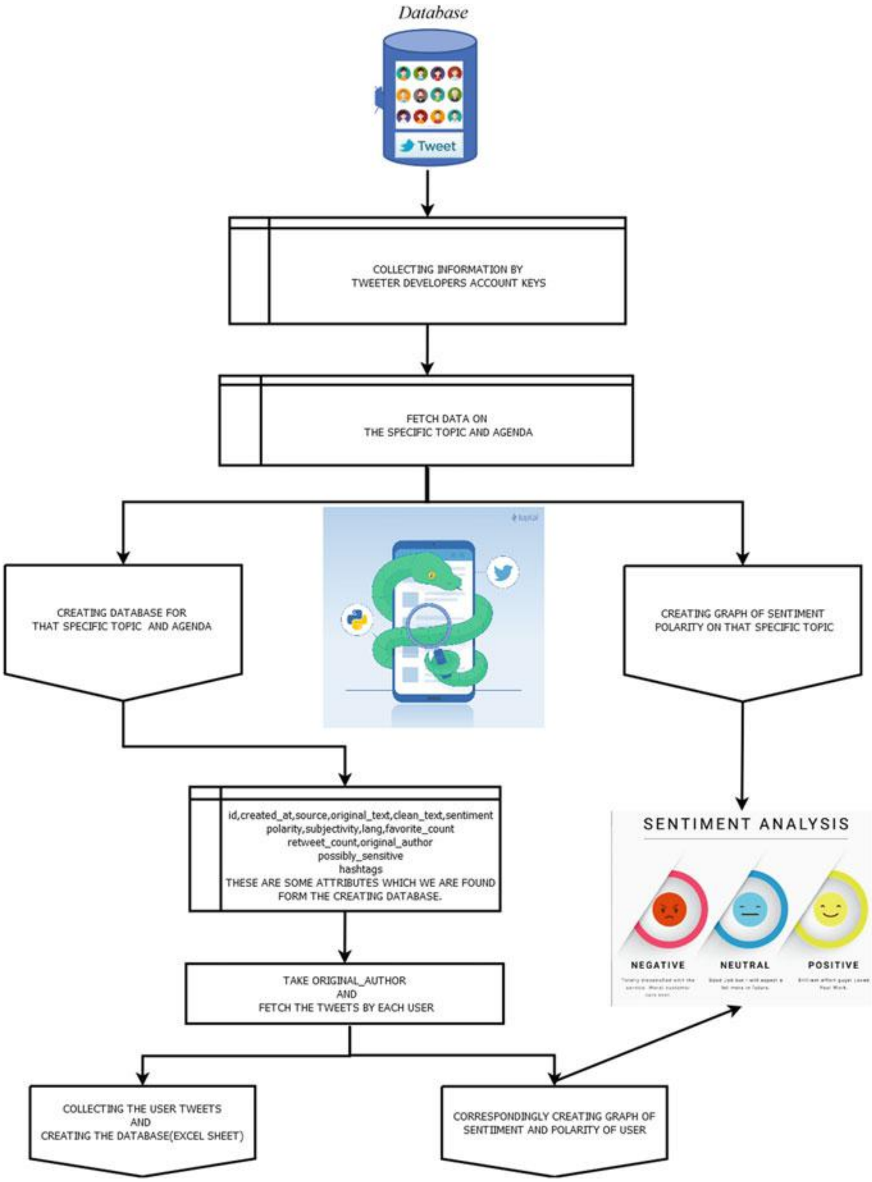


Fig. 1 Framework of the proposed approach

is general information and the second is tweet text information. The extraction of general information from each tweet is extracted, i.e., reputation score, timestamp, and user information. Location may depend on the availability in the tweets if the user shared its location. In the second module, we have calculated the tweet content score [3, 4].

The steps involved to calculate the score in the second module are: stop word removal using POS tagging, stemming using Porter's stemming algorithm, and at the end, sentiments of the tweets are extracted.

2.2 Tweet Content

If we are very much concerned about the tweet content module, in this module we deal with the stop word removal and stemming portion. After applying this procedure, we tend to follow up the tweet verification process. This tweet verification is a necessary process as it gives the following outcome in a negative or positive score. After calculating the tweet score from the positive and negative words, it manipulates the reputation score as it tends to update the score. Afterward, it tends to check the tweet authenticity which is directly possessed by reputation score.

Porter's algorithm was made in the assumption that we don't have a stemming dictionary and that the purpose of the task is to improve IR performance. The program is given an explicit list of suffixes and with each suffix, the criterion under which it may be removed from a word to leave a valid stem.

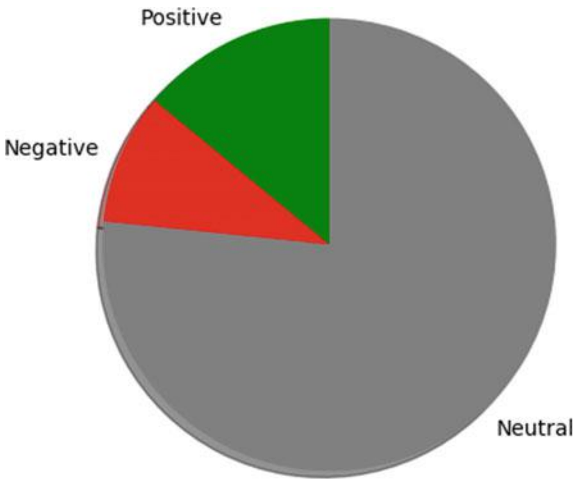
2.3 User Behavior Analysis

After extracting the information about tweets, user's behavior is analyzed. To identify the user's behavior, we have extracted the tweets of the user. After extracting tweets, tweets are cleaned by removing the stop words, emogies, and special symbols. Once the cleaning is completed, we have text information. By using POS tagging and sentiments, analysis score is assigned to every tweet. After assigning the score to every tweets, the overall score is analyzed, which shows the user's behavior.

We have considered the example of the tweets of very famous topic "Chandrayaan2." Atotal of 852 tweets are extracted and analyzed and sentiments of the tweets as shown in Fig. 2. For every tweet, tweet users are also extracted and their previous tweets are extracted and analyzed. According to their previous tweets, the tweet behavior is analyzed of the users'.

After analyzing the sentiments about the topic, the analysis of users is done as shown in Figs. 3, 4, and 5. There are different methods to calculate the sentiments of sentences which are listed in equations.

Fig. 2 Analysis of "Chandrayaan2" tweets



– Absolute proportional difference

$$\text{sentiment} = \frac{P - N}{P + N + Nu} \tag{1}$$

It is bounded between 0 and 1. Score calculated by using this method is affected by non-sentiment-related contents.

– Relative proportional difference

$$\text{sentiment} = \frac{P - N}{P + N} \tag{2}$$

It is bounded between −1 and +1. Score calculated by using this method may tend to cluster very strongly near the scale endpoints.

– Logit scale

$$\text{Sentiment} = \log(P + 0.5) - \log(N + 0.5) \tag{3}$$

It is bounded between −inf and +inf. Score calculated by using this method has the smoothest properties and is symmetric around zero.

Figure 3 depicts the user behaviors according to their previous tweets, and Fig. 3a shows that the most of the previous tweets of the user is neutral followed by the positive than negative, but on this topic, the tweet is analyzed as negative. Figure 3b shows that most of the previous tweets of the user is positive followed by the neutral tweets, whereas in Fig. 3c, d shows that the user's previous tweet sentiments are more negative tweets than the positive tweets.

Figure 4 depicts the user behaviors according to their previous tweets, and Fig. 4a shows that the most of the previous tweets of the user is neutral followed by the negative than positive, but on this topic, the tweet is analyzed as positive. Figure 4d

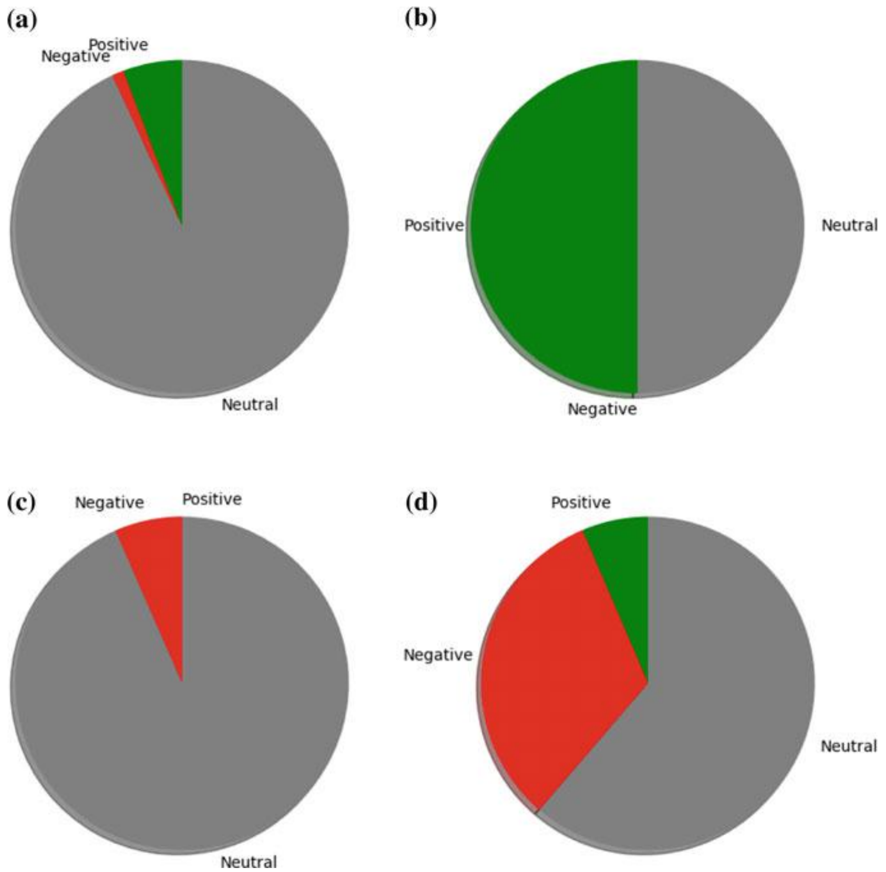


Fig. 3 Sample graphs of user's behavior analysis posted negative tweets on "Chandrayaan2"

shows that the most of the previous tweets of the user is more negative than positive tweets, whereas in Fig. 4b, c shows that the user's previous tweet sentiments are more positive tweets than the negative tweets.

Figure 5 depicts the user behaviors according to their previous tweets, and Fig. 5a, b, d shows that the user's previous tweets sentiments are the same as the user current tweet about the topic. However, Fig. 5c shows that the user behavior is positive according to their previous tweets.

There are the following cases possible between the user behavior and their tweets.

1. Positive User Negative Tweet
2. Positive User Neutral Tweet
3. Positive User Positive Tweet
4. Negative User Negative Tweet
5. Negative User Neutral Tweet
6. Negative User Positive Tweet

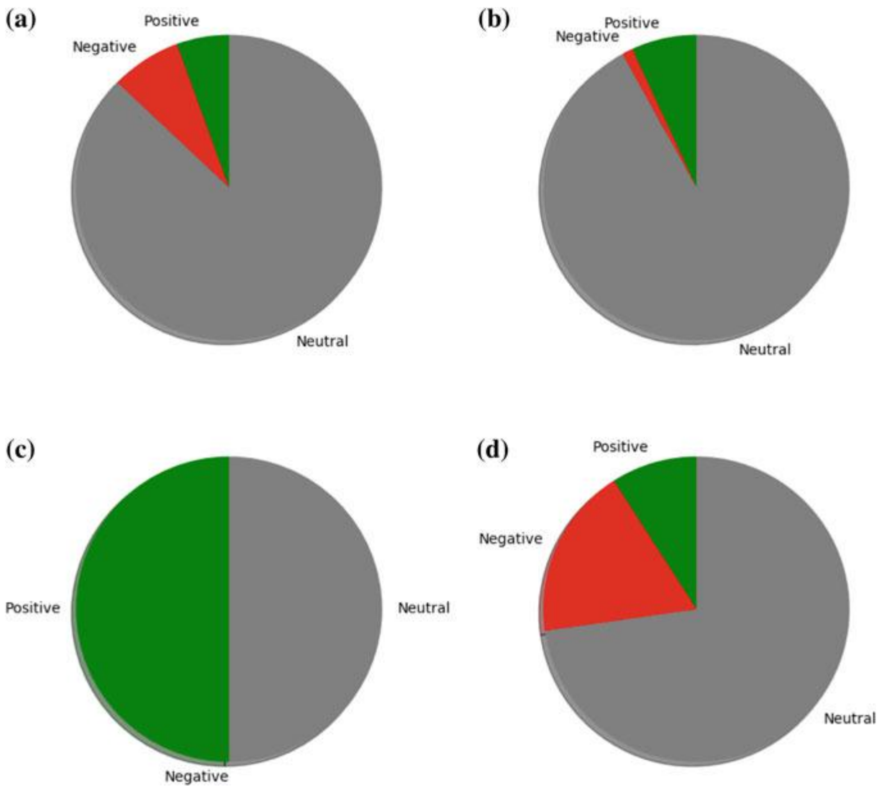


Fig. 4 Sample graphs of user's behavior analysis posted positive tweets on "Chandrayaan2"

7. Neutral User Negative Tweet
8. Neutral User Neutral Tweet
9. Neutral User Positive Tweet.

If the user behavior and their tweet sentiments lie in the same category, then we consider it as normal. In other cases, the user behavior and their current tweets are different, and it can be analyzed in different ways. With the application of this method, we can give the score to the tweets and prioritize it while displaying it in the social media.

3 Conclusion

In this paper, we proposed a model to analyze the user behavior on the micro-blogging site like Twitter. Analyzing the tweets on a particular topic, the users of those tweets are also analyzed. User's previous tweets are analyzed to identify their tweet pattern.

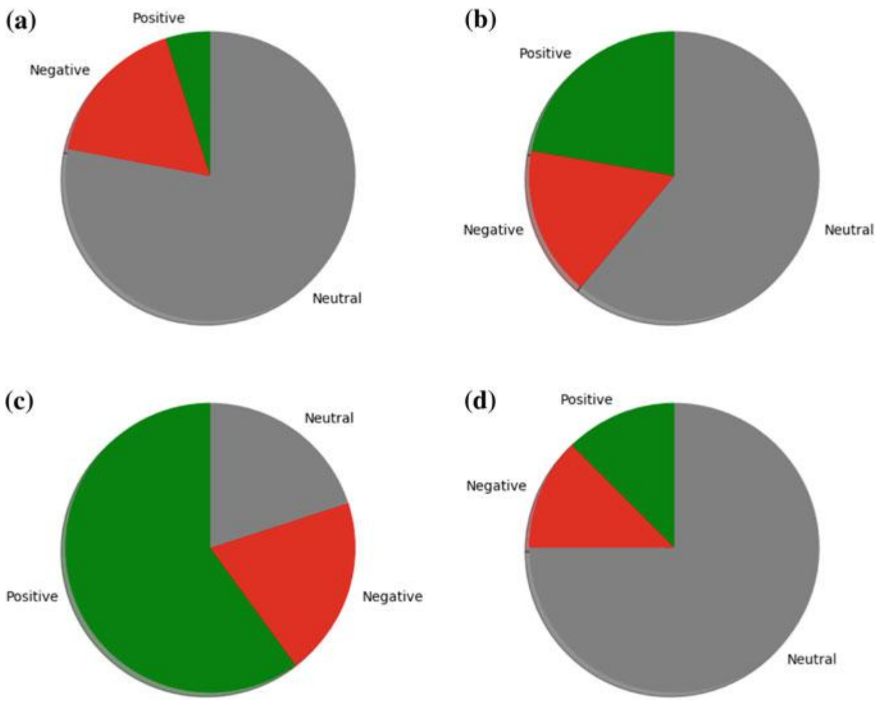


Fig. 5 Sample graphs of user's behavior analysis posted neutral tweets on "Chandrayaan2"

With the application of this approach, we can give the user score and prioritize it on the social media platform. Also, it is used to analyze the topic importance. In the future work, with the help of sentiments and user's behavior, authenticity, likes, and dislikes of the topics are being analyzed.

References

1. Atodiresei, C.S., Tănăselea, A., Iftene, A.: Identifying fake news and fake users on twitter. Procedia Comput. Sci. **126**, 451–461 (2018)
2. Chakra, A., Harpur, L., Rice, J.: Association of an emotional influencer to a post in a social medium (2019). US Patent 10,380,254
3. Gannon, W.: Build a sentiment analysis tool for twitter with this simple python script (2019). http://blog.aylien.com/build-a-sentiment-analysis-tool-for-twitter-with-this-simple-python-script/
4. Jayasekara, D.: Extracting twitter data, pre-processing and sentiment analysis using python 3.0 (2019). https://towardsdatascience.com/extracting-twitter-data-pre-processing-and-sentiment-analysis-usingpython
5. Li, Y., Huang, J., Wang, H., Feng, L.: Predicting teenager's future stress level from micro-blog. In: 2015 IEEE 28th International Symposium on Computer-Based Medical Systems, pp. 208–213. IEEE (2015)

6. Mogadala, A., Varma, V.: Twitter user behavior understanding with mood transition prediction. In: Proceedings of the 2012 workshop on Data-driven user behavioral modelling and mining from social media, pp. 31–34. ACM (2012)
7. Neto, V.R., Medeiros, D.S., Campista, M.E.M.: Analysis of mobile user behavior in vehicular social networks. In: 2016 7th International Conference on the Network of the Future (NOF), pp. 1–5. IEEE (2016)
8. Sa, B.P., Davis, S.M., Vilimonovic, N., Tan, J.Y., Goldsmid, A.P.: Searching for ideograms in an online social network (2018). US Patent App. 10/102,295
9. Shah, U., Finin, T., Joshi, A., Cost, R.S., Matfield, J.: Information retrieval on the semantic web. In: Proceedings of the eleventh international conference on Information and knowledge management, pp. 461–468. ACM (2002)
10. Tien, S.: Top twitter demographics that matter to social media marketers (2019). https://blog.hootsuite.com/twitter-demographics/
11. Verma, P.K., Agarwal, S., Khan, M.A.: Opinion mining considering roman words using jaccard similarity algorithm based on clustering. In: 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–4. IEEE (2017)
12. Vitkova, L., Saenko, I., Tushkanova, O.: An approach to creating an intelligent system for detecting and countering inappropriate information on the internet. In: International Symposium on Intelligent and Distributed Computing, pp. 244–254. Springer (2019)

Machine Learning Techniques for Short-Term Forecasting of Wind Power Generation

Yogesh Gupta and Amit Saraswat

Abstract In recent years, many countries have established their ambitious renewable energy targets to satisfy their future electricity demand with the main aim to foster sustainable and low-emission development. In meeting these targets, the changes to power system planning and operations involve the significant consideration of renewable energy generation through mainly wind energy and solar energy which are more variable and uncertain as compared to the conventional sources (i.e. Thermal and Nuclear Energy). In the present paper, three machine learning methods named as Support Vector Machine, Artificial Neural Network and Multiple Linear Regression are applied to forecast wind power generation on basis of past data of wind direction and wind speed. The impact of input variables such as wind speed and wind direction on wind power generation is investigated and compared.

Keywords Wind power · Artificial neural network · Support vector machine · Regression · Multilayer perceptron

1 Introduction

Nowadays, wind power is one of the most popular clean resources, which is widely used around the world. As the wind is irregular in nature and the production of wind power is uncontrollable. Due to these reasons, the forecasting of wind power generation becomes more problematic and demands high attention for intense research [1–3]. Recently, few notable approaches for wind power generation forecasting have been introduced, especially for countries such as Denmark [2], Germany [3] and America [4] where there is a high portion of wind power in their total capacity. These approaches are based on machine learning techniques. Researchers have used

Y. Gupta (✉)
Depatment of Computer Science and Engineering, Manipal University Jaipur, Jaipur, Rajasthan, India
e-mail: er.yogeshgupta@gmail.com

A. Saraswat
Depatment of Electrical Engineering, Manipal University Jaipur, Jaipur, Rajasthan, India

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_39

artificial neural network, naïve Bayes method and support vector machine in various applications including wind power generation forecasting [5, 6].

Though there is no proper categorization of forecasting of wind power, but it can be divided into two major categories on the basis of utilization: short-term (usually from minutes to hours) and long-term (usually from days to months). Short-term wind power forecasting mainly emphasis on forecasting wind power for minutes or hours ahead, while long-term forecasting of wind power focuses on prediction for days, weeks or months ahead.

The short-term wind power forecasting is helpful in deciding strategy for load scheduling, dispatching decision-making and maintenance scheduling [7]. Therefore, several works have been reported in literature for short-term wind speed forecasting. Generally, all the short-term wind power forecasting techniques can be classified into four major categories as physical type techniques, statistics type techniques, hybrid physical–statistical techniques and artificial intelligence techniques [4]. Physical type techniques forecast the wind power using multiple physical considerations. The historical operational data is not required in this strategy and the forecasting is required to be initiated after the fixing of the wind farm. However, these methods require detailed coordinates of the wind turbine locations with demographic map and power curves of generators to predict the short-term wind power forecasting.

Statistical type techniques use time series historic data to forecast the wind power. Few researchers have also used both types of techniques for forecasting. They used the results of physical techniques as input variables with historical operational data of wind speed, to train the system according to statistical theories. These techniques are called as hybrid techniques. Few approaches have been reported using artificial intelligence, artificial neural networks and fuzzy logic to forecast the wind power generation.

The remaining paper is structured as follows: Sect. 2 discusses related work in this field. The developed methods are presented in Sect. 3. Section 4 experimental results and their discussions. At last, Sect. 5 concludes the paper.

2 Related Work

In this section, some of the recent methods are enlisted which are recently proposed for forecasting of wind energy, i.e., real power production of wind farms in a modern power system. In [8], a most recent and exhaustive literature review on various deep learning methods for forecasting of renewable energy which includes both wind energy and solar energy. However, few detailed reviews on the various reported statistical methods and advancements for wind energy forecasting are presented in [9, 10]. A hybrid method based on adaptive neuro-fuzzy inference system (ANFIS) and an artificial neural network (ANN) is proposed for hourly forecast of wind energy [8]. Two distinct methods named as ANN and hybrid strategy based on statistical and physical methods are presented in [11]. A hybrid evolutionary-adaptive methodology is proposed for wind power forecasting which combines various recent methods such as wavelet transform, the adaptive neuro-fuzzy inference system, mutual information and evolutionary particle swarm optimization [12].

3 Methodology

In this paper, the class of methods based on machine learning concepts is used for determining dependencies and predicting the active power output of wind farm by considering three distinct input variables called as predictors. A brief description over the working of the tree adopted machine learning method is presented in the subsequent subsections.

3.1 Multiple Linear Regression

It is an extension of a simple form of a linear regression which replicas the relationship between two or more independent variables called predictors and a dependent variable. It is applied to forecast the value of an interested output variable which depends on two or more input variables. It is also capable to determine the relative contribution of all predictors on the occurrence of an interested dependent variable. A multiple linear regression model having one response (i.e., output dependent variable) and k number of predictors (i.e., input independent variables) can be mathematically expressed as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k + \varepsilon \quad (1)$$

where Y is a response, i.e., output dependent variable, whereas X_i is the input independent variables called as predictors. ϵ is the term which represent a model residual. However, the terms β_0, β_1, β_2,..., β_k are the coefficients of regression model.

3.2 Support Vector Machine

A support vector machine (SVM) is based on a supervised kind of learning which is developed based on the principles of statistical learning theory. It has also a capability to handle nonlinear binary classification problems. It generates a set of possible hyperplanes in a higher-dimensional space to obtain a suitable class separation. A suitable separation is obtained through a hyperplane having the largest distance to the nearest training data points of any class. It is well-established fact that a larger margin corresponds to a smaller generalization error for a given classifier.

3.3 Artificial Neural Networks

An artificial neural network (ANN) is another machine learning algorithm based on the model of a human brain which consists of millions of neurons. A neuron is a smallest processing unit which sends and processes various electrical and/or chemical signals. These neurons are arranged in a special structure through synapses which allow neurons to pass signals. ANN processes information in same manner as human brain processes information. It includes many interconnected processing units which work together for processing information to generate meaningful results. ANN is not only used for classification but also apply for regression of continuous target attributes. It is typically organized in layers of many interconnected "nodes" which contain a specific activation function. The said layers of an ANN may be classified into three types known as: input layer, hidden layer and output layer as shown in Fig. 1. The activation functions for a node defines the relation between the output of that node with a set of inputs. In the present work, a specific type of activation function is used which is commonly known as "sigmoid" having a characteristic "S" shaped curve or sigmoid curve. The sigmoid activation function is widely used in binary classification [5]. It generates a set of probability outputs between 0 and 1 corresponds to a set of inputs and mathematically defined as follows (Fig. 1):

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{(-x)}} = \frac{e^x}{1 + e^x} \tag{1}$$

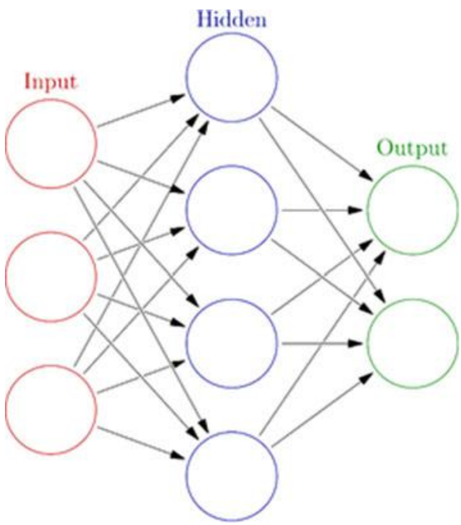


Fig. 1 Representation of artificial neural network

4 Experiments and Analysis

There are five variables in the used dataset: date/time, LV active power, wind speed, theoretical power and wind direction. The data is collected over 10 min of interval. LV active power variable is measured in kW and presents the power generated by the turbine for that moment. Wind speed is measured in meter/second and shows the wind speed at the hub height of the turbine. Theoretical power is measured in KWh. The theoretical power values that the turbine generates with that wind speed which is given by the turbine manufacturer. Wind direction is given in degree, which tells the wind direction at the hub height of the turbine. The statistics of this dataset are presented in Table 1.

Here, the LV active power is assumed as target variable or dependent variable, whereas the other four types of variables are considered as independent variables, i.e., predictors. In this work, the correlations among the variables are determined first, and then the variable LV active power is forecasted based on the historical values of predictors.

Table 2 presents the summary of correlation of predictors with target variable, i.e., VL active power. This table clearly indicates that the wind information such as wind speed and wind direction are useful predictors to predict the value of LV active power. These predictors are highly correlated with target variable. Theoretical power is also correlated with LV active power with value of 0.8123. The other variable date/time is negatively correlated with LV active power, which means target variable and date/time are not correlated, or date/time does not play any significant role in forecasting of LV active power values. Therefore, this paper considers real wind speed, wind direction and theoretical power for forecasting.

Tables 3 and 4 tabulate the summary and statistics of regression for target variable, i.e., LV active power. Table 3 presents that how the target variable (LV active power) co-relates with all dependent variables as a whole and individually as well. Here, R is a coefficient of co-relation and it specifies the co-relation of target variable with all predictors as whole. Then after, two more variants of R are determined from R: R^2 and adjusted R^2. The adjusted R^2 has a value of 0.8362 as given in Table 3. It demonstrates that LV active power has the dependency of 83.62% over all the input independent variables. This value is more than 50% which specifies that almost all the predictors are playing a significant role in forecasting the values of LV active power except one or two. Concurrently, *P value* is also computed and it has a zero value as depicted in Table 3. It directs that R^2 is rejecting null hypothesis.

Table 4 presents the obtained results in terms of a raw coefficients of regression (b), a standard error of b, a standardized coefficient of regression ($b*$), a standard error of $b*$, along with p value and t test (t). The obtained coefficients are compared for an evaluation of their relative contribution of each input independent variable for a desired forecasting of LV active power value. The predictors such as wind speed and wind direction are significant predictors for LV active power as p value is smaller than 0.05 which is 95% of confidence interval. In contrast, date/time is not a significant predictor as p value is greater than 0.05. The unstandardized regression

Table 1 Statistics of used wind dataset

Variable	Mean	Median	Min	Max	Lower quartile	Upper quartile	Std. dev.	Skewness	Std. Err. skewness	Kurtosis	Std. Err. kurtosis
LV active power (kW)	1307.6	825.8	2.4	3618.7	50.67	2482.52	1312.4	0.600	0.010897	−1.164	0.021
Wind speed (m/s)	7.5	7.1	0.0	25.2	4.20	10.30	4.2	0.619	0.010897	0.058	0.021
Wind direction	123.6	73.7	0.0	360.0	49.32	201.70	93.4	0.696	0.010897	−0.747	0.021

Table 2 Correlation matrix

Parameter	LV active power (kW)
Date/time	−0.0118
Wind speed (m/s)	0.9127
Theoretical power (kW)	0.8123
Wind direction	0.8306

Table 3 Statistics of regression summary for LV active power

Variable	Value
R	0.914398939
R^2	0.836125419
Adjusted R^2	0.836115689
p value	0

coefficient (b) for wind speed is 284.8, and it infers that if the other input variables are controlled then increment of wind speed increases LV active power by 284.8 units. Similarly, the value of b is 1.1 for wind direction, which specifies that any change in wind direction alters the value of LV active power by 1.1 units, if all other predictors are constant.

Table 5 shows the distribution fitting to evaluate the fit of observed data to some theoretical distributions. The distributions of target variable LV active power are examined and presented in Table 5. The values of LV active power are divided into 9 slabs ranging from 0 to infinity. Then, observed frequency, expected frequency and observed expected frequency are determined. Further, Chi-Square is determined and found 14278.63920 with p value approximately zero, which indicates Chi-Square is significant at the 0.05 level.

Moreover, two more machine learning methods, called as ANN and SVM, have also been applied to forecast the LV active power. Table 6 represents the ANN-based regression analysis of input dataset. Here, 3-6-1, 3-9-1 and 3-10-1 multilayer perceptron networks are considered. The forecasted values against original values of LV active power for few random observations are depicted in Table 6. It is a very

Table 4 Regression summary for LV active power

	$b*$	Std. Err. of $b*$	b	Std. Err. of b	t (50526)	p value
Intercept			−30423.3	987.6827	−30.8027	0.000000
Date/time	0.054097	0.001807	0.7	0.0228	29.9434	0.000000
Wind speed (m/s)	0.917393	0.001811	284.8	0.5624	506.4834	0.000000
Wind direction	0.009663	0.001807	1.1	0.1254	35.3462	0.000000

Table 5 Distribution fitting for LV active power

(a)

Upper boundary	Observed frequency	Cumulative observed	Percent observed	Cumulative % observed
<=0.00	10,838	10,838	21.448	21.448
500.00	10,196	21,034	20.178	41.626
1000.00	6101	27,135	12.074	53.700
1500.00	4352	31,487	8.612	62.313
2000.00	3419	34,906	6.766	69.079
2500.00	3099	38,005	6.132	75.212
3000.00	3163	41,168	6.259	81.472
3500.00	4118	45,286	8.149	89.622
<Infinity	5244	50,530	10.377	100.000

(b)

Upper boundary	Expected frequency	Cumulative expected	Percent expected	Cumulative % excepted	Observed excepted
<=0.00	8061.414	8061.41	15.953	15.953	2776.59
500.00	5538.537	13,599.95	10.960	26.914	4657.46
1000.00	6982.143	20,582.09	13.817	40.732	−881.14
1500.00	7626.217	28,208.31	15.092	55.824	−3274.22
2000.00	7217.010	35,425.32	14.2823	70.107	−3798.01
2500.00	5917.423	41,342.74	11.710	81.818	−2818.42
3000.00	4203.710	45,546.46	8.319	90.137	−1040.71
3500.00	2587.344	48,133.80	5.120	95.257	1530.66
<Infinity	2396.201	50,530.00	4.742	100.000	2847.80

Table 6 Forecasted values of LV active power using ANN -based regression analysis

Observation no.	Original LV power active value	Forecasted value (3-6-1 MLP)	Forecasted value (3-9-1 MLP)	Forecasted value (3-10-1 MLP)
1	380.048	303.341	340.537	322.464
10	526.816	485.172	487.704	539.888
181	2679.110	2481.816	2503.737	2463.422
4669	1187.152	1082.771	1079.617	1043.185
9592	1035.917	1115.239	1097.196	1117.332
21730	2622.129	2509.983	2494.502	2532.794
32102	3601.659	3557.291	3500.215	3577.035

Table 7 Summary of ANN-based regression analysis

Index	Network name	Training perf.	Testing perf.	Validation perf.	Training algorithm	Error function	Output activation
1	MLP 3-6-1	0.957540	0.955508	0.957379	BFGS 584	SOS	Exponential
2	MLP 3-9-1	0.966029	0.963798	0.966236	BFGS 161	SOS	Exponential
3	MLP 3-10-1	0.956746	0.954553	0.956667	BFGS 202	SOS	Logistic

Table 8 Forecasted values of LV active power using SVM analysis

Observation No.	Original LV power active value	Forecasted value
1	380.048	530.362
10	526.816	592.927
181	2679.110	2391.463
4669	1187.152	1007.351
9592	1035.917	1165.173
21730	2622.129	2499.532
32102	3601.659	3514.703

clear illustration that MLP 3-10-1 network provides more accurate forecasting of LV active power.

A summary of the said ANN-based regression is also presented in Table 7. It clearly depicts that MLP 3-9-1 network provides smaller values of testing error, training error and validation error as compared to other structures of ANN networks. Moreover, the forecasted values of LV active power using SVM analysis are also tabulated in Table 8. Here, a type-1 of SVM regression with total 738 support vectors is applied with radial basis function.

Finally, the accuracy of ANN and SVM for wind dataset is presented in Table 9. This table clearly shows that MLP 3-9-1 network gets 73.29% accuracy, which is higher than the accuracy of SVM, i.e., 69.42% as artificial neural network handles the interaction between the variables effectively as compared to SVM in wine dataset.

Table 9 Comparison of artificial neural network and SVM

Models	Accuracy (%)
Artificial neural network (MLP 3-9-1)	73.29
Support vector machine	69.42

5 Conclusion

Wind power generation forecasting is an important requirement to get prepare for future electricity demand. In this paper, three distinct machine learning methods named as multiple linear regression, artificial neural network (ANN) and support vector machine (SVM) are compared to forecast wind power generation. The said forecasting is based on two significant input variables such as wind speed and wind direction. This paper has also investigated the impact of input variables such as wind speed and wind direction on wind power generation. The presented comparative study shows that ANN (MLP 3-9-1) provides better accuracy as compared to SVM.

References

1. Liu, H., Tian, H.Q., Li, Y.F.: Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms. Energy Convers. Manage. **100**, 16–22 (2015)
2. Lydia, M., Kumar, S.S., Selva Kumar, A.I., Kumar, G.E.: Linear and non-linear autoregressive models for short-term wind speed forecasting. Energy Convers. Manage. **112**, 115–24, 2016
3. Li, P., Guan, X., Wu, J.: Aggregated wind power generation probabilistic forecasting based on particle filter. Energy Convers. Manage. **96**, 579–587 (2015)
4. Hu, X., Martinez, C.M., Yang, Y.: Charging, power management, and battery degradation mitigation in plug-in hybrid electric vehicles: a unified cost optimal approach. Mech. Syst. Signal Process. (2016). https://doi.org/10.1016/j.ymssp.2016.03.004
5. Colak, I., Sagiroglu, S., Yesilbudak, M.: Data mining and wind power prediction: a literature review. Renew. Energy **42**, 241–247 (2012)
6. Gupta, Y.: Selection of important features and predicting wine quality using machine learning techniques. Procedia Comput. Sci. **125**, 305–312 (2018)
7. Wu, Y.K., Hong, J.S.: A literature review of wind forecasting technology in the world. In: Power Tech. Lausanne, pp. 504–509. IEEE (2007)
8. Wanga, H., Leia, Z., Zhangb, X., Zhouc, B., Penga, J.: A review of deep learning for renewable energy forecasting. Energy Convers. Manage. **198**, 111799 (2019)
9. Okumus, I., Dinler, A.: Current status of wind energy forecasting and a hybrid method for hourly predictions. Energy Convers. Manage. **123**, 362–371 (2016)
10. Foley, A.M., Leahy, P.G., Marvuglia, A., Mc-Keogh, E.J.: Current methods and advances in forecasting of wind power generation. Renew. Energy **37**, 1–8 (2012)
11. Peng, H., Liu, F., Yang, X.: A hybrid strategy of short term wind power prediction. Renew. Energy **50**, 590–595 (2013)
12. Osorio, G.J., Matias, J.C.O., Catalao, J.P.S.: Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information. Renew. Energy **75**, 301–307 (2015)

Integrated Process Management System of Smart Substation Secondary Side Based on Practical Scheme

Shengyang Lu, Tongwei Yu, Hai Qian, and Liqun Li

Abstract The development of smart substations is the future trend of power systems. This paper proposes a scheme for building a whole process management system for the secondary system of the intelligent substation. In the secondary system management system, a specialized model of four roles management of individuals, engineering, equipment, and materials in the building and administrative intelligent substation was established, and a specialized process of the secondary system management system was established. The key technologies such as graphic display technology and model file design consistency check are analyzed. Then, we created a pilot project in the new 220 kV Gaohua smart substation in Shenyang. The management and control of the intelligent substation construction process, intelligent substation equipment, and configuration files are realized. We can clarify the responsibilities of each department to ensure the consistency of the work content of each link and the design of the model documents.

Keywords Integrated process · Management consistency · Intelligent substation · Management building

1 Introduction

(1) In recent years, smart substations have played a pioneering role in the development of smart grid. Among them, digital information is the core of intelligent substation [1]. Data are measured and analyzed through real-time collection and

S. Lu (✉) · T. Yu
Electric Power Research Institute, State Grid Liaoning Electric Power Supply Co. Ltd, Shenyang 110003, China
e-mail: 1025751326@qq.com

H. Qian
State Grid Liaoning Electric Power Supply Co, Ltd, Shenyang 110004, China

L. Li
Anshan Electric Power Supply Branch, State Grid Liaoning Electric Power Supply Co. Ltd, Aanshan 114000, China

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_40

transmission. The combination unit of intelligent substation is the most fundamental difference and innovation compared with ordinary substations. With its high degree of automation and high degree of uniformity and digital model, the second system realizes network mode. The corresponding relationship in this system, the secondary system have stand for main problems. Control systems based on secondary loop equipment and managed objects are not suitable for intelligent substations [2, 3].

(2) Most importantly, the design of secondary equipment for smart substations focuses on SCD files and related configurations [4], which need to be standardized. However, due to the immaturity of the substation intelligent secondary equipment, the actual operation and maintenance process are caused.

(3) At present, the lack of coordination of the intelligent substation system management system has led to the phenomenon of simultaneous design during the commissioning process in a series of pilot constructions [5].

This paper establishes the whole process control system and model profile of substation intelligent function through management. The platform effectively controls the substation intelligent secondary system and implements key control points for some key links in the process [6].

2 Establishment of Intelligent Substation Control System Model

2.1 Modeling of Control System

The secondary system covers the management of the intelligent substation to cover the intelligent substation system, and an object-oriented management system was adopted [7, 8]. The object management of intelligent substation, according to the different classification of its business model, it can be summarized as follows [9].

1. Personnel of working organization: divided into the construction department and substation management department, both departments have main responsibility concerning the employee with the organization [10].
2. Construction: Construction and renovation of intelligent substations. The "Technical Regulations for Intelligent Substation Engineering" has been strictly divided into various construction situations and other work into processes and tasks organized in different stages.
3. Information: The substation includes design materials, whole process documents, and intelligent substation configuration files for new construction, reconstruction, and expansion projects.
4. Installation: Control and management systems have clarified the equipment involved in secondary equipment [11]. The equipment is managed in a unified manner. The type of intelligent substation equipment is shown in Fig. 1.

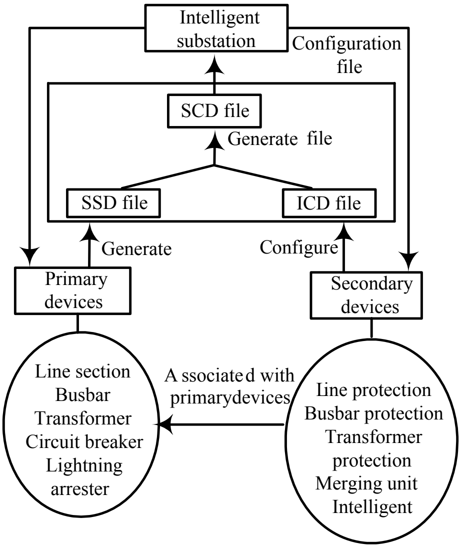


Fig. 1 Intelligent substation equipment type and configuration files

2.2 *Control System Business Process Establishment*

In the intelligent substation business model, because of each role has different functions, each role of the management device is set to a different level, and each role can read information of the device, share information, and set permissions separately [12]. At various stages of the project, all the different departments in the organization must carry out their responsibilities according to their specific duties in order to protect the smart substations system in the correct operation of information and equipment, etc., [13] (Fig. 2).

A logical node can be set up to send the sample data one to one. The matrix expression is shown in Formula 1 [14].

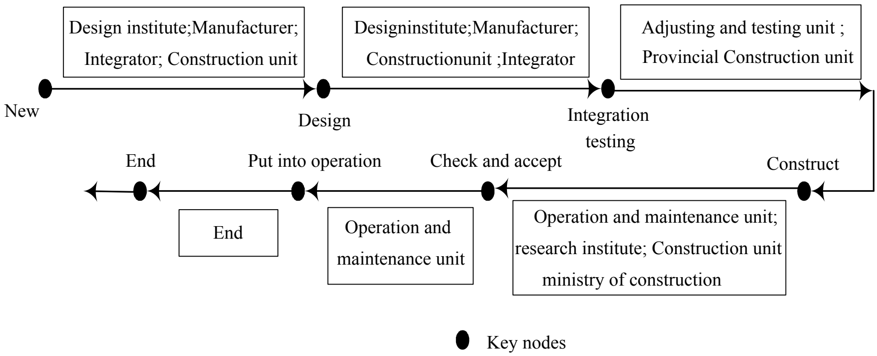


Fig. 2 Management system audit process

$$p_{m\times n}=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{bmatrix} \quad (1)$$

m is the number of logical nodes in the smart substation that has the function of sending and receiving sampled messages, and *n* is the number of all data fields in all messages [15].

3 Research on Management System and Its Control System

3.1 Object Analysis

The intelligent substation not only contains the management modules and control modules of the secondary system in various stages, but also some design and configuration information. It follows these requirements:

1. Upload file request first. Otherwise, no further work can be done.
2. In this process, only the role responsible for uploading files has the right to upload and update.
3. The file should be updated with the associated file.
4. Changing the file version will be logged.

3.2 Imaging Technology

The SCD file describes the instance configuration and communication parameters of all IEDs, and the communication configuration between IEDs and the primary system structure of substation is completed by the system integrator.

Analysis of SCD files divides IED files into different forms according to different functions; that is, the IED function categories can be filtered.

1. IED line

Integrated IED is simple to install and reliable to operate in practical application, which reduces construction cost, saves installation space, and improves the stability of monitoring system.

2. SV transceiver data set

The SV transceiver information and configuration information version are introduced in depth (Fig. 3).

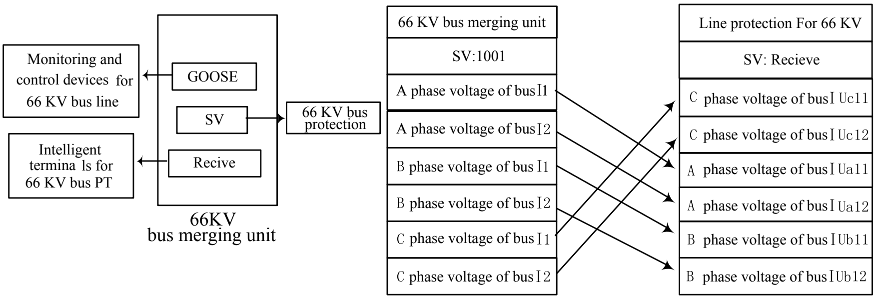


Fig. 3 Linked diagram IED and virtual terminal

3. Data set sent by Goose

Goose sends a wide range of information, including system ip and other information.

3.3 Check the Modeling of the Configuration File

Intelligent substation configuration files are checked uniformly. The contents and methods of inspection are as follows:

Code corresponding to the CID file and the SCD file needs to match the mutual check code. The manufacturer can upload the CID file only through the system's CRC code detection. If the CRC code of the uploaded cid file matches the CRC code of the SCD file, the check succeeds; if it does not match, the operator needs to confirm, as shown in Fig. 4, the smart substation configuration file consistency check diagram.

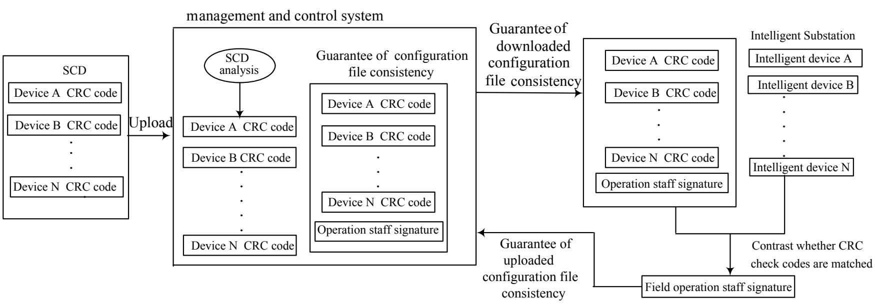


Fig. 4 Smart substation configuration file consistency check diagram

4 Engineering Applications

In December 2015, the Liaoning Provincial Dispatching Center organized the application of "Shenyang 220 kV High Intelligent Substation Construction Process," which involved Dalian Power Company and Liaoning Electric Power Design Institute. The Liaoning Electric Power Company started the construction of intelligent substation and appointed the contractor. The process has entered the early design phase. The MOC organizes various contractors to carry out engineering design, commissioning, acceptance, and commissioning. Within one month after the application was implemented, the operating unit reviewed and examined the relevant links of the secondary system electronic completion drawing.

The successful implementation of the application demonstrates the utility of the control system. The process displays all complete information and ensures system consistency. The construction process is concise and clear, effectively ensuring the normal operation of the substation.

5 Conclusion

This paper proposes a full process controller for substation intelligent secondary system, which fundamentally control functions of the secondary equipment configuration source. Therefore, in the future practical application, the advantages of intelligent substation standardization and networking should be fully utilized to improve effects and clear. As the only centralized source of the system, the primary goal of the management system is the specific work and scenarios of each project. The two aspects complement each other to ensure the stable and accurate operation of the entire substation.

References

1. Yu, L., Wu, H., Huang, J.: Development and application on virtual secondary circuit automatic testing system of smart substation. Inner Mongolia Electr. Power **33**(5), 53–57 (2015). (in Chinese)
2. Sun, P., Zhang, D., Wang, F.: Debugging and operation technology for intelligent substation. China Electr. Power Press, Beijing (2014). (in Chinese)
3. Wang, T., Pan, L.: Debugging technology of secondary system for intelligent substation. China Electr. Power Press, Beijing (2013). (in Chinese)
4. Singh, S., Chana, I.: QoS based machine learning algorithms for clustering of CloudWorkloads: a review. Int. J. Cloud Comput. Super Comput. **3**(1), 13–24 (2016)
5. Kumar, T.: Advanced approach and management in cloud trust. Int. J. Urban Des. Ubiquitous Comput. **4**(2), 7–12 (2016)
6. Konduri, B.: Cloud security based analysis in hybrid system. Int. J. Private Cloud Comput. Environ. Manage. **3**(1), 11–16 (2016)

7. Cho, Y., Jung, G., Cha, J.: Proposal and evaluation of webpage menu interface for people with low vision. Int. J. Adv. Sci. Technol. **119**, 33–52 (2018)
8. Kong, L.B., Qin, H.W., Zhang, X., Fan, Q.Y.: The development of Chemico-mechanical practices in polymer material machining by cutting. Int. J. Control Autom. **8**(3), 37–44 (2015)
9. Xie, X.L., Wang, A.X., Schroeder, H.: Study of the grid resource neural network reservation technology based on SLA. Int. J. Grid Distrib. Comput. **10**(1), 21–30 (2017)
10. Ou, Y.Y., Xie, J.B., Ling, J.: An improved evaluation index system for the host information security evaluation system. Int. J. Secur. Appl. **9**(3), 49–62 (2015)
11. Chen, B., Lu, G., Fang, H, Z., Ai, L., Fan, H.Z.: IGBT neural network prediction method of radar transmitter based on Levenberg-Marquard optimization. Int. J. Future Gener. Commun. Network. **9**(9), 1–14 (2016)
12. Ji, J.F., Wang, L.H., Yuan, Y.B., Yang, Y.F.: Analysis of electromagnetic immunity experiment for field installation of intelligent equipment in substation. Int. J. Signal Process. Image Process. Pattern Recogn. **9**(1), 21–32 (2016)
13. Geng, W.L., Jing, T.: Construction and simulation of competitiveness evaluating model for E-commerce enterprise based on DEA. Int. J. U E Serv. Sci. Technol. **8**(2), 61–70 (2015)
14. Kim, M.: A study on a comparison of diagnostic domain between SNOMED CT and Korea standard terminology of medicine. Int. J. Database Theory Appl. **9**(11), 49–60 (2015)
15. Gehlot, A., Singh, R., Mishra, R.J., Kumar, A., Choudhury, S.: IoT and Zigbee based street light monitoring system with LabVIEW. Int. J. Sens Appl. Control Syst. **4**(2), 1–8 (2016)

Machine Learning and IoT Applications

Lightweight Access Control Algorithm for Internet of Things

Aravind Sreekumar, Ruthvik Vegunta, and S. Deepthi

Abstract While Internet of Things is an upcoming technological breakthrough and a fast-developing field of technology, the access control to restricted resources devices still remains a concern. The conventional solutions require a lot of computational and bandwidth capabilities, which are not available in constrained devices. Also, there is very limited interoperability against standard communication protocols for the Internet of Things. These issues are being solved with the help of OAuth-IoT, an access control framework. Though the OAuth-IoT solves all the major issues with computational and bandwidth requirements, there are still a few unsolved issues. This contribution presents a flexible and simple method using open-standards to implement access control and authentication in Internet of Things.

Keywords Internet of Things · Access control · Authentication

1 Introduction

The Internet of Things (IoT) is a fast growing and emerging technology that has implications in every field and hence emerging as a promising communication paradigm allowing different devices to communicate with each other seamlessly. The success or failure of any revolutionary paradigm is determined by how secure is the technique and whether the system is able to keep the confidentiality intact [1]. With the introduction of the use, internet in every possible device brings in the concern for security of the data. Though there are several methods and measures that are in use

A. Sreekumar · R. Vegunta · S. Deepthi (✉)
Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Manipal 576104, Karnataka, India
e-mail: deepthi.s@manipal.edu

A. Sreekumar
e-mail: aravind.sreekumar@learner.manipal.edu

R. Vegunta
e-mail: vegunta.ruthvik@learner.manipal.edu

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_42

to keep the data safe, the authentication and authorization services of devices are still a discussed issue.

Access control is the most important security aspect in IoT. In IoT, there will be interaction between multiple devices. Using access control techniques, system can decide who can access what, when and in which condition [2].

Currently, the OAuth 2.0 [3] solves the problem that is faced in authentication and authorization services. OAuth 2.0 is a token-based access control framework. It helps users by logging in automatically with the services required by the users without the need to login for accessing the service each time the user requires to use the services.

Though OAuth 2.0 solves the issue of authentication and authorization services to some extent, since we are talking about Internet of Things here as the device used, the resource available is a concern. IoT devices have very restricted resources and power that can be used up, which means the software that is used must be very lightweight and less resource-intensive. OAuth 2.0 being a very resource-intensive framework cannot be used in all the IoT devices which have serious restrictions in terms of the resources that are being used.

In line with these premises, this paper presents an alternative, less resource-intensive approach to OAuth 2.0 which can be used in any low-powered, resource-restricted device. This approach uses open-source approaches, making it a cost-effective method too. This methodology uses access control lists to give users access to data that is available. Each user is granted permission by an administrator who reviews the user first and then approving his/her request to access the required data. Here, the requirement of an authorization server is eliminated as an administrator manually approves the request, thus making it more secure. The data transfers here take place using REST APIs and the entire backend is handled by PHP.

The rest of the paper is organized as follows. Literature review is discussed in Sect. 2. Section 3 describes the algorithms used in the proposed access control protocol. Sect. 4 discusses the results that are achieved. Finally, Sect. 6 draws conclusions and future works.

2 Literature Review

Many access control models are proposed in the literatures to address the authentication issues in IoT.

Authors in [4] give an abstract about access control matrices. In role-based access control (RBAC), each user is associated with roles and roles define permissions. Each user is assigned the access rights based on the roles, the user is associated with. Each resource is also associated with permission and access control decision is made based on user permissions and resource permissions.

Attribute-based access control is described by E. Yuan in “attributed-based access control (ABAC) for Web services” [5]. In this method, access control matrix is used. Access policies are associated with each user based on their attributes, and each resource has attributes that describe the conditions that must be met before access is granted. Access control decision is made based on equivalent user's attributes against demands associated with the resource. In the IoT context, updating the system is more difficult to implement.

In IACAC [6], each device established communication is verified by its capability access. Each device has capability. If a device needs to access a resource, IACAC model verifies whether device has capability to communicate with the resource. Accordingly, the request is either accepted or rejected.

Access control as described by Sergio Gusmeroli in "a capability-based security approach to manage access control in the Internet of Things [7]" lets only authorized users to access a resource, such as a file, IoT device, sensor or URL. All modern operating systems limit access to the file system based on the user. In the IoT context, access control is required to make certain that only trusted parties can update device software, access sensor data or command the actuators to perform an operation. In this access control method, each user is assigned with different capabilities. Based on capabilities, user can access the resources.

Access control solutions based on usage controlled proposed by authors in [8–10] support mutability of attribute that is if suppose a user is currently accessing an attribute and access levels of an attribute are changed, if this change in access level is higher than the allowed access privilege of the user, then the access grants are revoked and user is not anymore having any access rights to the attribute [11].

Organizational-based access control [12] brings the new dimension "organization" into consideration. This policy provides security policies for the organizations.

3 Methodology

The Raspberry Pi is a series of small single-board computers. It features a Broadcom system on a chip (SoC) with an integrated ARM compatible central processing unit (CPU) and on-chip graphics processing unit. Raspberry Pi here in our approach uses the below-mentioned sensors to detect temperature and humidity in the real-time environment, the sensor used in this approach is DHT-11 temperature and humidity sensor. After detecting the temperature and humidity, Raspberry Pi uses RSA algorithm [13] to encrypt that data before sending it to a cloud platform.

For accessing live data from the IoT, the following algorithm was used.

Step 1: User requests for live data.
Step 2: Mail is sent to the administrator which contains links to grant or decline access rights to the requested user.
Step 3: If administrator grants access go to Step 4. Else, go to Step 6.
Step 4: Access level of the user is incremented to grant access to the cloud data.
Step 5: Now, user accesses the cloud to pull the most recent data which is pushed by the IoT device.
Step 6: A message is displayed saying that the user is declined access.

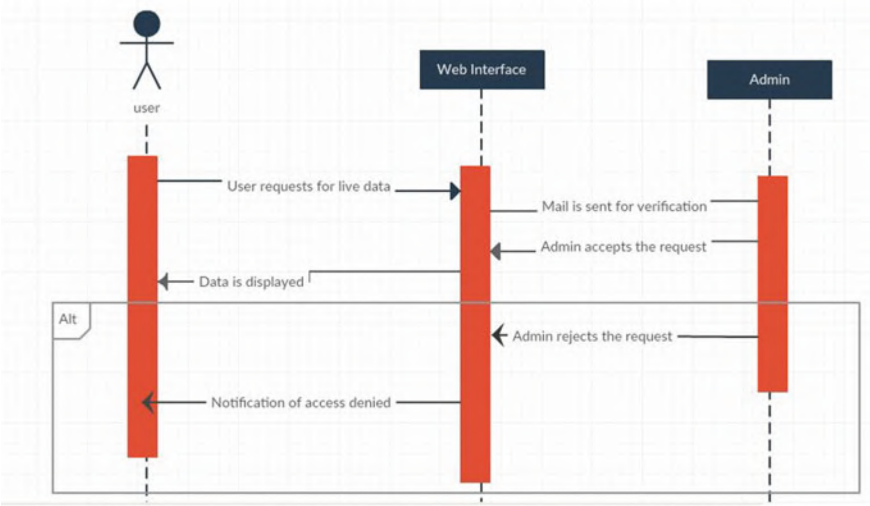


Fig. 1 Access control for live data

The diagrammatic representation of this method is shown in Fig. 1.

For accessing previous data from the IoT, the following algorithm was used.

Step 1: User requests for previous data.
Step 2: Users email Id is validated. The users email ID is fetched and checked with the database if the user's email's domain is present in the database for providing higher access rights.
Step 3: If email Id is from an accepted domain go to Step 4, else go to Step 5.
Step 4: Previous data is displayed.
Step 5: Notification is displayed saying the user does not have enough credentials to access the data.

The diagrammatic representation of this method is shown in Fig. 2.

Confidentiality of messages is protected using an encryption algorithm. This approach uses RSA algorithm to encrypt the messages to provide information security.

Figure 3 depicts the overall design of the approach.

4 Results and Discussion

4.1 *Experimental Setup*

Following entities are used in the experiment:

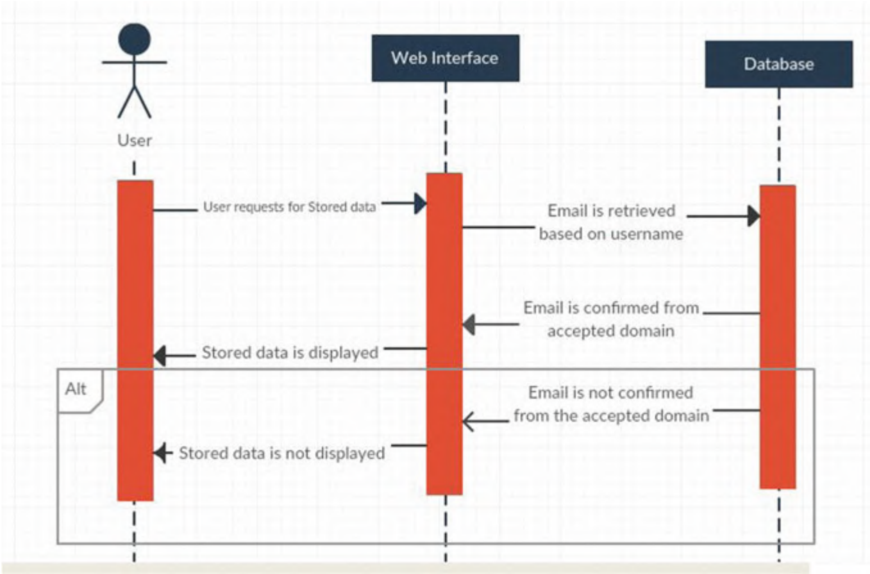


Fig. 2 Domain-based access control

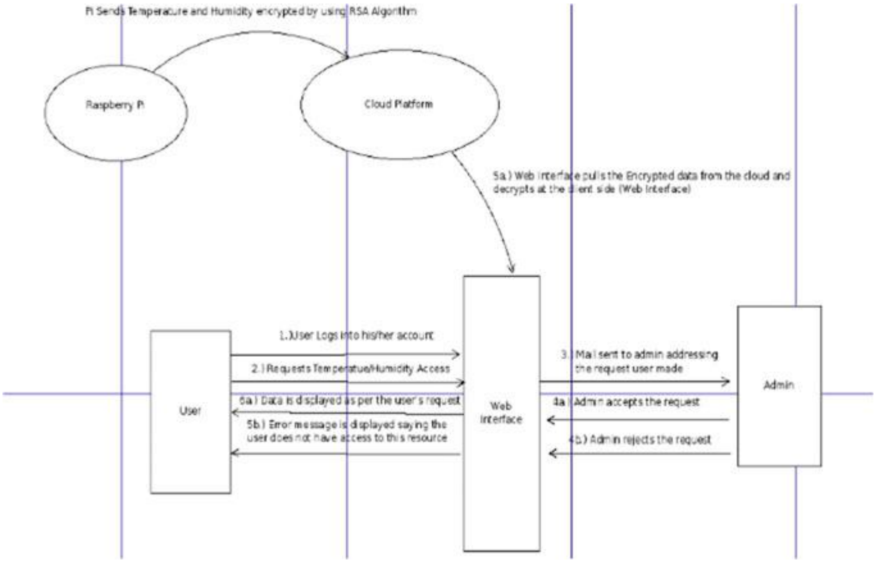


Fig. 3 Design of the approach

- IoT device: The Raspberry Pi has to be set up by installing the operating system, Raspbian (A Debian-based operating system made for Raspberry Pi). The software used and installed in the Pi for the approach was Python 3. The Raspberry Pi is connected to an external display for its initial setup which involves setting up of the operating system and setup of Wi-Fi.
- Sensor device: The DHT 11 sensor used to sense temperature and humidity has three pins which are: a 5 V power, data output and ground from left to right, respectively. These pins are connected to the general purpose I/O (GPIO) pins of the Raspberry Pi in above-mentioned manner. IoT device sends captured data to cloud.
- Client: An application which allows user to login and send requests to access data.
- Admin: An application which authorizes the client and grants the access rights accordingly.
- FTP: Windows Secure Copy (WinSCP) is used to remotely access the device; the python code is transferred to the device memory using WinSCP and run through the terminal using the Linux commands. When the code is run, the device sends data to Thingspeak IoT platform; the code accesses the channel in Thingspeak using a token key and an API for pushing data provided by the platform.
- Cloud: The Thingspeak platform provides a graph view of all the parameters send to it. The data can be read from the platform using a set of access token and API.

4.2 Results

The Raspberry Pi keeps sending data to the cloud platform by capturing data from the DHT11 temperature/humidity sensor. The Raspberry Pi keeps sending data at a predefined interval set by the administrator. The data sent is encrypted using a RSA algorithm. The data sent is time of capture of data, ambient temperature, encrypted temperature value, ambient humidity, encrypted humidity and the API used for sending data. The Web interface is used to access the data that is being captured by the IoT and which is stored in the cloud. A user must first signup with the interface with his details to be able to access the data that is being stored in the cloud platform. Once the user has signed up with the page, then he/she shall login with the provided username and password combination. Once logged, the user will be able to view the data that his access right provides him with. Once the data is viewed, the user may sign out from the page.

Once a user logs in, he/she is asked to choose the access right required by him/her. Once the choice is made, a mail is sent to the administrator wherein the administrator chooses to grant access or decline access by clicking on the appropriate link. Once the administrator grants the access, the access level of the user is incremented in the database and then the temperature is displayed. If access is not given to a user, then he/she receives an error message. If a user has an accepted domain from a predefined list, then the user can view the previous temperature by clicking on the

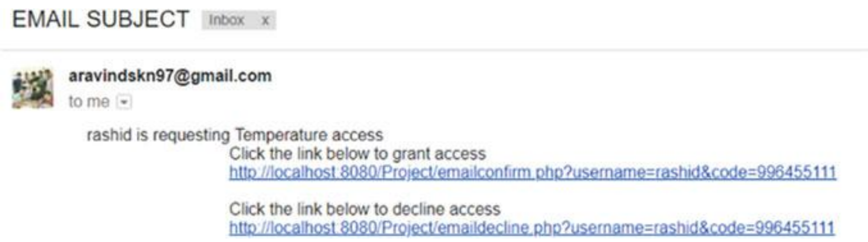


Fig. 4 Access request mail

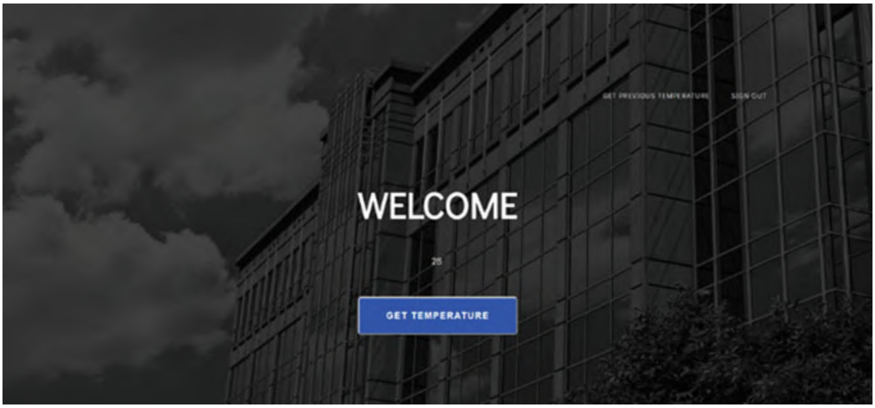


Fig. 5 Displaying temperature

GET PREVIOUS TEMPERATURE link, where the users domain name is checked first, if approved he can view the data. If access is denied, then the user is given an error message. Figs. 4, 5 and 6 show snapshots of the access control which is implemented.

Time taken by IoT device to fetch the data, encrypt and send to cloud is captured as shown in Fig. 7. It shows that time taken is nominal.

5 Conclusion

The main objective of this approach was to provide access control to the devices which have resource constraints. The access control which is being implemented in this approach uses less resources as when compared with other access control frameworks which are used on both Internet of Things devices and with devices which do not have any resource constraints.

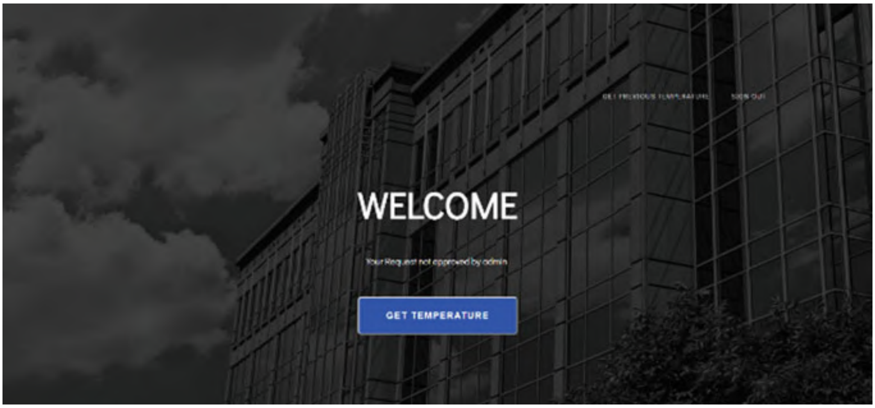


Fig. 6 Access denied for use

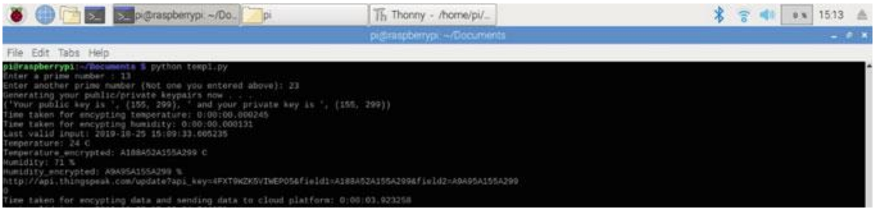


Fig. 7 Time taken by IoT device

The approach also gives the implementation of security algorithms that are used to send and receive data in an encrypted format, thus keeping the data secure from any kind of attack.

This implementation can be further improved in the future by the addition of many new features to it. A few noteworthy ones are mentioned below. The Web interface can be made into a mobile application, thus making it more portable and the user can access the data from anywhere with ease. The IoT can have more sensors added to it, thus giving the user more information of the surrounding environment. For example, a GPS sensor could be added, thus giving the location of from where the temperature or humidity is being collected.

References

1. Orange, the future of digital trust a European study on the nature of consumer trust and personal data, Technical Report (2014)
2. Ouaddah, A., Mousannif, H., Elkalam, A.A., Ouahman, A.A.: Access control in The Internet of Things: Big challenges and new opportunities. Comput. Netw. **112**. https://doi.org/10.1016/j.comnet.2016.11.007

3. Sciancalepore, S., Piro, G., Caldarola, D., Boggia, G., Bianchi, G.: OAuth-IoT: an access control framework for the Internet of Things based on open standards. In: IEEE Symposium on Computers and Communications, Italy (2017)
4. Kulkarni, D., Tripathi, A.: Context-aware role-based access control in pervasive computing systems. In: Proceedings of the 13th ACM symposium on Access control models and technologies, Colorado State University, USA, pp. 113–122, June 2008
5. Yuan, E., Tong, J.: Attributed based access control (ABAC) for Web services. In: Proceedings IEEE International Conference on Web Service (2005)
6. Mahalle, P.: Identity authentication and capability based access control (IACAC) for the internet of things. J. Cyber Secur. Mobility **1**, 309–348 (2013)
7. Gusmeroli, S., Piccione, S., Rotondi, D.: A capability-based security approach to manage access control in the Internet of Things. Math. Comput. Model. **58**, 1189–1205 (2013)
8. Zhang, X., Parisi-Presicce, F., Sandhu, R., Park, J.: Formal model and policy specification of usage control. ACM Trans. Inf. Syst. Secur. **8**(4), 351–387 (2005)
9. Park, J., Sandhu, R.: Towards usage control models: beyond traditional access control. In: Proceedings of the Seventh ACM Symposium on Access Control Models and Technologies, p. 57. ACM Press, New York, New York, USA (2002)
10. Park, J.: Usage control: a unified framework for next generation access control (2003)
11. Lazouski, A., Martinelli, F., Mori, P.: Usage control in computer security: a survey. Comput. Sci. Rev. **4**(2), 81–99 (2010)
12. Kalam, A., Baida, R., Balbiani, P., Benferhat, S., Cuppens, F., Deswarte, Y., Miege, A., Saurel, C., Trouessin, G.: Organization based access control. In: Proceedings of POLICY 2003. IEEE 4th International Workshop on Policies for Distributed System Networks, pp. 120–131 (2003)
13. Zhou, X., Tang, X.: Research and implementation of RSA algorithm for encryption and decryption. In: Proceedings of 2011 6th International Forum on Strategic Technology, Harbin, Heilongjiang, 2011, pp. 1118–1121

Named-Entity Recognition for Legal Documents

Harsh Vardhan, Nitish Surana, and B. K. Tripathy

Abstract The law has language at its core, so it is not surprising that software operating on natural language has played a role in certain areas of the legal industry for a long time. The last few years have seen a significant upsurge of interest in this area, including an increasing number of start-ups applying deep learning techniques in the context of specific legal applications. In this paper, we present a simple yet powerful method that is applied to legal documents from different legal bodies to correctly recognize a numerous entity to find relevant information for some specific matter at hand. To the best of our knowledge, no attempt has been made in this direction so far and as such our work opens a new direction of research, which will be very much useful for the society.

Keywords Named-entity recognition · Deep learning · Natural language processing · Legal tech

1 Introduction

Language and law are meticulously connected [1]. It is fundamental to create techniques for legal natural language processing (NLP) so as to comprehend legal language, to create tools supporting the misuse of legal sources for law imposition, as well as to empower straightforward, universal, and interoperable legal frameworks over the Web. LegalTech refers to technologies from computer science that is applied

H. Vardhan · N. Surana (✉)
FindMind Analytics Pvt. Ltd., Vellore, India
e-mail: nitishsurana@findmind.in

H. Vardhan
e-mail: harshvardhan@findmind.in

B. K. Tripathy
SITE, Vellore Institute of Technology, Vellore, India
e-mail: tripathybk@vit.ac.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_43

to a range of areas related to legal practice and materials [2]. It refers to the utilization of technology and software to provide legal assistance. It has a wide range of application areas to help law firms and organizations with daily activities related to document support, legal proceedings, and more generally all aspects related to the dematerialization of legal services from text and paper to digital form [3].

Since there are over hundreds and thousands of legal cases all across the country happening every minute, it becomes really hard for any lawyer if he wants to find details about any case which he can use in his arsenal as an advantage. Presently, paralegals are commonly appointed to give a summary of any legal document as well as the important aspects of the document that might be useful. These are generally handwritten documents by them and became really hard to manage as the number of documents increase. Hence, it becomes physically impossible to keep providing these documents with the same efficiency in quality.

Over the last decade, the use of the regular expression to ship out these attributes and entities has become widely popular. Various organizations have started implementing this and are trying to find out the patterns that are generally followed in the legal documents so as to get the required entities as soon as possible. But there exists an anomaly in this being that this pattern is never consistent and can change at any time, greatly increasing the risk factor in following this approach.

Various attempts were also made to understand the importance of classifying texts [4] in legal documents through the use of machine learning. While some preferred using a semi-supervised convolutional neural network [5] in which unlabeled data is combined with a small set of labeled data to create better models for classifications, others preferred focusing more on keyword-based extraction techniques using the sequence to sequence modeling [6]. Multi-label text classification [7] through the use of annotations and word vectors [8] were also tried by different groups of people. LegalTech advancements also encouraged a wide range of individuals to understand the significance of text summarization too in the later years [9].

We have created a model that understands the legal documents and recognizes the important entities that are present therein. These entities are then returned back to the user. The main advantage of using this model is that we do not have to worry about any patterns that are present in the legal documents as the model does not look for any pattern in the documents but rather tries to understand the data and gives out the correct output based on it.

We have used spaCy [10] to create our model. SpaCy is an open-source Python library that is available online. It is structured mainly for production usage, as it helps users create applications that can be used to process and understand enormous volumes of content. It is used to perform advanced NLP tasks in Python. It also provides an effective statistical framework for NER in Python, which can relegate labels to a group of contiguous tokens. It uses a default model that can perceive a wide scope of named or numerical elements, which incorporate person, organization, language, event, etc. Aside from these default elements, spaCy additionally gives us the freedom to add subjective classes to the NER model, by training the model to update it with newer trained examples. The few other features of spaCy are tokenization, part-of-speech (POS) tagging, lemmatization, rule-based matching, text classification, etc.

In Sect. 2 of this paper, we talk about how the model was built and its respective implementation. In Sect. 3, we show the results that we were able to obtain using this model. In Sect. 4, we conclude our findings, and lastly, in Sect. 5, we present some future scopes for the model and the directions in which it can be extended.

2 Methodology

We have created a model that can correctly recognize specific entities like "court name," "date of judgment," "petitioner," "respondent," "name of the judge," and "acts" from any legal document and present the output to the user. Knowing these entities before actually viewing the entire document can come in handy for the lawyers, can greatly reduce the number of documents that he needs to go through, and can utilize his time effectively only viewing the important documents. Hence, this gives them more time to prepare before the actual hearing of the case. We perform this using the named-entity recognition (NER). It is a subprocess of data extraction that searches out and arranges determined entities in a body or assortments of writings. The entire methodology followed for recognizing various entities present in the legal documents is represented below (Fig. 1).

The above flowchart shows the step-by-step procedure that was followed to create the NER model. The methodology can be split into two sections, data preparation and model architecture. Data preparation talks about the different steps that need to be taken before the NER model can be made, while the model architecture talks about the main algorithm through which the entities are recognized from the testing data.

2.1 Data Preparation

We get our majority of the training data from various different sites that provide legal documents. The basic preprocessing of the data is also done before actually feeding it to the model. There are various steps that need to be performed before we can actually create the model based on our training data. The first step would involve getting the legal documents present in the training data to text format so that

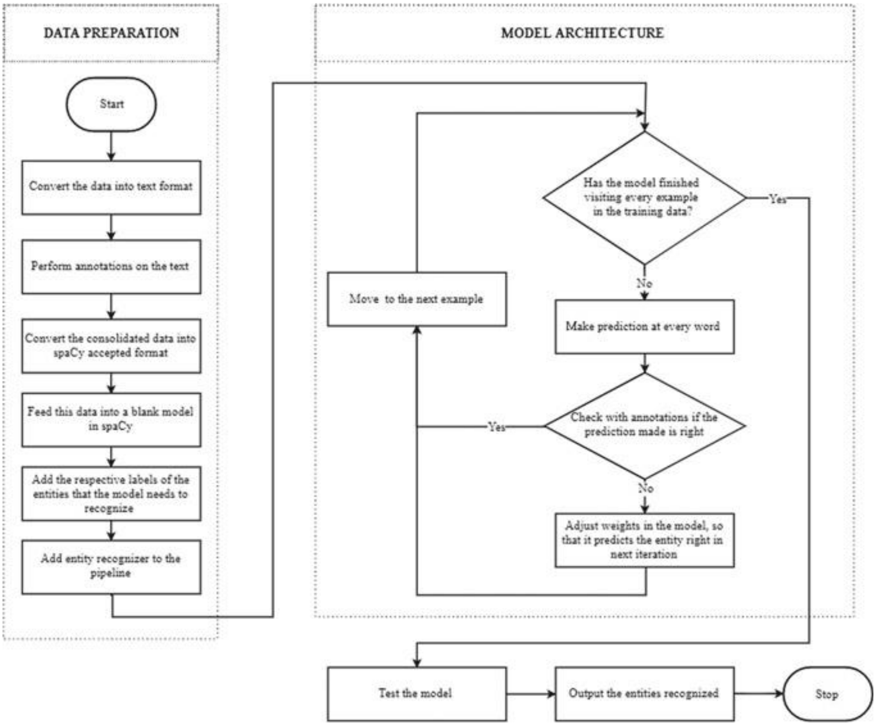


Fig. 1 Step-by-step procedure for creating the NER model

we can perform annotations on them. The annotations refer to all the entities that we want the model to recognize from the documents present in the testing data. The annotations should have the starting and the ending location of the entity, along with the respective entity name. This is one of the most important steps as, if the entities are wrongly tagged in the training data, the model will not be able to recognize the entities from the testing data. Respective annotations are made for all the entities that we want the model to recognize for all the documents present in the training data. An example of the annotations that can be performed for one legal document is shown below:

```
{
      "content":"Bombay High Court\nCommissioner Of In-
      come-Tax, ... vs Carona Sahu Co. Ltd. on 21 Octo-
      ber, 1983\nEquivalent citations: (1984)",
      "metadata": {

      },
      "annotations": [
      {
            "start":0,
            "end":17,
            "tag": "CourtName"
      },
      {
            "start":250,
            "end":262,
            "tag": "NameOfJudge"
      },
      {
            "start":18,
            "end":73,
            "tag": "RespondentvsPetitioner"
      },
      {
            "start":77,
            "end":93,
            "tag": "DateOfJudgement"
      }]

}
```


Since we are using spaCy to create our NER model, we need to convert our data into the spaCy accepted format. We have to remove any extra keywords like, “start,” “end,” and “tag” keywords that may be present in the annotations. SpaCy only requires that we provide the location for the start and the end, along with the respective entity name. We also need to provide data in a Python list/tuple format. We must remove any other extra keywords that may be present in the annotations like “content,” “classification,” “metadata,” etc. as spaCy does not allow them. After changing the data into spacy accepted standard, a blank model is created. Since we are not using any pretrained model, entity recognizer is added in the NLP pipeline. After this, the respective entity labels that we have used while performing the annotations on the training data are added to the model. An example of a document that is converted into spaCy format is shown below.

```
DATA = [
      (
            'Bombay High Court\nBasantibai Fakirchand
            Khetan And ... vs
            State Of Maharashtra And Anr. on 8 Novem-
            ber, 1983\nEquivalent citations: AIR 1984
            Bom 366',

            {'entities': [(0, 17, 'CourtName'), (18,
            87, 'RespondentvsPetitioner')]}
      ),
      (
            'Bombay High Court\nCommissioner Of Income-
            Tax, ... vs Carona Sahu Co. Ltd. on 21 Oc-
            tober, 1983\nEquivalent citations: (1984)
            38 CTR Bom 219, 1984 ',

            {'entities': [(0, 17, 'CourtName'), (18,
            73, 'RespondentvsPetitioner')]}
      )
]
```

2.2 Model Architecture

After data preparation, we can now build the NER model. We are using a blank model with the ID of the English language class. This NER model uses four convolutional neural network (CNN) layers with residual connections and the max out activation function to perform the "encode" step. It first converts all the words present in the document into its feature vectors. These feature vectors have four attributes—norm, prefix, suffix, and shape. The norm represents the normalized form of the string, prefix stores the prefix of the string and has a maximum capacity of three, suffix stores the suffix of the string and also has the maximum capacity of three, and the shape uses a fuzzy zoomed out view which is good at recognizing unknown words based on the patterns. The CNN uses a window-like architecture for the encoding. A multi-layer perceptron (MLP) is also used to map the feature vectors to 128 dimensions. These are then used to construct CNN layers. All the four CNN layers are stacked together. In this way, a receptive field of four tokens is used to “encode” in the NER model. Through this field, a single model that works generally well on detecting entities is built.

The entity recognizer present in the model will be trained to recognize the different entities present in the testing data. We then loop over the examples in the training data and call an update function provided in the NLP package, which ventures through the words of the input. At each word, it makes a prediction. It then checks the respective annotation for that word, to see whether it was correct. If it does not predict the right outcome, it changes its weights so that the correct action will score higher next time. In this way, it checks its respective results with each loop iteration and changes weight accordingly so that it can recognize correctly in the next iteration. This way the model is built and finally tested on testing data to check how well the model performs, respectively.

3 Results and Discussion

We have split the total dataset into an 80:20 ratio, i.e., 80% of the data will be used as training data for the model and the remaining 20% of the data will be used as testing data for the model. The model was made to run for 10 epochs before the output was generated. After 10 epochs, the model was able to recognize various entities that are present in the document from the testing data.

The metrics that we used for evaluating the model are precision, recall, and *F*-score. These were generated by using the scorer method from the spaCy library. On calling the scorer script, it returns six values. They are:

- uas (Unlabeled Attachment Score)—It is the proportion of tokens whose head has been correctly assigned. It is used to evaluate dependency parsing.
- las (Labeled Attachment Score)—It is the proportion of tokens whose head has been correctly assigned with the right dependency label (subject, object, etc.). It is also used to evaluate dependency parsing.
- ents_p (Precision)—It is the percentage of named entities found by the model that is correct.

- $$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

  where *TP*—True positive (the model correctly predicts the respective class), *FP*—False positive (the model incorrectly predicts the respective class).
- ents_r (Recall)—It is the percentage of named entities present in the corpus that is found by the model.
- $$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$
- ents_f (*F*-score)—The *F*-score is a measure of a test's accuracy. The *F*-score is defined as the weighted harmonic mean of the test's precision and recall.
- $$F\text{−score} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \tag{3}$$
- tags_acc—It is the part-of-speech (POS) tagging accuracy.
- token_acc—It is used to calculate the precision for token segmentation.

Since our task was to recognize named entities, we need to check for ents_p, ents_r and ents_f scores only. On using scorer script for our model, the output generated is shown below:

```
{'ents_p': 72.88135593220339, 'ents_r': 50.0,
'ents_f':59.31034482758621}
```

From this result, we observe that our model has 72.88% precision, 50% recall and F-score of 59.31%. A sample test document from the test set that is not labeled is shown below:

Bombay High Court

Kailash Laxman Khamkar vs State Of Maharashtra on 9 February, 2010 Bench: S.C. Dharmadhikari

IN THE HIGH COURT OF JUDICATURE AT BOMBAY CRIMINAL APPELLATE JURISDICTION

CRIMINAL APPEAL NO.159 OF 2004

Kailash Laxman Khamkar.. Appellant Versus

State of Maharashtra .. Respondent

Mr.B.G.Tangsali for appellant Ms.A.T.Jhaveri, APP for State.

ig CORAM : S.C.DHARMADHIKARI, J.

DATE : 9th February 2010.

ORAL JUDGEMENT:-

1] The appellant is the original accused in Sessions Case No.282 of 1998, which was on the file of the District and Sessions Court, Thane. The appellant challenges the judgement and order dated 31st December 2002 delivered by the Third Additional Sessions Judge, Thane convicting him for the offence punishable under section 366-A of Indian Penal Code and sentencing him to suffer R.I., for three years and fine of Rs.2000/-. In default, he was to suffer R.I., for two months.

This text is sent to our model and the output generated in the JSON format is shown below:

```
{
        "CourtName": [
              "Bombay High Court"
        ],
        "Petitioner": [
              "Kailash Laxman Khamkar"
        ],
        "Respondent": [
              "State Of Maharashtra"
        ], "DATE": [
              "9 February, 2010"
        ],
              "Name of Judge / Bench": [ " S.C. Dhar-
        madhikari"
        ], "ACTS": [
              "CRIMINAL APPEAL NO.159 OF 2004"
        ]
}
```


4 Conclusion

In this paper, we presented how we can create our own model for NER for legal documents. This can be used in the production environment where we can analyze large volumes of data in a much better manner. We successfully created the NER model using the methodology mentioned in the paper. We then trained this model using the training data, followed by the testing performed on the testing data. The model was able to recognize all the entities present in the documents from the testing data. Even though we have only trained the model for a fairly small number of legal documents, we can make the model even more robust by entering more data and also train it on varied types of legal documents like acts and judgments.

5 Future Work and Scope

Our model only aimed at recognizing a few entities that we believed were important when going through any legal document. We can further refine it by training it for even more entities that are present in the document. We can add our very own custom entities that we wish to find out from the legal document and train the model based on them. The choice of entities is not limited to any particular number and can be further implemented into a large-scale model that can be used to separately differentiate between all the different acts that are present. There is always a scope of improvement in this aspect as we can customize our own model according to our needs to recognize a large number of entities from the document.

References

1. Nazarenko, A., Wyner, A.: Legal NLP introduction. In: Association pour le Traitement Automatique des Langues (2017)
2. Legal technology. https://en.wikipedia.org/wiki/Legal_technology
3. Chalkidis, I., Kampas, D.: Deep learning in law: early adaptation and legal word embeddings trained on large Corpora. Artif. Intell. Law **27**(2), 1–28 (2018)
4. Sulea, O.M., Zampieri, M., Malmasi, S., Vela, M., Dinu, L.P., van Genabith, J.: Exploring the Use of Text Classification in the Legal Domain (2017)
5. Li, P., Zhao, F., Li, Y., Zhu, Z.: Law text classification using semi-supervised convolutional neural networks. In: 2018 Chinese Control and Decision Conference, pp. 309–313. IEEE (2018)
6. Daojian, Z., Guowei, T., Yuan, D., Feng, L., Bing, H., Songxian, X.: Keyphrase extraction for legal questions based on a sequence to sequence model. J. Tsinghua Univ. (Sci. Techn.) **59**(4), 256–261 (2019)
7. Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., Androutsopoulos, I.: Extreme Multi-Label Legal Text Classification: A Case Study in EU Legislation (2019)
8. Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Androutsopoulos, I.: Large-Scale Multi-Label Text Classification on EU Legislation (2019)
9. Kanapala, A., Pal, S., Pamula, R.: Text Summarization from legal documents: a survey. Artif. Intell. Rev. **51**(3), 371–402 (2019)
10. spaCy. https://spaCy.io/usage

Visual Speech Processing and Recognition

U. B. Mahadevaswamy, M. Shashank Rao, S. Vrushab, C. Anagha, and V. Sangameshwar

Abstract Lip reading is the ability to understand what a person is communicating using just the video information. Due to the advent of Internet and computers, it is now possible to remove human intervention from lip reading. Such automation is only feasible because of a couple of developments in the field of computer vision: availability of a large-scale dataset for training and use of neural network models. The applications to this are numerous. From dictating messages to a device in a noisy environment to improving speech recognition in the current technologies, visual speech recognition has proved to be pivotal. In this paper, the lip-reading models are based on deep neural network architectures that capture temporal data which are created for the task of speech recognition.

Keywords CNN · Viseme · ConvLSTM · PyDrive · HOG

1 Introduction

Since the dawn of mankind, we have been born or cursed with disabilities, but till recent technological advancements, we have not been able to alleviate these problems. Throughout the world, there are more than 450 million people who suffer

U. B. Mahadevaswamy · M. Shashank Rao (✉) · S. Vrushab · C. Anagha · V. Sangameshwar
Department of Electronics and Communication Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, India
e-mail: shashankmrao@gmail.com

U. B. Mahadevaswamy
e-mail: mahadevaswamy@sjce.in

S. Vrushab
e-mail: sangamvrushab@gmail.com

C. Anagha
e-mail: anaghacraj98@gmail.com

V. Sangameshwar
e-mail: sangameshvenu@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_44

from hearing loss. With only 30–40% of all spoken sounds visible on the lips or tongue in the English language, it can be extremely problematic for those dependent on lip reading to understand what is being said on a day-to-day basis. Now, we can make the lives of these unfortunate souls and help them experience the world as we do. When we have the technology to help anyone, especially those who deserve it the most, then it is our fundamental duty to do it. Recent advancements in the fields of computer vision, pattern recognition and signal processing help brings us closer to building a robust machine for lip reading. Automating the human ability to lip read, a process referred to as visual speech recognition (VSR), opens a lot of opportunities in redubbing silent films, differentiating multi-talker speech and also increasing the performance of speech recognition technology itself.

2 Summary of Literature Review

In paper [1], HMM and DBN have been used to classify the temporal changes of the appearance of the mouth. Similarly, in [2], HMM-based systems were used to report a recognition rate of 70%. But in [3], it was proposed to use a spatiotemporal descriptor to capture temporal changes in video. In [4], a generative process that analyzes sequences of talking faces and classifies them as visual speech units called visemes was implemented to provide an effective graphical model. Even though many previous researches have been built upon one another, new novel features define the innovations which have advanced the visual speech recognition and processing technology one step at a time, and similarly, every small advancement is learned from and further implemented in this paper.

3 Methodology

As mentioned earlier, a dataset for training the deep learning model was needed. It was obtained from BBC, which consisted of many news reporters saying different words. Each word in the dataset consisted of 1000 videos, and each video consisting of a single news reporter saying the word. 200 words were selected (totaling up to 200 * 1000 = 2,00,000 videos) as the raw dataset for training the model.

A. **Data Preprocessing**

Having obtained the dataset of the various words that will be used for training, it had to be converted into a form that can be used for training. The number of videos in each word was split into the training set and the test set, such that the training set consisted of 1000 videos, the test set consisting of 50 videos and validation set consisting of 50 videos.

Each video in the dataset was modified by extracting only the lip component of the person speaking in the video. All other features in the video (the person's other facial landmarks apart from the lips, the background, etc.) did not contribute to the training of the lip-reading model and were hence cropped off.

In order to extract only the lip movement of the person speaking in the video, OpenCV in Python was used. Before the extraction of the facial features, the video is first converted into a series of frames/images, by selecting one frame for every three frames, so as to get 9 frames for every video. This ensures that there is neither loss of information between the frames in the temporal dimension nor does it increase the computational burden.

Facial landmarks consist of the various components of the face which contribute to expressions, blink detection and related features. They include eyes, eyebrows, nose, mouth and jawline. Given an input image or a region of interest that includes the object of interest, a shape predictor is used to localize the key points of interest along the shape. Detection is a two-step process:

(a) The face must be localized from the image.
(b) Detection of key facial structures on the ROI (region of interest).

In Python, the face alone can be detected from the image using Haar cascade. Haar cascade is an object classifier model that detects the object that it is trained for. But a pre-trained model of HOG + Linear SVM object detector, for face detection is used for our purpose. The pre-trainer facial landmark detector is used to detect 68 two-dimensional coordinates that map the facial structures of the face (Fig. 1).

Thus, on obtaining the various facial landmarks, only the ones which are under consideration must be isolated. The landmarks indicated by the coordinates from 49 to 68 were taken into consideration. Since this is carried out on a single image, the same procedure was iterated throughout the entire video, for each video of every word in the dataset.

Fig. 1 68 facial landmark coordinates

After isolating the lip component of the face, a separate image of the lip alone was created and appended into a new video. Thus, the new video consists of a concise, cropped version of the original video of only the person's lips saying the word.

Figures 2 and 3 give us the flow of data preprocessing. The first phase converts all the videos into videos with just the lips and puts it a different directory. For this, the video directory is navigated and the frame of a video is read. Then, the face in the video is detected. Using facial landmarks predictor, the lips are cropped with enough scaling such that the end frame will be of a constant dimension of 40 * 60. The second phase is to convert the lip videos into arrays which are locally stored in hdf5 file format. For this, the modified video directory is navigated, and one out of three frames is read. This is done to reduce the computational burden on the system. This frame is stored as arrays, after all the videos are converted into arrays, and the arrays are converted into hdf5 files.

B. The Training Setup

In order to train a model consisting of this magnitude, a substantial amount of processing power was needed. Training on the generic processor would be time-consuming, so a good Nvidia GPU was required. Google Colab was thus used as our training environment.

Google Colab is a free cloud service for AI developers, where machine learning models can be trained by using GPUs free of cost. It is based on the Jupyter project, and all the code here is Python based. The work done, while engaging the virtual machine assigned, can be saved for future use or reference on the Google Drive account. The specifications of the virtual machine are:

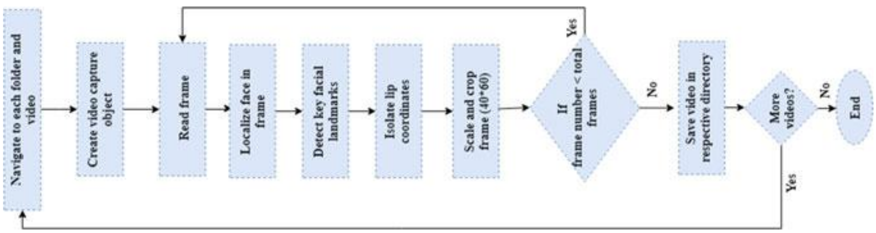


Fig. 2 Flow of data preprocessing (phase 1)

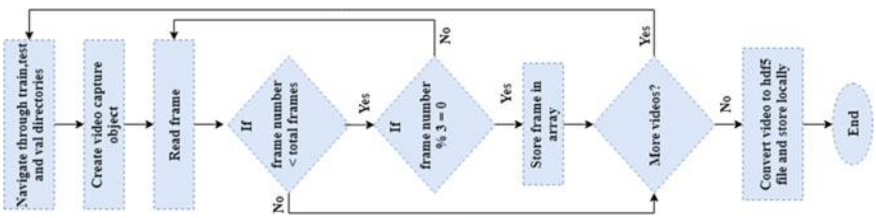


Fig. 3 Flow of data preprocessing (phase 2)

- GPU: 1x Tesla K80, compute 3.7, having 2496 CUDA cores, 12 GB GDDR5 VRAM
- CPU: 1x single-core hyper-threaded Xeon processors @2.3 Ghz (i.e., 1 core, 2 threads)
- RAM: 12.6 GB
- Disk: 33 GB.

The Google Colab virtual machine can be engaged for a maximum duration of 12 h. Since our training takes more than 12 h, the partially trained model at every step had to be saved back onto the Google Drive. The datasets are stored in an array data structure. Since the arrays are stored in the RAM, it was converted to hdf5 format for exporting. The transferring of files to and from the Jupyter Notebook on the virtual machine was carried out using the 'PyDrive' library in Python. 'PyDrive' enables transferring the model and the datasets saved in hdf5 format between Google Drive and Colab.

C. **Architecture**

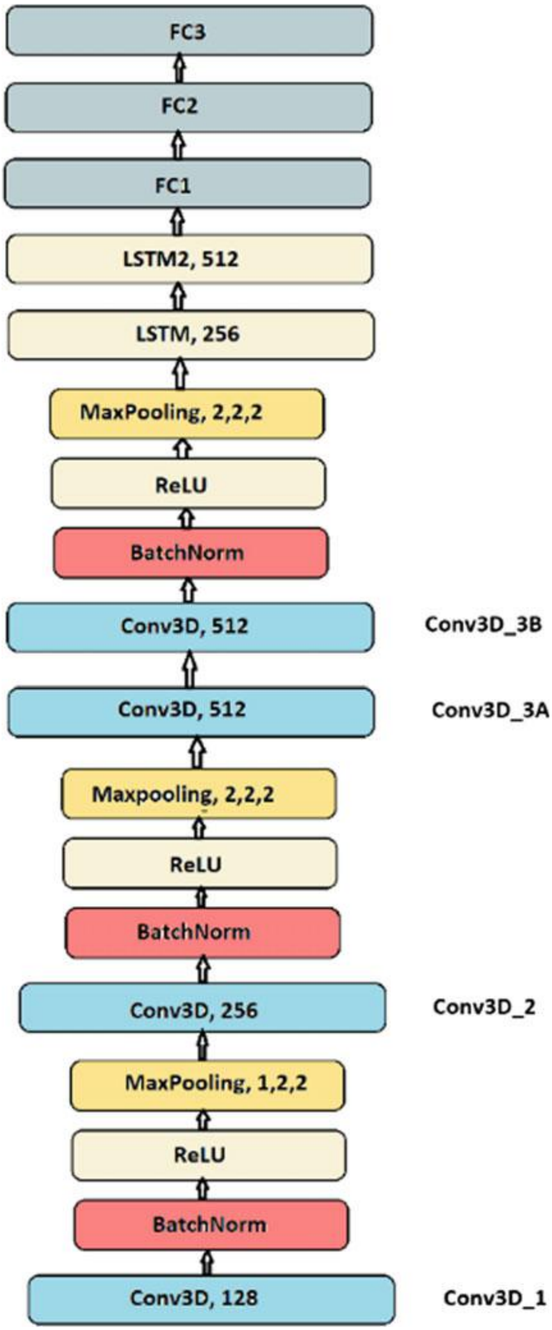
The architecture that was used is a combination C3D model [5] and ConvLSTM model [6]. The use of 3D CNN was to factor in the temporal data along with the spatial data of the frame. 3D CNN layer is followed by a BatchNorm layer. This particular layer is used to improve the speed, performance and stability of the neural network by normalizing the output of each layer and scaling the activation. We use the activation layer ReLU to improve the performance of batch normalization. A final layer, Max Pooling layer, is added before the repetition of the above layers. This layer reduces the spatial size and hence reduces computation required in the network. It preserves the important feature of the space and delivers it to the next layers.

The next part of the architecture is the ConvLSTM layers. A ConvLSTM is a variant of long short-term memory which has a convolution operation inside an LSTM cell. This layer can learn long-term spatiotemporal features. Since lip reading requires temporal data as a critical parameter, this type of layer proves to be very useful.

The last part of the architecture is the fully connected layers as seen in Fig. 4. This is a basic artificial neural network. This Dense layer has a specified number of neurons connecting the previous layers to upcoming layers.

Between each Dense layer, we have a dropout layer whose main intention is to reduce overfitting of the model by dropping a significant percentage of the neurons from the fully connected layer.

Fig. 4 3D CNN–ConvLSTM architecture



4 Mathematics Behind Choice of Layers

Adam optimizer
Adam is an acronym for adaptive moment estimator. This is a method that calculates an adaptive learning rate for each of its parameters, which is described by,

$$m_t = \beta_1 m_t - 1 + (1 - \beta_1) g_t \tag{1}$$

$$v_t = \beta_2 v_t - 1 + (1 - \beta_2) g_{2t} \tag{2}$$

m_t is the decaying average of the past gradients, v_t is the decaying average of the past squared gradients.
ReLU activation function
ReLU which stands for rectified linear unit is the most popular activation function used in almost all CNNs or DNNs. It is defined by,

$$R(z) = \max(0, z) \tag{3}$$

Batch normalization
The normalization of input as a preprocessing is not sufficient. Intermediate layer also needs normalization. Another problem is covariate shift. To solve this, batch normalization is used. To perform batch normalization, the following steps are done.

(1) The mean and variance of the input for a given layer are calculated as given in Eqs. (4) and (5).

The batch mean is obtained as:

$$\mu_B = \frac{\sum_{i=1}^{m} x_i}{m} \tag{4}$$

The batch variance is obtained as:

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_B)^2 \tag{5}$$

(2) The layer inputs are normalized using the calculated batch mean and variance as,

$$\bar{x}_i = \frac{(x_i - \mu_B)}{\sqrt{\left(\sigma_B^2 - \varepsilon\right)}} \tag{6}$$

(3) The output of the layer is obtained by scaling and shifting the normalized mean.

$$y_i = \gamma \bar{x}_i + \beta \tag{7}$$

3D CNN
3D CNN can also be represented by,

$$Yi, k, x, y, z = \sum_{c=0}^{C-1} \sum_{t=0}^{T=1} \sum_{r=0}^{R-1} \sum_{s=0}^{S-1} I_{i,c,x+t,y+r,z+s} F_{k,t,r,s,c'} \tag{8}$$

ConvLSTM
Convolution long short-term memory identifies spatiotemporal correlations better and outperforms fully connected long short-term memory.

The main equations of ConvLSTM are given below:

$$f_t = \sigma\left(W_{xf} * X_t + W_{hf}^{*} H_{t-1} + b_f\right) \tag{9}$$

$$i_t = \sigma\left(W_{xi}^{*} X_t + W_{hi}^{*} H_{t-1} + b_i\right) \tag{10}$$

$$\tilde{C}_t = \tanh\left(W_{x\tilde{C}} * X_t + W_{h\tilde{C}} * h_{t-1} + b_{\tilde{C}}\right) \tag{11}$$

$$O_t = \sigma(W_{xo} * X_t + W_{ho} * h_{t-1} + b_o) \tag{12}$$

$$C_t = f_t^s \circ C_{t-1} + i_t \circ \tilde{C}_t \tag{13}$$

$$H_t = o_t \circ \tanh(C_t) \tag{14}$$

where σ and tanh are the sigmoid and hyperbolic tangent functions, "o" is the Hadamard product and "∗" is the convolution operator. C_t the memory cell is the key module. It acts as an accumulator of the state information which are controlled by gates. [7] Gives a more description regarding Eqs. 9–14.

5 Results

Two sets of data were used, one with 20 classes and the other one with 200 classes. The top-1 and top-10 accuracies were calculated (Tables 1 and 2).

Top-1 accuracy is the conventional model accuracy, i.e., the model answer must match exactly with the expected answer. For top-10 accuracy, the expected answer must be contained in the 10 highest probable answers predicted by the given model.

Table 1 Accuracies for different number of classes

Number of classes	Top-1 accuracy (%)	Top-10 accuracy (%)
20	75.4	99.1
200	45.4	88.4

Table 2 Output at different stages for five selected words

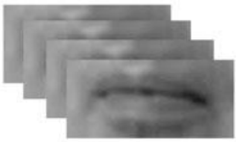
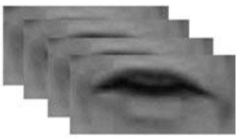
SL. NO	Total Frames (25-29)	Pre-processed data (9 frames)	Word predicted
1.			EDUCATION
2.			COMPANY
3.			DECISION
4.			AGREEMENT
5.			BUSINESS

Table 3 Comparison of the previous and proposed work

Sl. no.	Ref no.	Approach	Drawbacks
1	[8]	✔ Background subtraction was the technique used in which foreground image is extracted for further processing ✔ Detects a moving object from the difference between the frame and the frame of reference ✔ Horizontal and vertical projection of image is histogram over horizontal and vertical way of grayscale level	✔ The model is directly trained with ANN which does not take temporal data into account ✔ The segmentation, classification process are cumbersome and also uses a lot of computing ✔ Classifications of correct words were 67% for 3 classes
2	[9]	✔ 4 Different model architecture. EF-3, EF, MT-3 and MT ✔ Multiple pipelines or towers to capture temporal data effectively ✔ Data augmentation to reduce overfitting	✔ 3D convolutions not used to their advantage ✔ Training time is slower or large computational power required to train the model quicker
3	Proposed work	✔ Dynamic lip segmentation to reduce computational burden ✔ Conversion of image to grayscale from RGB ✔ Combination of 3D CNN and ConvLSTM to effectively capture spatial as well as temporal data ✔ 200 words uttered by 1000 different people	MERITS: ✔ Simple but deep model. ✔ Higher top-10 of 88.4% accuracy when compared to other papers ✔ Quicker training ✔ Diverse dataset Small size of trained model so that it can be incorporated in smaller memory devices

6 Comparison of Proposed Work with the Existing Works

Table 3 summarizes the different approaches and drawback of the referred papers and also gives an insight to the novelty of our work.

7 Novelty

The proposed work comes with dynamic lip segmentation, which is done so that redundant regions of the face as well as the background are completely removed. The trained model is also smaller and therefore could be easily put into smaller memory devices. The architecture is also a large improvement to the C3D as well as a normal 3D CNN model. Tweaking of hyper-parameters will lead to a higher accuracy.

8 Future Scope

So far, different models for lip reading have been trained and compared in this paper. The architectural parameters and hyper-parameters, namely learning rate, momentum, regularization effect, etc., are set based on trial and error or prior knowledge and experience in building the layers. These aspects on proper tuning can result in more accurate prediction. The trained models can be implemented in resource-constrained platforms such as Raspberry Pi, Android and iOS to make this system more accessible for whoever requires it for implementing in their own applications. We leave these tasks as areas for future work.

References

1. Chou, G., Hwang, J.: Lipreading from color video. TIP **6**(8), 1192–1195 (1997)
2. Saenko, K., Livescu, K., Siracusa, M., Wilson, K., Glass, J., Darrell, T.: Visual speech recognition with loosely synchronized feature streams. In: ICCV, pp. 1424–1431 (2005)
3. Zhao, G., Barnard, M., Pietikainen, M.: Lipreading with local spatial temporal descriptors. TMM **11**(7), 1254–1265 (2009)
4. Benhaim, E.; Sahbi, H., Vittey, G.: Continuous visual speech recognition for audio speech enhancement. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)
5. Tran, D., Bourdev, L.D., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3d convolutional networks. In: ICCV (2015)
6. Learning Spatiotemporal Features using 3DCNN and Convolutional LSTM for Gesture Recognition
7. Xingjian, S.H.I., Chen, Z., Wang, H., Yeung, D.Y.: Convolutional LSTM network: a machine learning approach for precipitation nowcasting (2015). arXiv: 1506.04214[cs.CV]
8. Arifin, F., Nasuha, A., Hermawan, H.D.: Lip reading based on background subtraction and image projection. In: 2015 IEEE International Conference on Information Technology Systems and Innovation (ICITSI)
9. Chung, J.S., Zisserman, A.: Lip reading in the wild. In: Asian Conference on Computer Vision (ACCV)

Predictive Analytics for Cardiovascular Disease Diagnosis Using Machine Learning Techniques

Anandhavalli Muniasamy, Vasanthi Muniasamy, and Roheet Bhatnagar

Abstract Medical data can be mined for effective decision making in the presence of disease analysis. Globally, cardiovascular alias heart disease is one of the highly rated causes of death [1] disease which will lead to 76% of the deaths [2] by the year 2030. Currently, the techniques of machine learning and predictive analytics have proven importance in medical data analysis. In a nutshell, this paper aims to apply six classifiers namely artificial neural network, support vector machine, decision tree, nearest neighbor, linear discriminant analysis, random forest, and to predict the presence of heart diseases in the patient's datasets. Moreover, the performance of these classifiers on three heart disease datasets is compared. The results reveal that RF, LDA, DT, and ANN have performed better than SVM and KNN in terms of accuracy, recall, F1-score, confusion matrix, and error rate.

Keywords Predictive analytics · Healthcare · Machine learning · Artificial neural network (ANN) · Support vector machine (SVM) · Decision tree (DT) · Nearest neighbor (NN) · Linear discriminant analysis (LDA) · Random forest (RF)

1 Introduction

The accelerating growth of medical data due to technological advancement enforces many challenges that involve imprecision and uncertainty in the process of decision making in medical analysis. Agarwal and Dhar [3] stated that the disease diagnosis has its own challenges in healthcare data analytics in terms of cost of treatment, diagnosis, and treatment at right time. Cardiovascular disease (CVD) refers to the disorder of heart and blood vessels, such as heart attack, cardiomyopathy, heart stroke, cardio arrhythmia, and congenital heart disease. American Heart Association (AHA) reported that the leading global cause of death is due to heart diseases [4].

A. Muniasamy (✉) · V. Muniasamy
College of Computer Science, King Khalid University, Abha, Kingdom of Saudi Arabia
e-mail: anandhavalli.dr@gmail.com

R. Bhatnagar
Department of CSE, Manipal University Jaipur, Jaipur, India

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_45

In addition, the World Health Organization (WHO) estimated that 76% of the death would be due to CVD by 2030 [5]. So, the identification and prevention of CVD should reduce this growing death rate [6, 7] and also the accurate and timely data analytics may support medical practitioners in the diagnostic process of CVD.

The medical industry today generates massive complex electronic patient data and their disease diagnosis information. The timely, accurate, and reliable analysis of this data will save cost and help practitioners in making valid diagnostic decisions. The applications of predictive analytics, such as optimizing the cost of resources, the accuracy of disease diagnosis, and enhancement of patient care improve clinical outcomes [8]. Machine learning techniques have proven importance in disease diagnosis to predict the presence of disease and have better understandings of it [9].

The paper aims to apply six classifiers namely artificial neural network, support vector machine, decision tree, nearest neighbor, linear discriminant analysis, random forest on three CVD datasets, and compare that which classifiers have better performance using various measures, like accuracy, precision, error rate, recall, and F1-score, confusion matrix in the prediction of the presence of heart disease.

In this paper, Sect. 2 reviews the application of machine learning classifiers in heart disease data analysis. The experimental methodology is described in Sect. 3. Section 4 reports a comparative analysis of the six machine learning classifiers. Lastly, conclusions are provided in Sect. 5.

2 Literature Review

Due to the complexity of healthcare data, the application of machine learning algorithm is a challenging task. In [9, 10], the authors described the new challenges in the machine learning domain due to the emergence of healthcare digitization. The applications of various machine learning classifiers have great impact on the identification and the prediction of various leading death rate diseases globally [11].

For the dataset with 'n' number of attributes, SVM maps each sample as a point or coordinates in an n-dimensional space for finding the class of the sample [12]. SVM finds a hyperplane to differentiate the two target classes for the sample classification. The classification process involves the mapping of the new sample into the n-dimensional space, based on which side of the hyperplane the new sample falls in. Burges [13] described SVM as the best tool to address bias–variance tradeoff, overfitting, and capacity control to work within complex and noisy domains. However, the quality of training data [13] decides the accuracy of SVM classifier. Moreover, [6, 13, 14] concluded the scalability is the main issue in SVM. In addition, the results reported in [6, 14, 15] stated that the use of optimization techniques can reduce SVM's computational cost and increase its scalability.

For the k-closet training samples with k as a user-defined constant, KNN classifies the test samples by assigning it to the nearest among the k-training samples which has closest distance to that sample in the attribute space as it is a nonparametric method. In [5], the authors stated that KNN is a lazy learner algorithm due to its easy understanding and implementation, but still it has various issues like the number of neighbors, appropriate distance measure selection, and preparation of class labels which affect its performance.

RF classifier can handle huge datasets with missing values. This classifier generates many decision trees in which each tree is selected based on voting. Then the selected tree is used to improve for the accuracy. Singh et al. [16] applied RF on the Cleveland heart disease dataset and achieved 85.81% accuracy, due to the nonlinear nature of this dataset. However, this ensemble classifier may give wrong decision on the classification of new sample due to the generation of many noisy decision trees [17].

LDA [18] classifies the dataset by applying dimensionality reduction and statistical methods [10]. It classifies the given samples appropriately into various classes using a linear combination of target values. Rossouw et al. [19] and Li et al. [20] concluded that LDA is one of the best choices for multi-class problems. Cako[21] applied LDA on Cleveland dataset and concluded that LDA is suitable for the smaller datasets with limited targets.

ANN is a mathematical model. The input layer of ANN receives the input sample, the backpropagation iteratively checks the errors and adjusts the weight values in the middle layers till achieving the similar results for the actual and the expected outputs, then the forward process will continue to minimize the errors, and finally, the output layers generate the accurate results. ANN algorithms are appropriate for machine learning tasks with higher accuracy, but it is too complex and requires high computation for solving the problems [4, 10, 22–25].

Verma et al. [7] report the applications of methods such as multinomial logistic regression model, multilayer perceptron fuzzy induction algorithm, and DT on Cleveland dataset and achieved 88.4%.

In [26], the authors applied decision tree classifier and concluded 69.5% accuracy for Long Beach dataset and 78.54% accuracy for Cleland dataset. Verma et al. [27] applied multilayer perceptron and DT on the Z-Alizadeh Sani dataset and got 77% accuracy for multilayer perceptron and DT did not offer higher value. The above-discussed research works reveal the applications of predictive analysis using DT, RF, DT, LDA, and SVM classifiers. We tried to improve these results by using more machine learning techniques.

3 Experimental Methodology

For the experiments, we used three different CVD datasets (Table 1). The first dataset includes the following: (1) Cleveland (303 records with 139 patients with positive diagnostics and 164 healthy people). (2) Switzerland (123 records with 115 patients with positive diagnostics and eight healthy people). (3) Hungary (294 records with 106 patients with positive diagnostics and 188 healthy people). (4) Long Beach VA (200 records with 149 patients with positive diagnostics and 51 healthy people) from the UCI data repository, in which each record has 14 variables. The second dataset

Table 1 Description of datasets

Cleveland, Switzerland, Hungary, Long Beach VA Dataset-1		Z-Alizadeh Sani Dataset-2		Western Cape, South Africa, Dataset-3	
Age	(28–77) years	Age	(30–86) years	Age	(15–54) years
Sex	0—female 1—male	Diabetes Mellitus	0—No 1—Yes	sbp	Blood pressure (101–218)
cp	Chest pain type 1—typical angina 2—atypical angina 3—non-anginal pain 4—asymptomatic	Fasting Blood sugar (mg/dl)	62–400	Alcohol	Alcohol consumption (0.00–147.19)
tresbps	Blood pressure (0–200)	Pulse rate	50–100		
chol	Serum cholesterol (0–603) (mg/dl)	ST elevation	0—No 1—Yes	Tobacco	Cumulative tobacco (kg) (0–31.2)
fbs	Fasting blood sugar 0—false(<120 mg/dl) 1—true (>120 mg/dl)	ST depression	0—No 1—Yes	LDL	Low-density lipoprotein cholesterol (0.98–15.33)
restecg	Resting electrocardiographic results 0—normal 1—ST-T wave abnormality 2—showing left ventricular hypertrophy	Low-density Lipoprotein (mg/dl)	18–232	Adiposity	Measure of % body fat (6.74–42.49)
thalach	Maximum heart rate achieved (60–202)	White Blood Cell	3700–18 000	Obesity	Measure weight-to-height rations (14.70–46.58)
exang	Exercise-induced angina 0—no 1—yes	Obesity	0—No (BMI < 25) 1—Yes (BMI > 25)	Famhist	Family history of heart disease (present, absent)
oldpeak	ST depression induced by exercise relative to rest (mm) (–2.6–6.2)	Creatine (mg/dl)	0.5–2.2		
Slope	The slope of the peak exercise ST segment 1—upsloping 2—flat 3—downsloping	Ex-Smoker	0—No 1—Yes	Type a	Type—an excessive competitive drive, impatience and anger/hostility

(continued)

Table 1 (continued)

Cleveland, Switzerland, Hungary, Long Beach VA Dataset-1		Z-Alizadeh Sani Dataset-2		Western Cape, South Africa, Dataset-3	
ca	Number of major vessels (0–3)	Hemoglobin (g/dL)	8.9–17.6		
Thal	3—normal 6–fixed defect 7-reversible defect	Non-anginal CP	0—No 1—Yes		
Num (target attribute)	0—absence; 1–4 presence; range least to most serious	Heart disease (target attribute)	0—Normal 1—Presence	chd (target attribute)	0—negative 1—positive

includes Western Cape, South Africa, [28] that has 462 records and 10 variables. The third dataset includes Z-Alizadeh Sani dataset [10] that has 303 records and 54 variables, in which 160 patients have positive diagnoses of CVD.

The target variable in dataset-1 refers to the negative diagnosis as 0 (zero), while a positive diagnosis is represented by an integer value from 1 to 4. Dataset-2 arranged the patients into four groups, namely demographics, ECG, laboratory, symptoms, and echo features based on 54 variables (only a few attributes are given in Table 1). The patient's diameter narrowing is greater than 50% for positive diagnosis results. In dataset-3, each patient was characterized by 10 variables.

Figure 1 depicts the experimental methodology used in this paper. Before simulating the algorithms, the datasets are preprocessed first to make them suitable for the classifiers. All the experiments are performed in Python platform. After that, every dataset has been divided into training and testing sets using the following ratios 80/20, 70/30, and 60/40, respectively. For the prediction models, the six algorithms

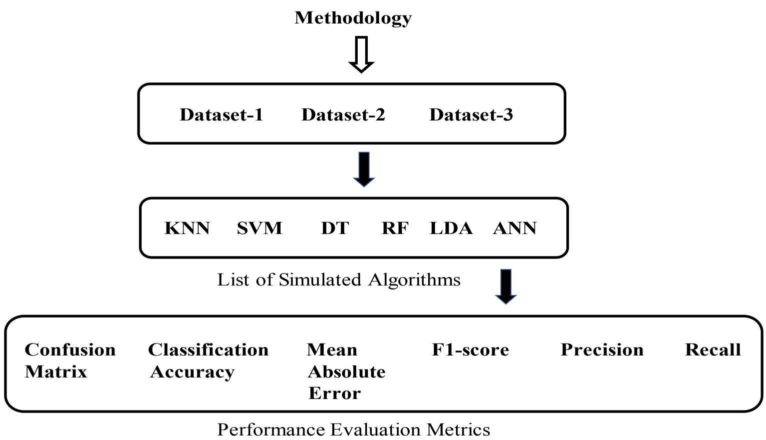


Fig. 1 Experimental methodology

are applied to the original datasets, and the performance of every classifier is analyzed using the metrics given in Fig. 1.

4 Results and Analysis

We analyzed and evaluated these six machine learning algorithms using the performance metrics namely confusion matrix, classification accuracy, precision, recall, F1-score, and error rate. The formulas for calculating these metrics are given in Table 2.

Confusion matrix (CM) shows the classification measures of all the samples of the given dataset. The CMs of all six algorithms simulated on three different datasets are given in Figs. 2, 3, and 4, respectively.

Performance measures, for the prediction of absence of CVD, namely classification accuracy percentage, precision, recall, F1-score, mean absolute error, of all six algorithms are presented in Table 3. The measures precision, recall, F1-score values given in Table 3 are for predicting the no heart disease. The performance of ANN, LDA, RF algorithms is very close to each other in terms of classification accuracy with respect to Table 3. The algorithms ANN, LDA, RF performed far better than SVM, KNN, and DT algorithms for the prediction of the absence of CVD in all three datasets as mentioned in Figs. 2, 3 and 4, respectively.

The performance of SVM classifier is lesser because all three CVD datasets have multi-labels, which is the weakness of SVM for handling multi-label data. SVM has better performance than KNN with 2% classification accuracy. The choice of nearest neighbors and distance measure affects the classification accuracy of KNN. All the

Table 2 Description of metrics

<table>
<tr><th>Name of the metrics</th><th>Formula</th></tr>
<tr><td>Confusion matrix</td><td>
<table>
<tr><td></td><td></td><td colspan="2">Predicted</td></tr>
<tr><td></td><td></td><td>0</td><td>1</td></tr>
<tr><td rowspan="2">Actual</td><td>0</td><td>True Negatives</td><td>False Positives</td></tr>
<tr><td>1</td><td>False Negatives</td><td>True Positives</td></tr>
</table>
</td></tr>
<tr><td>Classification accuracy</td><td>$\frac{(\text{True Positives} + \text{True Negatives})}{\text{Total}}$</td></tr>
<tr><td>Precision</td><td>$\frac{\text{True Positives}}{(\text{True Positives}+\text{False Positives})}$</td></tr>
<tr><td>Recall</td><td>$\frac{\text{True Positives}}{(\text{True Positives}+\text{False Negatives})}$</td></tr>
<tr><td>F1-score</td><td>$2 * \frac{\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}}$</td></tr>
<tr><td>Mean absolute error</td><td>$\frac{1}{m}\sum_{i=1}^{m}\lvert\text{Predicted} - \text{Actual}\rvert$</td></tr>
</table>

Actual/ Predicted	0	1	2	3	4
0	101	30	0	2	0
1	28	86	2	1	0
2	2	20	1	0	0
3	6	18	1	1	0
4	1	4	0	0	0

(a)

Actual/ Predicted	0	1	2	3	4
0	103	30	0	0	0
1	33	84	0	0	0
2	2	21	0	0	0
3	10	16	0	0	0
4	1	4	0	0	0

(b)

Actual/ Predicted	0	1	2	3	4
0	95	34	2	1	1
1	30	72	8	5	2
2	5	4	9	4	1
3	3	6	11	4	2
4	0	1	2	2	0

(c)

Actual/ Predicted	0	1	2	3	4
0	95	35	1	2	0
1	27	84	4	2	0
2	2	15	5	1	0
3	3	11	9	3	0
4	0	2	2	1	0

(d)

Actual/ Predicted	0	1	2	3	4
0	102	29	0	1	1
1	24	79	7	5	2
2	1	9	4	3	6
3	4	5	6	8	3
4	0	0	1	1	3

(e)

Actual/ Predicted	0	1	2	3	4
0	115	35	1	2	0
1	22	63	4	7	1
2	2	10	5	4	2
3	5	4	14	2	1
4	0	0	2	2	1

(f)

Fig. 2 Confusion matrix for dataset-1 **a** KNN; **b** SVM; **c** DT; **d** RF; **e** LDA; **f** ANNs

Actual/ Predicted	0	1
0	64	21
1	42	26

(a)

Actual/ Predicted	0	1
0	4	27
1	1	68

(b)

Actual/ Predicted	0	1
0	4	27
1	1	68

(c)

Actual/ Predicted	0	1
0	11	20
1	3	66

(d)

Actual/ Predicted	0	1
0	61	16
1	40	25

(e)

Actual/ Predicted	0	1
0	21	10
1	3	66

(f)

Fig. 3 Confusion matrix for dataset-2 **a** KNN; **b** SVM; **c** DT; **d** RF; **e** LDA; **f** ANNs

classifiers except ANN, RF, and LDA have poor performance for other metrics in dataset-2. The remaining algorithms, RF, LDA, DT, and ANN, performed the better in terms of other metrics in dataset-1 and dataset-3.

Actual/ Predicted	0	1
0	64	21
1	42	23

(a)

Actual/ Predicted	0	1
0	64	21
1	42	26

(b)

Actual/ Predicted	0	1
0	53	32
1	38	30

(c)

Actual/ Predicted	0	1
0	62	23
1	43	25

(d)

Actual/ Predicted	0	1
0	54	31
1	33	35

(e)

Actual/ Predicted	0	1
0	51	14
1	32	67

(f)

Fig. 4 Confusion matrix for dataset-3 **a** KNN; **b** SVM; **c** DT; **d** RF; **e** LDA; **f** ANNs

Table 3 Predicted metrics

Dataset	Method	Accuracy	Precision	Recall	F1-score	Mean absolute error
Cleveland, Switzerland, Hungary, Long Beach VA	KNN	0.71	0.73	0.76	0.75	0.43
	SVM	0.73	0.69	0.77	0.73	0.44
	DT	0.73	0.71	0.71	0.71	0.43
	RF	0.73	0.75	0.71	0.73	0.39
	LDA	0.74	0.78	0.77	0.77	0.37
	ANN	0.85	0.79	0.75	0.76	0.33
Z-Alizadeh Sani	KNN	0.70	0.6	0.8	0.69	0.41
	SVM	0.72	0.70	0.75	0.67	0.41
	DT	0.79	0.70	0.80	0.69	0.31
	RF	0.77	0.73	0.65	0.49	0.23
	LDA	0.76	0.65	0.48	0.56	0.24
	ANN	0.80	0.70	0.72	0.70	0.21
Western Cape, South Africa	KNN	0.71	0.60	0.80	0.69	0.30
	SVM	0.73	0.60	0.75	0.67	0.31
	DT	0.74	0.58	0.62	0.60	0.35
	RF	0.79	0.59	0.73	0.65	0.33
	LDA	0.79	0.69	0.74	0.76	0.32
	ANN	0.88	0.74	0.74	0.73	0.30

5 Conclusion

The medical domain generates massive heart disease data due to the advancement of information technologies. Finding the hidden knowledge in this data is a challenging task in machine learning. This paper reports the applications of some of the well-known machine learning algorithms including ANNs for the prediction of the presence of heart disease. We found that LDA, RF, DT, and ANN algorithms can give relatively higher accuracy results. The motivation of this paper is to support the medical practitioners for choosing the appropriate machine learning classifiers for the analysis of various heart disease samples.

Acknowledgements We acknowledge King Khalid University for providing the opportunity to initiate this paperwork and UCI data repository for datasets.

References

1. World Health Organization (WHO). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
2. World Health Organization: Prevention of diabetes mellitus-report of a WHO study group. WHO Technical Report Series no. 844, World Health Organization, Geneva (1994). http://whqlibdoc.who.int/trs/WHO_TRS_844.pdf
3. Agarwal, R., Dhar, V.: Big data, data science, and analytics: the opportunity and challenge for is research. Inf. Syst. Res. **25**(3), 443–448 (2014)
4. Heart Disease and Stroke Statistics-2019. https://healthmetrics.heart.org/wp-content/uploads/2019/02/At-A-Glance-Heart-Disease-and-Stroke-Statistics-%E2%80%93-2019.pdf
5. Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., et al.: Top 10 algorithms in data mining. Knowl. Inf. Syst. **4**(1), 1–37 (2012)
6. Alizadehsani, R., Habibi, J., Hosseini, M.J., Mashayekhi, H., Boghrati, R., Ghandeharioun, A., Bahadorian, B., Sani, Z.A.: A data mining approach for diagnosis of coronary artery disease. Comput. Methods Programs Biomed. **111**(1), 52–61 (2013)
7. Verma, L., Srivastaa, S., Negi, P.C.: A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data. J. Med. Syst. **40**(178) (2016). https://doi.org/10.1007/s10916-016-0536-z
8. Nithya, B.: Study on predictive analytics practices in health care system. IJETTCS **5** (2016)
9. Chowdhury, D.R., Chatterjee, M., Samanta, R.K.: An artificial neural network model for neonatal disease diagnosis. Int. J. Artif. Intell. Expert Syst. (IJAE) **2**(3) (2011)
10. Vanisree, K., Singaraju, J.: Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks. Int. J. Comput. Appl. (0975 – 8887) **19**(6) (2011)
11. Srinivas, K., Rani, B.K., Govrdhan, A.: Applications of data mining techniques in healthcare and prediction of heart attacks. Int. J. Comput. Sci. Eng. (IJCSE) **2**(2), 250–255 (2010)
12. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
13. Burges C.: A Tutorial on Support Vector Machines for Pattern Recognition. Bell Laboratories and Lucent Technologies (1998)
14. Burbidge, R., Buxton, B.: An introduction to Support Vector Machines for Data Mining. Computer Science Departmentm, UCL (2001)
15. Bardhan, I., Oh, J., Zheng, Z., Kirksey, K.: Predictive analytics for readmission of patients with congestive heart failure. Inf. Syst. Res. **26**(1), 19–39 (2014)
16. Singh, Y., Sinha, N., Sanjay, S.: Heart disease prediction system using random forest, pp. 613–623 (2017). https://doi.org/10.1007/978-981-10-5427-3_63
17. Fawagreh, K., Gaber, M.M., Elya, E.: RF.: From early developments to recent advancements. Syst. Sci. Control Eng. **2**(1), 602–609 (2014)
18. Pang, S., Ozawa, S., Kasabov, N.: Incremental linear discriminant analysis for classification of data streams. IEEE Trans. Syst. Man Cybern. B. **35**(5), 905–914 (2006)
19. Rossouw, J.E., du Plessis, J., Benade, A., Jordaan, P., Kotze, J., Jooste, P.: Coronary risk factor screening in three rural communities. S. Afr. Med. J. **64**, 430–436 (1983)

20. Li, T., Zhu, S., Ogihara, M.: Using discriminant analysis for multi-class classification: an experimental investigation. Knowl. Inf. Syst. **10**(4), 453–472 (2006)
21. Cako, S., Njeguš, A., Matić, V.: Effective diagnosis of heart disease presence using artificial neural networks. In: International Scientific Conference on Information Technology and Data Related Research, Belgrade, Singidunum University, Serbia, pp. 3–8 (2017)
22. Hongjun, Lu: Setiono R, Huan Liu: Effective data mining using neural networks. IEEE Trans. Knowl. Data Eng. **8**(6), 957–961 (1996)
23. Kamruzzaman, Sarkar A.: A new data mining scheme using artificial neural networks. Sensors **11**(12), 4622–4647 (2011)
24. Kumari, M., Godara, S.: Comparative study of data mining classification methods in cardiovascular disease prediction. IJCST **2**(2) (2011)
25. Singh, Y., Chauhan, A.S.: Neural network in data mining. J. Theor. Appl. Inf. Technol. 37–42 (2009)
26. El-Bialy, R., Salamay, M.A., Karam, O.H., Khalifa, M.E.: Feature analysis of coronary artery heart disease data sets. Procedia Comput. Sci. ICCMIT **265**, 459–468 (2015). https://doi.org/10.1016/j.procs.2015.09.132
27. Verma, L., Srivastaa, S., Negi, P.C.: A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data. J. Med. Syst. **40**(178) (2016)
28. Alizadehsani, R., Hosseini, M.J., Sani, Z.A., Gandeharioun, A., Boghrati, R.: Diagnosis of coronary artery disease using cost-sensitive algorithms. In: IEEE 12th International Conference on Data Mining Workshop, pp. 9–16 (2012)

A Novel Approach for Smart-Healthcare Recommender System

Richa Sharma and Shalli Rani

Abstract With the expeditious development of big data and internet of things (IoT), technology has successfully associated with our everyday life activities with smart healthcare being one. The global acceptance toward smart watches, wearable devices, or wearable biosensors has paved the way for the evolution of novel applications for personalized e-Health and m-Health technologies. The data gathered by wearables can further be analyzed using machine learning algorithms and shared with medical professionals to provide suitable recommendations. In this work, we have analyzed the performance of different machine learning techniques on public datasets of healthcare to select the most suitable one for the proposed work. Based on the results, it is observed that random forest model performs the best. Further, we propose a quantified self-based hybrid model for smart-healthcare environment that would consider user health from multiple perspectives and recommend suitable actions.

Keywords Big data · IoT · Machine learning · Recommender systems

1 Introduction

In recent years, the alliance of healthcare with technology has taken a global upsurge. Present-day healthcare applications have abridged the patient–physician distance by enabling real-time patient monitoring and healthcare management. IoT has achieved a strong foothold in almost every aspect of our lives by facilitating real-time observation and analysis in a ubiquitous manner. The exponential rise in the number of wearables, smart watches, fitness trackers, health monitoring devices and apps further provides provision for gathering real-time health-related data and physical monitoring [1, 2]. IoT wearables are connected with the patients' body to monitor various

R. Sharma · S. Rani (✉)
Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India
e-mail: shalli.rani@chitkara.edu.in

R. Sharma
e-mail: richa.sharma@chitkara.edu.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_46

health parameters including body temperature, blood circulation, blood pressure, glucose level, heart rate, breathing rate, etc. Therefore, multiple readings or measurements are taken dynamically in order to monitor the health conditions proactively. However, there are two challenges: (a) identifying the most suitable wearable from among a plethora of devices and (b) handling massive volume of data produced by IoT devices or wearables and using that data for decision-making process.

Machine learning algorithms play a key role in decision-making process even while handling massive volume of data. Further, machine learning acts as the backbone of recommender system by enabling it to learn about user interests, find similar users, find hidden patterns between users or items, etc. In recent years, IoT technology has begun to utilize recommender systems to optimize offered options to IoT users, analyzing devices data and user behaviors. IoT devices can gather a huge volume of data like user mood, location, health conditions, environmental conditions, etc. which can further be analyzed by recommender systems to provide more personalized recommendations.

In this paper, we study and investigate the role of recommender systems in IoT environment. Further, we perform an experimental study on publicly accessible healthcare datasets to analyze and evaluate various machine learning algorithms, namely multilayer perceptron, simple logistic, Bayes net, naïve Bayes, decision table and random forest. Based on the results obtained, we further propose an IoT-based hybrid framework for smart healthcare.

The major contributions of this study include:

(1) We provide an insight into the state-of-the art recommender systems and machine learning algorithms in IoT-based healthcare applications.
(2) We study and compare the performance of six machine learning algorithms on multiple healthcare datasets.
(3) Based on the analysis of the results obtained by comparing the machine learning algorithms, we have proposed an IoT-based smart-healthcare recommender system.

The following section is focused on related work in the field of IoT-based e-health and m-health technologies. Section 3 primarily covers the framework for the proposed system. Section 4 focusses on the performance evaluation of various classification models and their comparative analysis. Finally, Sect. 5 summarizes the conclusion and future scope of this work.

2 Related Work

Considerable amount of work has been done toward IoT-based e-healthcare systems by various researchers in the past. Previous works were primarily focused on applying machine learning algorithms on data collected using sensors for disease predictions and health monitoring. Following this line of work, authors proposed smart algorithms based on machine learning techniques to predict diabetes [3], thyroid [4], heart diseases [5], AMD disease [6], stress detection [7], monitor health conditions like fall detection [8] and human activity recognition [9].

Data captured using IoT devices, sensors or wearables is now being used by recommender systems to provide suggestions to physicians or patients based on their real-time conditions. Some of such works are discussed as follows. Authors presented an intelligent health recommender system emHealth, for patients suffering from depression or emotional disorder toward retrieving personalized therapies in the shortage of medical resources [10]. Based on the factors that lead to depression and the level of depression, personalized recommendations are given to the user in terms of treatment solutions and emotional improvement suggestions. Similarly, authors proposed a novel pro-active and personalized recommender system that recommends wearables and IoT solutions to an individual [11]. A user's health conditions are predicted frequently by analyzing his/her medical history, based on text mining and machine learning classifiers. In another work, authors introduced a Personalized Health-centric Travel Recommender System called ProTrip which generates personalized recommendations based on the travel pattern, actions and demographic information of the user to suggest the food availability as per their personal choice and nutritive value [12].

Another work is focused on novel recommendation methods in the IoT-enabled m-health field following the principle of quantified self, to recommend healthcare devices, mobile apps and physical movement plans for the patients using the concept of virtual coach and virtual nurse [13]. Authors presented a hybrid reasoning-based technique for predicting diseases, based on fuzzy set theory, k-nearest neighbor and case-based reasoning [14]. However, the system only considers user symptoms for prediction and not the risk factors that might help in early disease prediction. In a similar work, authors proposed an IoT-based hybrid recommender system for detecting cardiovascular diseases using wireless sensor networks and deliver personalized recommendations based on a patient's age, gender and clinical test results [15].

Based on the literature, we observe that the existing IoT-based e-health or m-health systems defined for specific purposes like patient tracking or monitoring, assisted living, diet monitoring or predicting diseases. Therefore, we propose a hybrid model based on the concept of quantified self for smart-healthcare environment that would consider user health from multiple perspectives and recommend suitable actions.

3 Proposed Framework for Smart Healthcare

The proposed smart health monitoring system involves three stages. Stage-1 concentrates on accumulating data using IoT wearable devices. Stage-2 is dedicated for storing and analyzing the gathered data. Further, stage-3 uses a hybrid model to provide suitable recommendations in case of abnormal condition as predicted in stage 2. Figure 1 depicts the proposed framework for IoT-based e-health monitoring system.

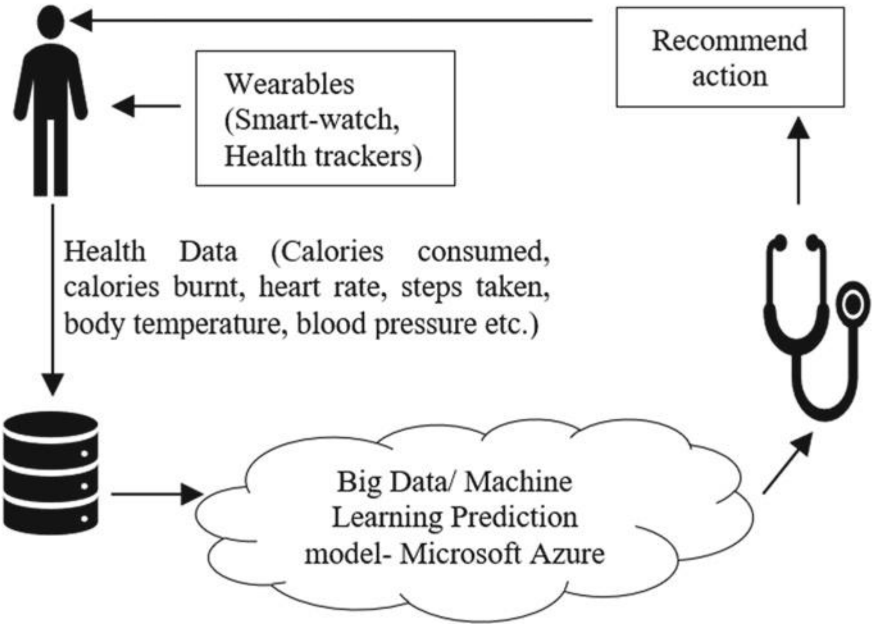


Fig. 1 Proposed framework

Stage 1: Collection
User physiological data is collected using wearable IoT sensors, health trackers and devices, which enables continuous data gathering, being attached to or worn by the user. IoT wearables attached to the human body gather the patient's physiological data in an active manner.

Stage 2: Storage and analysis
Since the attached sensors or IoT devices would gather data actively, a huge amount of data would be generated and storing such huge volume of data using traditional data processing tools is cumbersome. Therefore, we propose to use Microsoft Azure, a cloud framework which supports multiple machine learning algorithms, provides an environment to link IoT devices via the cloud and does not have scalability or security issues. The observations as gathered by the wearables would continuously be analyzed using random forest machine learning model. Random forest is a supervised learning algorithm that establishes multiple decision trees and fuses them to achieve better accuracy and stability. Previous works based on random forests have attained admirable results [7, 16]. Random forest, also referred to as ensemble model, creates an ensemble of decision trees based upon bagging mechanism to solve a classification problem.

Stage 3: Recommend action
When any of the measured physiological value of the user exceeds its normal value, an immediate notification would be sent to his mobile device recommending an action. In this work, we propose to use a hybrid recommendation model. Existing work on recommendations in smart healthcare follows feature-combination model for developing the hybrid model [13, 14, 17]. However, there are few other hybrid models which have not been used in smart healthcare yet [18].

A sample use-case is given in Fig. 2. Suppose a user Alex wants to go on a diet and has set a target on his smartphone application about how many calories he wishes to burn. His real-time observations about the calories he consumes a day and the calories he is yet to burn are displayed on his smart watch. If Alex has a cheat meal during lunch, the calories he consumed would automatically be fetched by his smart watch. The prediction model is continuously getting these readings, and as soon as it observes that the number of calories consumed by Alex is greater than the calories burnt or he was supposed to burn, it immediately sends this information to the recommendation model. Now, the recommendation model analyzes the whole scenario, and based on the target as set by Alex, his present health condition, his vitals, location and remaining time of the day, it recommends a series of activities to Alex and sends a notification on his smartphone about what he may do to get back on track and still follow the weight loss plan.

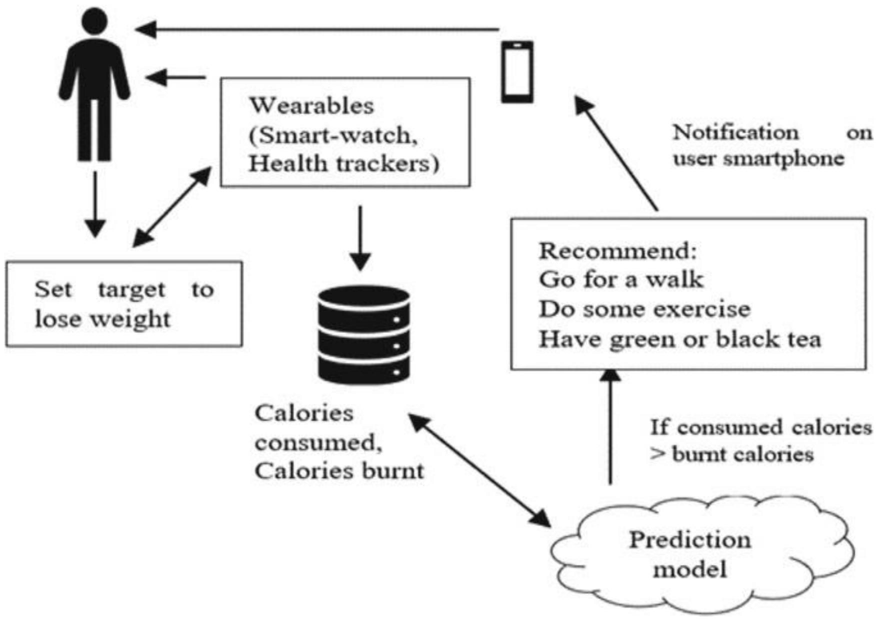


Fig. 2 Use-case example

4 Results and Discussions

In this section, we will evaluate and compare the performance of various machine learning algorithms. We perform an experimental study on chronic kidney disease and Autism_Child datasets, which are publicly accessible, to analyze and evaluate various machine learning algorithms, namely multilayer perceptron, simple logistic, Bayes net, naïve Bayes, decision table and random forest. The evaluation of the algorithms has been done based on mean absolute error (MAE), root mean square error (RMSE), precision, recall and *f*-score. The evaluation is done on WEKA tool using 10-fold cross validation.

(i) *Mean Absolute Error* (*MAE*): It is sometimes also referred to as absolute deviation. It is a measure of the deviation of observations from their true self. It is given by:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N} p_i - o_i}{N} \tag{1}$$

Here,

p_i predicted readings
o_i observed readings
N no. of observations considered

(ii) *Root Mean Square Error* (*RMSE*): It is a measure of standard deviation of the differences between the predicted and observed values. It is computed like MAE but squares the error before summing it:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (p_i - o_i)^2}{N}} \tag{2}$$

(iii) *Precision*: It is referred to as the gauge of goodness. It is the measure of fittingly predicted positive observations to the total predicted positive observations. It can be calculated as:

$$\text{Precision} = \frac{tp}{tp + fp} \tag{3}$$

(iv) *Recall*: It is referred to as the gauge of completeness. It is the ratio of fittingly predicted positive observations to all observations in definite class—yes. It can be calculated as:

$$\text{Recall} = \frac{tp}{tp + fn} \tag{4}$$

(v) *F-score*: It is the weighted average of precision and recall. Hence, this score considers both false positives and false negatives. It can be calculated as given in 5.

$$\text{F-score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{5}$$

The results obtained after applying the aforementioned algorithms on the dataset are tabulated in Table 1. From Table 1, it can be observed that random forest has performed equally well for both the datasets with precision and recall value being 1. Decision table follows next with 0.99 and 1 precision–recall in chronic kidney disease and Autism dataset, respectively. Further, Bayes net algorithm has the minimum MAE value as compared with other algorithms and also has a good precision and recall score of 0.988 and 0.99 for both the datasets, respectively. Therefore, we will use random forest algorithm in the proposed framework as it classified all the test values correctly with 0 misclassification for the two datasets considered. Figures 3 and 4 depict and compare the performance of the used algorithms on the basis of MAE, RMSE, precision, recall and *f*-score, evaluated for both the datasets considered, respectively.

Based on the results obtained, we can clearly see that random forest algorithm outperformed the other machine learning algorithms with 0% misclassification rate

Table 1 Results

Algorithm	MAE	RMSE	Precision	Recall	*F*-score
Chronic kidney disease					
Multilayer perceptron	0.0085	0.0622	0.998	0.998	0.998
Simple logistic	0.0222	0.1068	0.981	0.98	0.98
Bayes net	0.013	0.1035	0.988	0.988	0.988
Naïve Bayes	0.0479	0.2046	0.956	0.95	0.951
Decision table	0.1815	0.2507	0.99	0.99	0.99
Random forest	0.0414	0.0844	1	1	1
Autism_Child					
Multilayer perceptron	0.0091	0.0497	0.997	0.997	0.997
Simple logistic	0.2334	0.2682	1	1	1
Bayes net	0.0011	0.0041	1	1	1
Naïve Bayes	0.0444	0.1098	0.99	0.99	0.99
Decision table	0.0076	0.0076	1	1	1
Random forest	0.1039	0.1339	1	1	1

Fig. 3 Result analysis (chronic kidney disease)

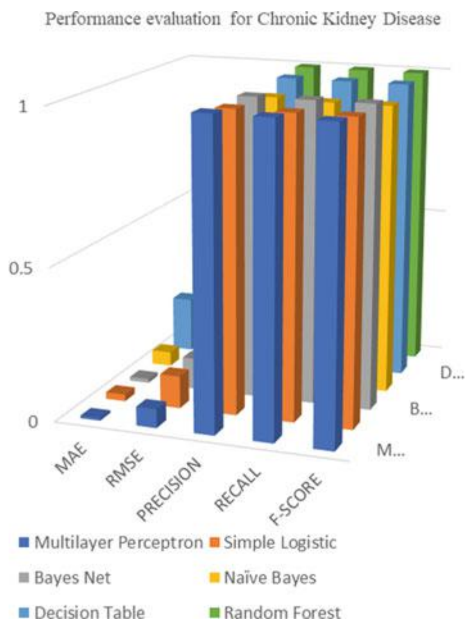
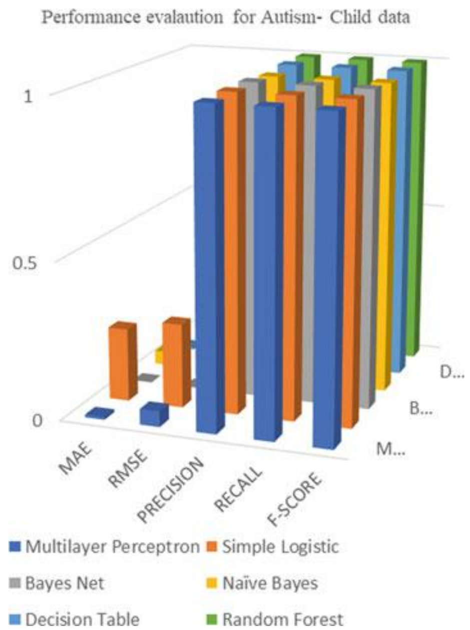


Fig. 4 Result analysis (Autism-Child)



for both the health datasets considered. Since the proposed framework deals with real-time health dataset, the information gathered is sensitive both in terms of user privacy and for intelligent decision making. Therefore, random forest algorithm is preferred over the other algorithms. However, our objective is to design a framework which is not just accurate but is reliable as well. For instance, the performance of an algorithm varies depending on the volume of data as well as the number of parameters considered. Therefore, the proposed framework does not rely solely on a single algorithm; rather, we intend to use a hybrid model for better reliability.

5 Conclusion and Future Scope

In this work, we have analyzed and compared the performance of six machine learning models on publicly accessible healthcare datasets. The models considered in this study are multilayer perceptron, simple logistic, Bayes net, naïve Bayes, decision table and random forest. The evaluation of the algorithms has been done based on mean absolute error (MAE), root mean square error (RMSE), precision, recall and *f*-score. The evaluation is done on WEKA tool. From the results, it was observed that random forests performed the best in terms of precision, recall and *f*-score with 100% performance rate for both the datasets considered.

Further, existing work on recommendations in smart healthcare follows feature-combination model for developing the hybrid model. However, there are many combinations of recommendation approaches which have not been explored yet. Therefore, we propose to design a hybrid recommendation model for the proposed framework that would consider user health from multiple perspectives, unlike the existing models. This work opens promising room for the development of quantified self-frameworks. Although we intend to test the proposed approach in coming future using a real-time learning environment to obtain feedback from the users, lack of data poses a huge research challenge for us.

References

1. Lorincz, K., Malan, D.J., Fulford-Jones, T.R., Nawoj, A., Clavel, A., Shnayder, V., Mainland, G., Welsh, M., Moulton, S.: Sensor networks for emergency response: challenges and opportunities. IEEE Pervasive Comput. **4**, 16–23 (2004)
2. Shafiq, M.Z., Ji, L., Liu, A.X., Pang, J., Wang, J.: A first look at cellular machine-to-machine traffic: large scale measurement and characterization. ACM SIGMETRICS Perform. Eval. Rev. **40**(1), 65–76 (2012)
3. Devi, M.R., Shyla, J.M.: Analysis of various data mining techniques to predict diabetes mellitus. Int. J. Appl. Eng. Res. **11**(1), 727–730 (2016)
4. Turanoglu-Bekar, E., Ulutagay, G., Kantarcı-Savas, S.: Classification of thyroid disease by using data mining models: a comparison of decision tree algorithms. Oxford J. Intell. Decis. Data Sci. **2**, 13–28 (2016)

5. Kumar, P.M., Gandhi, U.D.: A novel three-tier internet of things architecture with machine learning algorithm for early detection of heart diseases. Comput. Electr. Eng. **65**, 222–235 (2018)
6. Das, A., Rad, P., Choo, K.K.R., Nouhi, B., Lish, J., Martel, J.: Distributed machine learning cloud teleophthalmology IoT for predicting AMD disease progression. Future Gener. Comput. Syst. **93**, 486–498 (2019)
7. Pace, P., Aloi, G., Gravina, R., Caliciuri, G., Fortino, G., Liotta, A.: An edge-based architecture to support efficient applications for healthcare industry 4.0. IEEE Trans. Ind. Inf. **15**(1):481–489 (2019)
8. Yacchirema, D., de Puga, J.S., Palau, C., Esteve, M.: (2019) Fall detection system for elderly people using IoT and ensemble machine learning algorithm. Pers. Ubiquit. Comput. 1–17
9. Subasi, A., Radhwan, M., Kurdi, R. and Khateeb, K.: IoT based mobile healthcare system for human activity recognition. In: 2018 15th Learning and Technology Conference (L&T), pp. 29–34. IEEE (2018)
10. Yang, S., Zhou, P., Duan, K., Hossain, M.S., Alhamid, M.F.: emHealth: towards emotion health through depression prediction and intelligent health recommender system. Mobile Netw. Appl. **23**(2), 1–11 (2017)
11. Asthana, S., Megahed, A., Strong, R.: A recommendation system for proactive health monitoring using IoT and wearable technologies. In: International Conference on AI & Mobile Services, Honolulu, USA. IEEE, pp. 14–21 (2017)
12. Subramaniyaswamy, V., Manogaran, G., Logesh, R., Vijayakumar, V., Chilamkurti, N., Malathi, D., Senthilselvan, N.: An ontology-driven personalized food recommendation in IoT-based healthcare system. J. Supercomput. **74**, 1–33 (2018)
13. Felfernig, A., Ngoc, T.: Recommender systems for IoT enabled m-health applications. Artif. Intell. Appl. Innov. **520**, 227–237 (2018)
14. Malathi, D., Logesh, R., Subramaniyaswamy, V., Vijayakumar, V., Sangaiah, A.K.: Hybrid reasoning-based privacy-aware disease prediction support system. Comput. Electr. Eng. **73**, 114–127 (2018)
15. Jabeen, F., Maqsood, M., Ghazanfar, M.A., Aadil, F., Khan, S., Khan, M.F., Mehmood, I.: An IoT based efficient hybrid recommender system for cardiovascular disease. Peer-to-Peer Netw. Appl. **12**, 1–14 (2019)
16. Wang, Y., Wu, K., Ni, L.M.: Wifall: device-free fall detection by wireless networks. IEEE Trans. Mob. Comput. **16**(2), 581–594 (2016)
17. Felfernig, A., Polat-Erdeniz, S., Uran, C., Reiterer, S., Atas, M., Tran, T.N.T., Azzoni, P., Kiraly, C., Dolui, K.: An overview of recommender systems in the internet of things. J. Intell. Inf. Syst. **51**, 1–25 (2018)
18. Çano, E., Morisio, M.: Hybrid recommender systems: a systematic literature review. Intell. Data Anal. **21**, 1487–1524 (2017)

Heart Disorder Prognosis Employing KNN, ANN, ID3 and SVM

Jyoti Deshmukh, Mukul Jangid, Shreeshail Gupte, and Siddhartha Ghosh

Abstract Cardiovascular disease (CVD) stands as a type of disease that incorporates the heart or the blood vessels. Over the past few decades, it has become the common result of a death in underdeveloped, developing as well as in developed countries. This field is thus, "data rich" but "knowledge poor". Here, data mining holds great potential and immensely helps the health system to systematically use the data and analytics to recognize the inefficiencies thus, reducing the practice of burdensome tests. Author presents the Heart Disorder Prognosis System for accurate detection of heart disease which has been derived from distinctive analysis among several data mining algorithms. The presence of heart disorder in a sufferer is forecasted by digging out appealing patterns from the datasets. The datasets used for analysis are fetched from the UCI Machine Learning Repository, namely Cleveland Clinical Foundations and the Hungarian Institute of Cardiology. This paper tries to introduce the methodology, implementation and analysis of Decision Tree (ID3), Support Vector Machine (SVM), Artificial Neural Network (ANN) and K-Nearest Neighbor (KNN) algorithm for detection of heart diseases. The conclusion is induced on the basis of accuracy and ROC value. ID3 algorithm gives better performance over other algorithms for both the datasets.

Keywords Classifiers · Heart disease · Prediction · Sensitivity · Specificity

J. Deshmukh · M. Jangid (✉) · S. Gupte · S. Ghosh
Department of Computer Engineering, Rajiv Gandhi Institute of Technology, Andheri West, Mumbai 400053, Maharashtra, India
e-mail: mkljngd@gmail.com

J. Deshmukh
e-mail: jyotideshmukh11@gmail.com

S. Gupte
e-mail: shreeshail04@gmail.com

S. Ghosh
e-mail: sidd.ghosh9@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_47

1 Introduction

Deaths due to CVD are more frequent computing to approximate 17.3 million loss of life every year [1]. Doctor's experience is used to predict health issues of the patient which suppresses the knowledge abundant data resulting in very high medical costs and needless medical tests. To figure out these problems, data mining solutions along with the help of medical datasets was introduced. Results proved that the suggested system has its own exclusive capability in fulfilling the aim of the described mining target. The paper is concerned about accomplishing the most optimized technique amidst NN, DT, KNN, and SVM. The fuzzy K-NN [2], prediction output showed convincing nature, for eliminating the repeated data and to improve the efficiency of the classifier. There are various types of decision tree algorithm, but the most preferred is the ID3 [3, 4] algorithm that is Iterative Dichotomized 3 by J. R. Quinlan. ANN [5] is a powerful tool for data analysis and employed for developing a detailed nonlinear scheme. SVM [6–8] used for overcoming the high dimensionality problem which is one of the main advantages of SVM. High dimensionality problem arises when the number of observations is less than the number of input variables.

Suggested method is described in Module 2. ID3, KNN, SVM ANN are presented in Sect. 2. Module 3 gives practical results. Module 4 consists of the conclusion and future work.

2 Proposed Methodology

The predictions systems used by the authors are ANN, ID3, SVM, and KNN on the Python platform. It has the capability to find the probability of the presence of heart disorder in victims depending upon a few medical attributes such as chest pain, sex, age, and slope. In the previous research paper, the heart disease prognosis system has been advanced employing all the attributes. Proposed schemes employed a varied number of attributes that is 14 out of 15 attributes and 12 out of the 15 attributes to analyse the difference in the accuracies. Figure 1 describes the data mining process which is performed in six steps.

2.1 Classifiers

Classifiers are the functions that learn from the target attributes which are further used for classifying new records.

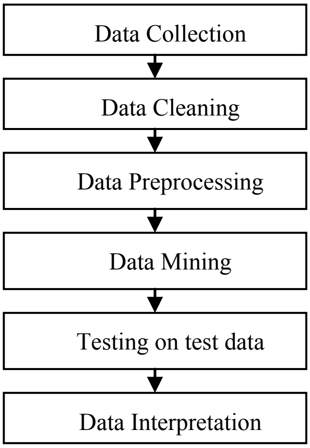


Fig. 1 Flow diagram for data mining process

2.2 ID3 Classifier

Decision tree [3, 9] is a scientific model used for prediction. The structure of the tree is described in Fig. 2. The gain ratio decision tree used is based on the scheme that chooses the dividing property thus reducing the value of entropy and increasing the information gain. The difference between the original information content and the amount of information is known as information gain. A comma separated value (CSV) file was given as input for training the algorithm where the train test split function was used using the train test Split library. Here, the dataset gets separated as training (80%) and testing (20%). The ID3 algorithms consist of various attributes with one target class (Here that target class is num).

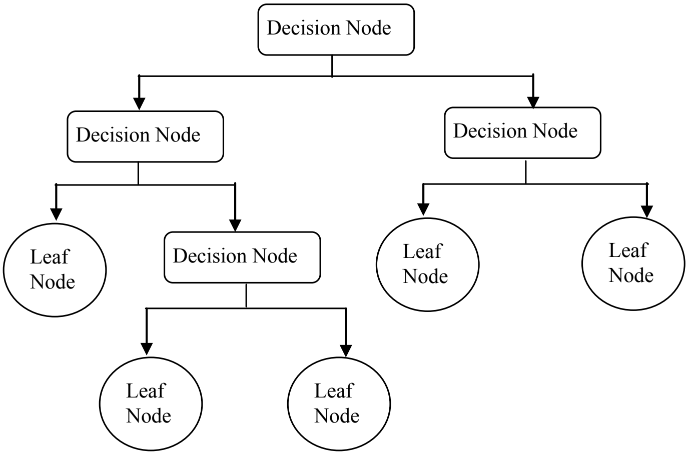


Fig. 2 Decision tree diagram

$$\text{Entropy } (F) = \sum_{i=1}^{n} P_i \log_2 P_i \tag{1}$$

$$\text{Gain } (F, A) = \text{Entropy } (F) - \sum_{k=1}^{m} \frac{|F(k)|}{s} \text{Entropy } (F_k) \tag{2}$$

$$\text{Gain Ratio } (F, A) = \frac{\text{Gain } (F, A)}{\text{Split information } (F, A)} \tag{3}$$

$$\text{Split Information } = -\sum \frac{|Fi|}{|F|} \times \log_2 \left[\frac{|Fi|}{|F|} \right] \tag{4}$$

$$\text{Info } (I) = -\sum_{i=1}^{n} P_i \log_2 P_i \tag{5}$$

$$\text{Info}_B(I) = \sum_{k=1}^{u} \frac{|Ik|}{|I|} \times \text{Info } (I_k) \tag{6}$$

$$\text{Gain } (B) = \text{Info } (I) - \text{Info}_B(I) \tag{7}$$

2.3 ANN Classifier

Neural network [3, 9, 10] is another important technique of data mining. The ANN structured is a multilayer perceptron neural network which consists of three layers. The diagram for heart disease prediction system using a neural network is described in Fig. 3. Rectifier function abbreviated by 'relu' is utilized as an activation function. The loss method utilized is logarithmic, while ADAM, a gradient descent procedure, abbreviated by 'adam', is utilized for optimization. The measure that we account for

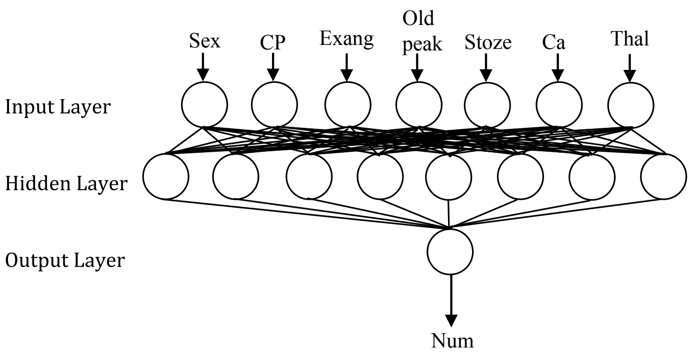
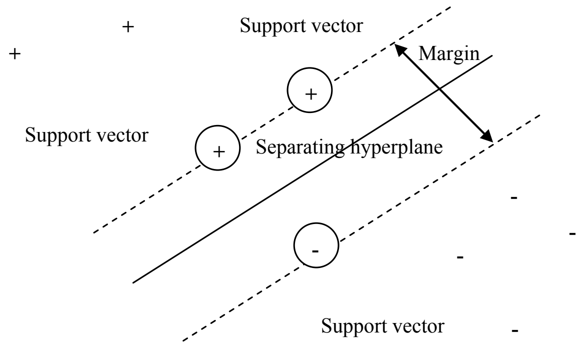


Fig. 3 Neural network diagram for heart disease prediction

Fig. 4 SVM classifier



this classification is accuracy. ANN scheme considered for a set size kept as 'auto'. Learning rates are used for weight updates, and here, the learning rate used in our research work is 'adaptive'.

The learning_rate_init is used for controlling the step-size in weight updation which is set to 0.001. Maximum number of iterations set are 500.

2.4 SVM Classifier

SVM [6–8, 11] a classification method that uses hyperplane. SVM is an example of points in space such that the separate categories are separated by a distinct gap which describes itself in Fig. 4. In this method, the hyperplane providing the maximum distance between the two classes is selected. It is used to overcome the high dimensionality problem which is one of the main advantages of SVM. In this research paper, the implementation of the SVM classifier is based on 'libsvm'.

2.5 KNN Classifier

KNN [2, 12] is the simplest algorithmic rule compared to different machine learning algorithms. The flow diagram of the KNN algorithm is described in the Fig. 5. KNN [2, 12] could be a non-parametric technique that is employed for classification and regression. In this research work, the number of neighbours used for k-neighbours queries is set to 5, the weight function used for prediction is set to 'uniform', and the metric function is set to 'None'.

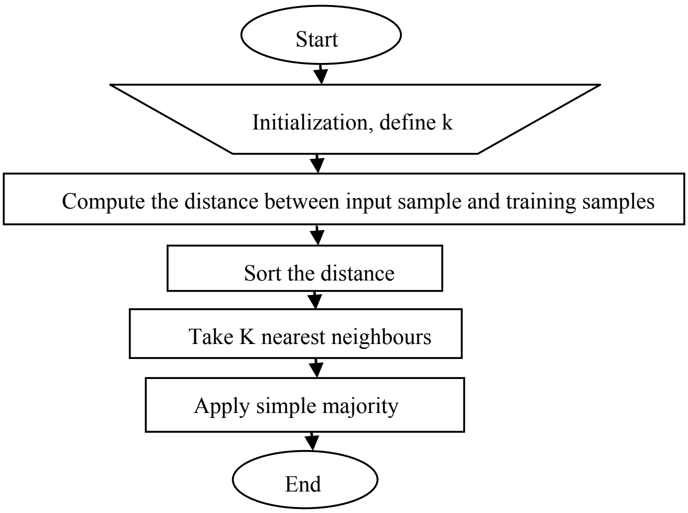


Fig. 5 Flow diagram of KNN algorithm

3 Experimental Results

The procedure is executed in Python language employing two different datasets. Dataset Hungary with a total of 457 samples (training = 365, testing = 92) and the dataset of Cleveland with total of 303 samples (training = 241, testing = 62) were used. Tables 1 and 2 are the accuracy tables wherein Table 1 gives the accuracy for Cleveland dataset and Table 2 gives the accuracy for Hungary dataset using ID3 followed by ANN, SVM and KNN classifiers. Figures 6, 7, 8 and 9 show ROC plot for Cleveland dataset, and Figs. 10, 11, 12 and 13 show ROC plot for Hungary

Table 1 Describes the result of Cleveland dataset

Sr No.	Algorithm used	Accuracy (%)
1	ID3	100
2	Artificial neural network	100
3	Support vector machine	67.00
4	K-nearest neighbour	93.44

Table 2 Describes the result of Hungary dataset

Sr No.	Algorithm used	Accuracy (%)
1	ID3	84.78
2	Artificial neural network	73.91
3	Support vector machine	80.03
4	K-nearest neighbour	76.01

Fig. 6 ID3–ROC (Cleveland data)

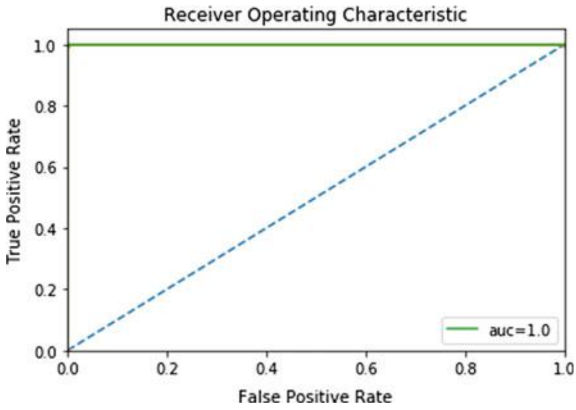


Fig. 7 ROC–ANN (Cleveland data)

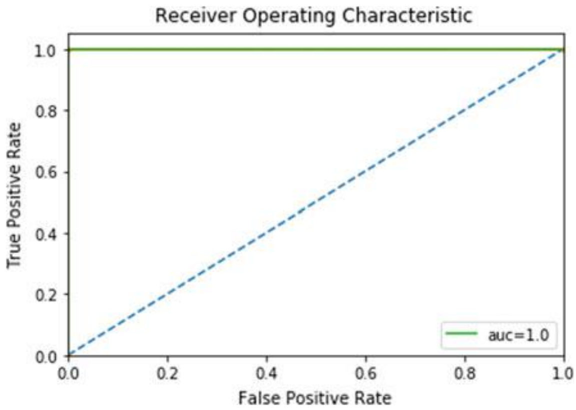
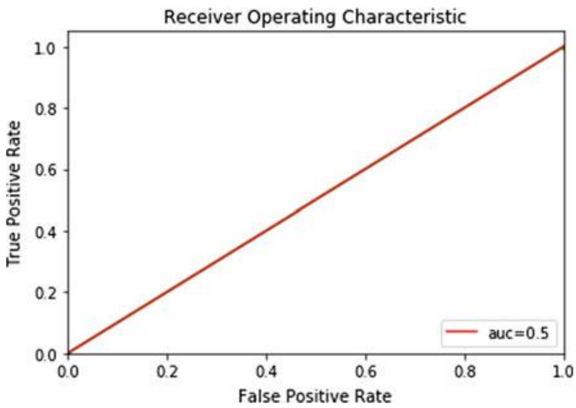


Fig. 8 SVM–ROC (Cleveland data)



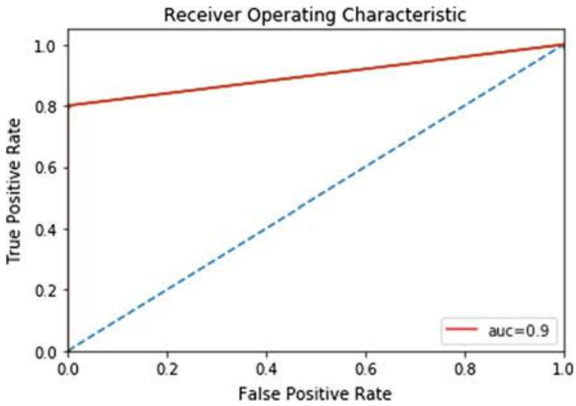


Fig. 9 KNN–ROC (Cleveland data)

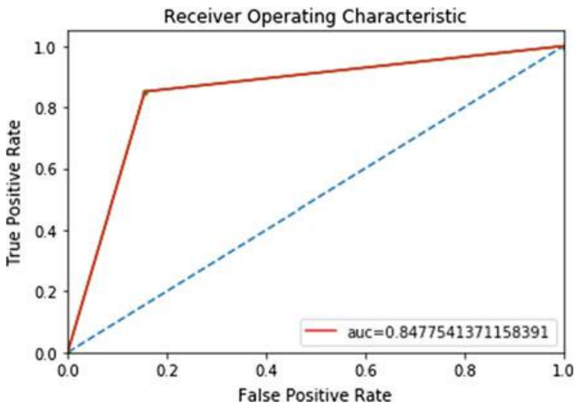


Fig. 10 ANN–ROC (Hungary data)

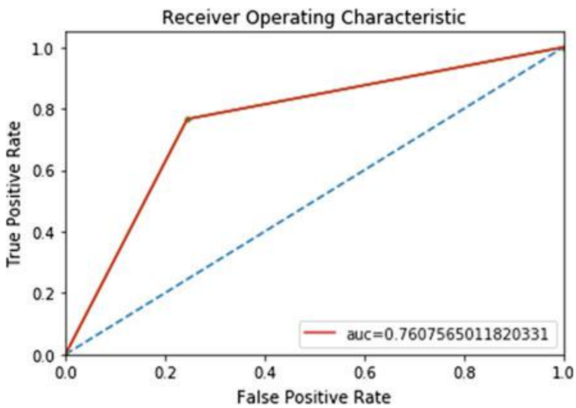


Fig. 11 KNN–ROC (Hungary data)

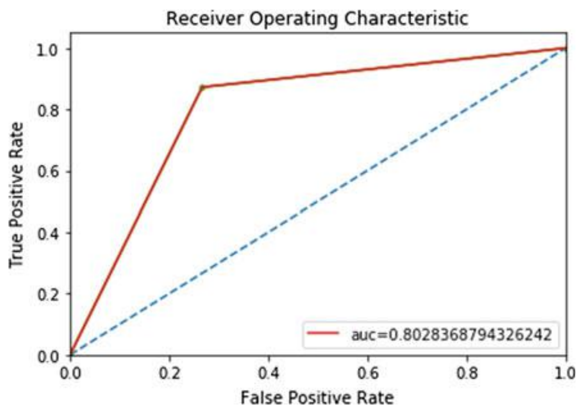


Fig. 12 SVM–ROC (Hungary data)

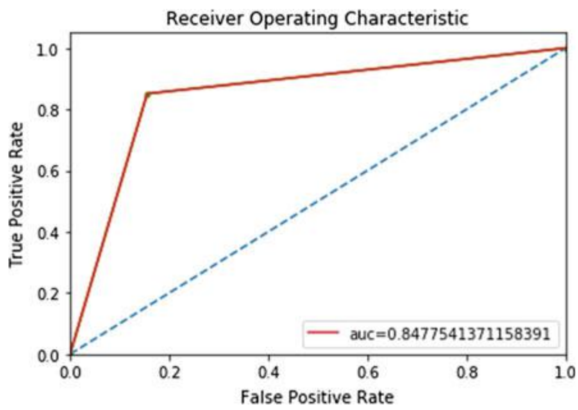


Fig. 13 ID3–ROC (Hungary data)

dataset using all the classifiers. Tables 3 and 4 present the performance metrics for the datasets of Cleveland and Hungary, respectively, with all the classifiers. Performance of these classifiers is evaluated on the basis of specificity, accuracy, sensitivity, recall, precision and AUC. It is observed that ID3 algorithm exhibits better performance over other algorithms in terms of accuracy and ROC measures for heart disease prediction.

Table 3 Performance measures with component classifiers for Hungary database

Classifier	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	Accuracy (%)	AUC
ID3	84.44	75.77	85.00	85.00	84.78	0.8477
ANN	72.34	75.55	74.00	74.00	73.91	0.8477
SVM	84.61	77.35	81.00	80.00	80.03	0.8028
KNN	75.55	72.34	76.00	76.00	76.01	0.7608

Table 4 Performance measures with component classifiers for Cleveland database

Classifier	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	Accuracy (%)	AUC
ID3	100.00	100.00	100.00	100.00	100.00	1.0
ANN	100.00	100.00	100.00	100.00	100.00	1.0
SVM	67.00	33.00	45.00	67.00	67.00	0.50
KNN	93.44	100.00	93.44	93.44	93.44	0.90

4 Conclusion and Future Scope

Early detection of heart disorder is crucial as it is one of the leading causes of death worldwide. According to the results, ID3 performed well on the Hungary dataset followed by SVM with the accuracy of 84.08% and 80.43%, respectively. Similarly, ANN and ID3 performed well for the Cleveland dataset with the accuracy of 100% for both the algorithms. The future scope of this paper is to make use of the genetic algorithm for improving the accuracy, in the near future. Also, the time required for execution can be minimized.

Acknowledgement The authors intend to convey gratefulness to Dr. Andras Janosi, M.D. and Dr. Robert Detrano, M.D., Ph.D., for permitting us to use the heart disease data set for our research work.

References

1. Chadha, R., Mayank, S.: Prediction of heart disease using data mining techniques. CSI Trans. ICT **4**(2–4), 193–198 (2016)
2. Krishnaiah, V., Narsimha, G., Chandra, N.S.: Heart disease prediction system using data mining technique by fuzzy KNN approach. In: Emerging ICT for Bridging the Future- Proceedings of the 49th Annual Convention of the Computer Society of India (CSI), vol. 1, pp. 371–384. Springer, Cham (2015)
3. Gandhi, M., Singh, S.N.: Predictions in heart disease using techniques of data mining. In: Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), International Conference, pp. 520–525. IEEE (2015).
4. Bhatia, S., Prakash, P., Pillai, G.N.: SVM based decision support system for heart disease classification with integer-coded genetic algorithm to select critical features. In: Proceedings of the World Congress on Engineering and Computer Science, pp. 34–38 (2008)
5. Deepika, M., Kalaiselvi, K.: A Empirical study on disease diagnosis using data mining techniques. In: Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 615–620. IEEE (2018)
6. Gnaneswar, B., Jebarani, M.E.: A review on prediction and diagnosis of heart failure. In: Innovations in Information, Embedded & Communication Systems (ICIIECS), International Conference, pp.1–3. IEEE (2017)
7. Karimifard, S., Ahmadian, A., Khoshnevisan, M., Nambakhsh, M.S.: Morphological heart arrhythmia detection using hermitian basis functions and KNN classifier. In: 2006 International

Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1367–1370. IEEE (2006).
8. Jabbar, M.A., Deekshatulu, B.L., Chandra, P.: Classification of heart disease using artificial neural network and feature subset selection. Global J. Comput. Sci. Technol. Neural Artif. Intell. **13**(3) (2013).
9. Saglain, M., Hussain, W., Saqib, N.A., Khan, M.A.: Identification of heart failure by using unstructured data of cardiac patients. In: 2016 45th International Conference on Parallel Processing Workshops (ICPPW), pp. 426–43. IEEE (2016)
10. Avci, E.: A new intelligent diagnosis system for the heart valve diseases by using genetic-SVM classifier. Expert Syst. Appl. **36**(7), 10618–10626 (2009)
11. Mane, T.U.: Smart heart disease prediction system using Improved K-means and ID3 on big data. In: 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI), pp. 239–245. IEEE (2017)
12. Maglogiannis, I., Loukis, E., Zafiropoulos, E., Stasis, A.: Support vectors machine-based identification of heart valve diseases using heart sounds. Comput. Methods Programs Biomed. **95**(1), 47–61 (2009)

IoT-Based Home Security System with Wireless Communication

Juhi Jha, Praharsh R. Dubey, Pratichi Pradhan, and Smitha N. Pai

Abstract The rise of Internet of Things (IoT) technology has revolutionized the world, and with this growth, additional features have being incorporated into the digital device. Security of individual and property is of utmost prominence, and robust means of handling is possible because of IoT. The approach to having access, automate, control, and provide alert has taken a new dimension to provide security to property and human being. A security system which encompasses various communication techniques has been developed to provide a safe environment for person at home and an alert system to others. Controlled access to premises is possible in unmanned area with the current system. Various ways to alert the person who is not a home premise or inside the house are possible. This system provides robustness as compared to some of the existing system.

Keywords Home security · IoT · Wi-Fi · Wireless communication · ZigBee

J. Jha · P. R. Dubey · P. Pradhan
Department of Electronics and Telecommunication, Manipal University Jaipur, Jaipur, Rajasthan, India
e-mail: juhijha365@gmail.com

P. R. Dubey
e-mail: praharsh.dubey05@gmail.com

P. Pradhan
e-mail: pradhanpratichi@gmail.com

S. N. Pai (✉)
Department of ICT, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India
e-mail: smitha.pai@manipal.edu

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_48

1 Introduction

A security detects intrusion and unauthorized entry and allows access to only the genuine person. Security alarms are used in residential area for protection against burglary, body or property harm, or damage. Security systems for homes are becoming more popular day the day, considering the alarming rise in the number of crimes. However, it is difficult to find a system that has all the necessary security features and not cost a fortune. IoT plays the part of the modern-day savior.

Hardware, software, and sensing devices comprise the parts of Internet of Things (IoT). It collects data and sends it to the processing unit. These data obtained are used to control and monitor the infrastructure. A bulk of process is done with the help of sensors in IoT. Sensors measure data in the raw form and convert into the digital form and send it to the control center. The architecture is chosen based on the context of the operation.

The current project incorporates the IoT features to develop a cheap, durable, and efficient home security system. It uses all the essential features such as an email alert, door control, image capturing, and buzzer.

2 Related Work

Various literatures speak in depth about the home security aspect. Ref. [1] discusses in length the various components that are used to sense, detect, assess, and respond to the threats that are possible in the home settings. The authors in ref. [2] have suggested various home automation products with their merits and demerits. Comparison of various controllers with their strengths and weakness is addressed. Literature [3] works from a different perspective of energy efficient data encryption method for security to data. In the literature [4], a framework is developed which provides data integrity, authentication, and availability of information to the user. In the work [5], the author emphasizes on the characteristic of users to design the smart home. Stress is also imposed on the importance of reliability and interconnectivity among devices. In general terms, the author [6] conveys how communication technology is selectively used in the case of smart homes. The system architecture for the smart home is discussed [7]. It also discussed the design of the electric service system. The author proposes [8] an edge-of-things to manage the house at the edge by a network operator. This is same as managing the set up box and streaming of multimedia instead of sending the data to the cloud. Currently, a modest prototype is carried out to implement the door opening and alert system using various network communication technologies and various means of communication.

3 Methodology

The presented home security system uses the technology of Internet of Things for controlling, monitoring, providing, and enhancing security at home from any remote location. Implementation of a low-cost, flexible, home security system is offered. It enhances the use of wireless communication which provides the user with remote control of their home's security.

The hardware along with software details and steps to carry out the project is as explained below.

3.1 Hardware

Arduino Uno: Fourteen of the existing pins are used for digital input/output. Among this, the microcontroller uses six pins for output purpose. There is a ceramic resonator of 16 MHz. It also has provision for USB connection. The other feature provided is for accepting analog input and adding a power jack. There is a scope for reset. It has a built-in voltage regulation between 5 and 3.3 v with direct power from USB port or a 12 v external power.

XBee–ZigBee Pro S1 Module: It uses a 2.4 GHz ZigBee module with 802.15.4 stack, along with serial command set. It offers reliable communication using the serial port between the microcontrollers, computers and any other system. It supports point-to-point and multipoint network. It works on 3.3 V at 50 mA offering with a maximum offered data rate of 250 Kbps, and the range of communication is 100 m. It has a wire antenna for line-of-sight communication. Six ADC inputs each of ten bits and eight digital IO pins are supported. It works on both AT and API command sets. This module is excellent for point-to-point as well as multipoint communication. It is easily configurable to a mesh network point. Figure 1a shows

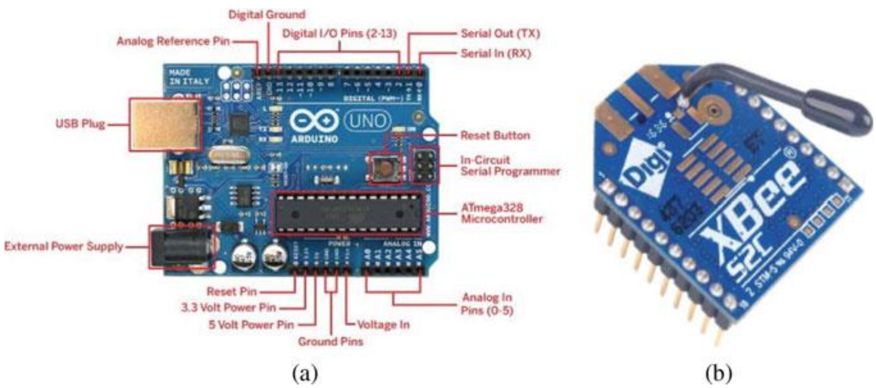


Fig. 1 **a** Arduino Uno board [9]. **b** XBee module [10]

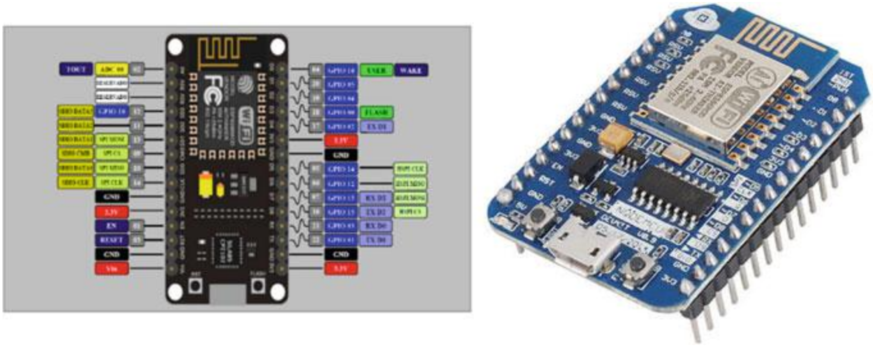


Fig. 2 NodeMCU Wi-Fi module [9]

the layout of the Arduino board, and Fig. 1b shows the XBee module used for short-range communication. As ZigBee modules consume less power, it is used at location where the sensing of data is taking place continuously. The module can be put to sleep when not in use.

NodeMCU (Wi-Fi Module): The module runs on the ESP8266 Wi-Fi with System on Chip (SoC). The hardware for this is based on the module ESP-12.

Figure 2 shows the ESP Wi-Fi system module with its pin diagram. This module is placed inside the house where the power supply is available on continuous basis.

3.2 Softwares

The software utilized includes the Arduino IDE, ESPlorer, ESP Flasher, and XCTU.

3.3 Setup

Two Arduino boards are connected using XBee modules to connect wirelessly between them. One of the boards is the master (sender), and the other one is the slave (receiver). A proximity sensor connected to the master Arduino acts as a source for data collection. It is also responsible for alerting when an object approaches the gate. Here object could be a vehicle or a person.

A NodeMCU (Wi-Fi module) is also connected to the master Arduino in order to send long-distance alerts. The slave Arduino has an LED connected to it and acts as an indicator when an object is approaching the premises.

3.4 Steps to Carry Out the Work

1. The proximity sensor senses the approaching object, and this information is transferred to the Arduino board. The serial interface port displays the message "Obstacle exists" if the object exists or approaches the proximity sensor. When there is no object a message, "No obstacle" is displayed.
2. Arduino with the proximity sensor is configured as the sender and the other Arduino board as receiver with XBee boards.
3. XBee boards are preconfigured using the XCTU software with sender acting as a "Coordinator" and the receiver as the "End device."
4. The sender Arduino also has the NodeMCU board and provides correct settings using the ESP Flasher software. The NodeMCU is connected to a Wi-Fi network.
5. The NodeMCU and the sender Arduino are connected such that when the sensor is activated by an approaching object, an email is sent to a predesignated registered email id. The receiver of the email alert now knows that someone or some object is approaching their door and can respond accordingly. Anybody inside the house is also indicated about the approaching person by the LED and a buzzer alarm.
6. While the email is being sent, the sender Arduino uses XBee and sends the sensor readings to the receiver Arduino as well. The response of the email is evaluated, and the LED is lit depending on the response. This process can be replaced by opening the door on recognizing the arriving person.

The process is demonstrated with the flow diagram as in Fig. 3.

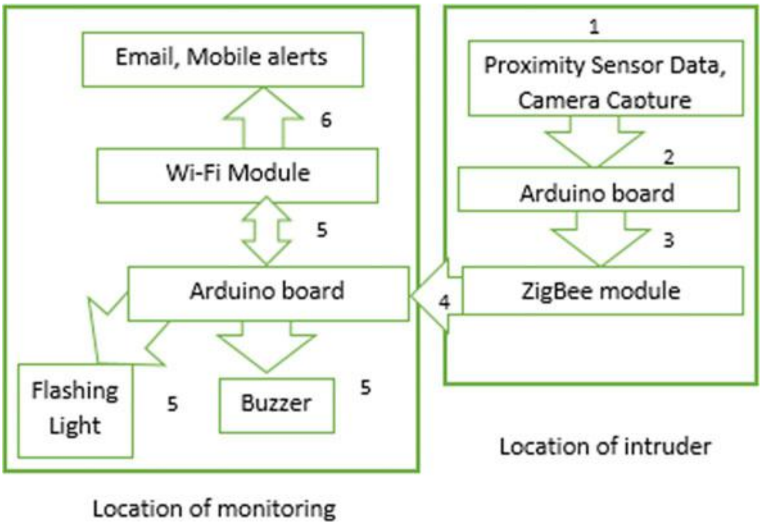


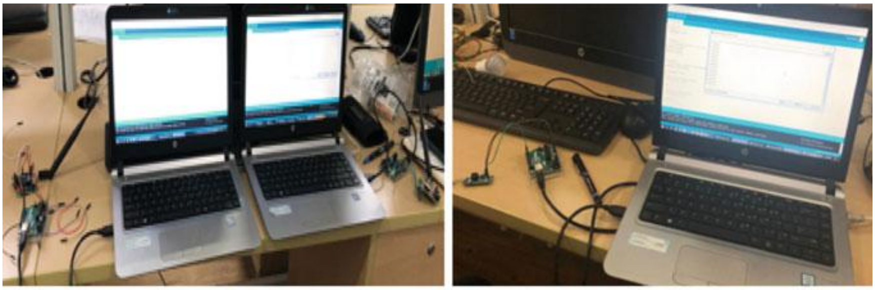
Fig. 3 Process flow

4 Results and Discussions

The screenshots show various steps carried out in implementing the project. Figures 4, 5, and 6 show the various steps of implementation.

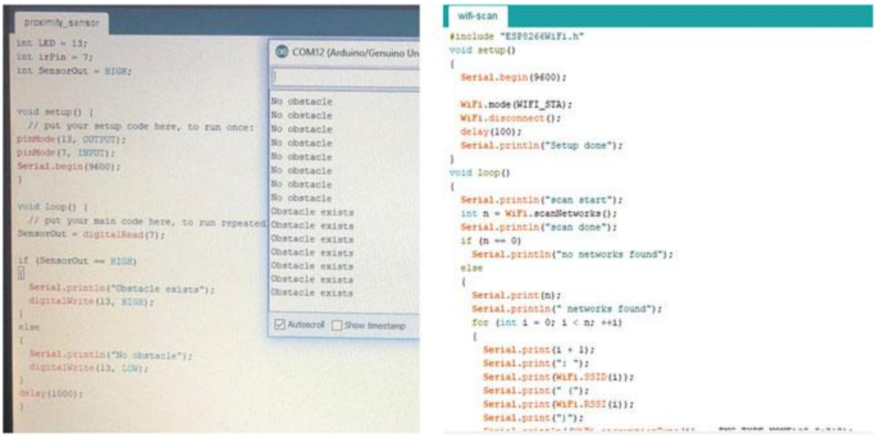
The door of the premises has the proximity sensor with an Arduino board. The object approaching information is passed on to the other board placed within the house. In order to send the email, a Wi-Fi module is used. Email is used to warn about intruder in areas where mobile access is not provided for safety reasons. Figure 7 shows that when an object (as shown by the finger tip) approaches the house premises then an ALARM is heard and LED light glows.

Next part of the work includes taking the photograph of the object or person entering the premises. This can be achieved using the Raspberry Pi which has an option for camera interface. The face detection and recognition algorithm using OpenCV



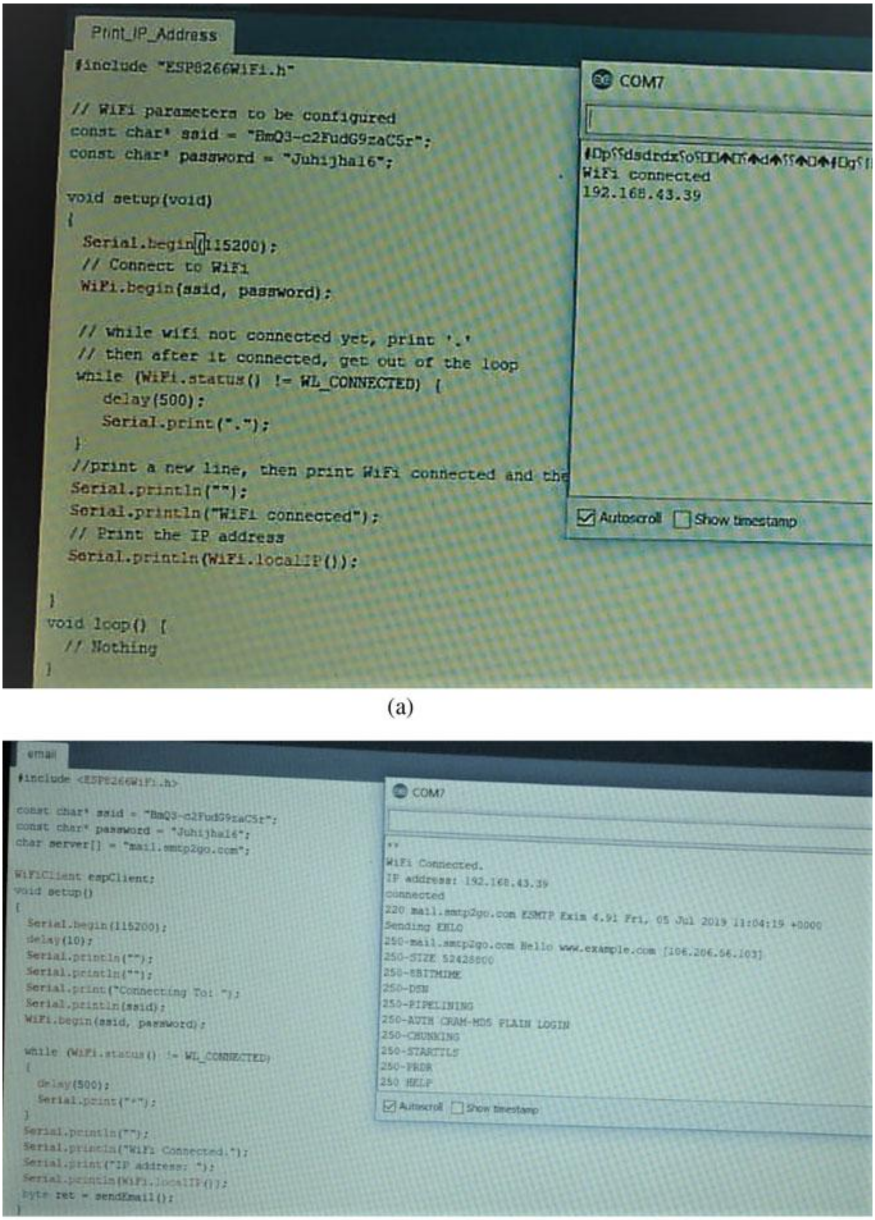
(a) (b)

Fig. 4 **a** Arduino boards connected to XBee boards. **b** Proximity sensor with the sender Arduino board



(a) (b)

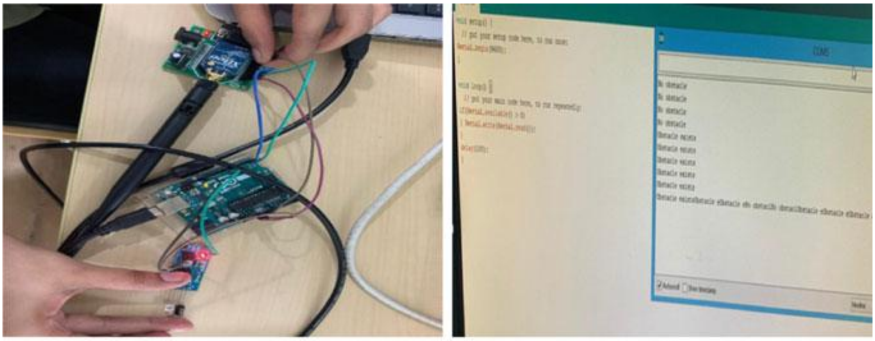
Fig. 5 **a** Warning about approaching object. **b** Scanning nearby Wi-Fi networks using NodeMCU



(a)

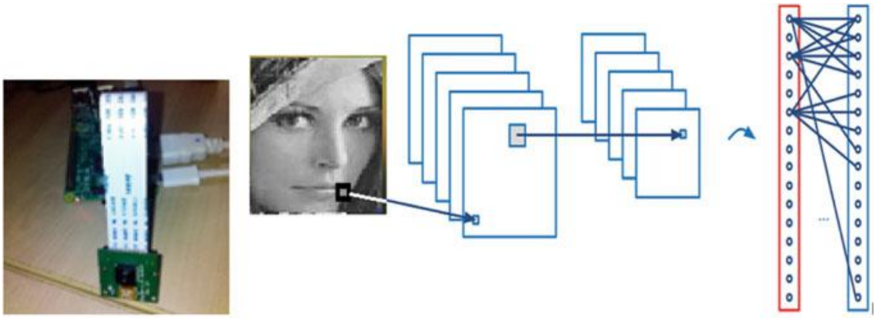
(b)

Fig. 6 **a** Connecting to the Wi-Fi networks and getting its IP address. **b** Sending email alert



(a) (b)

Fig. 7 **a** Sending and receiving sensor data across the two systems. **b** Receiver's serial monitor output



(a) (b)

Fig. 8 **a** Camera photo capture using Raspberry Pi. **b** Convolution neural network with max pool and vectorization for face recognition

and python can identify the person. Currently, the images of the face are stored in a file for face recognition. The interface between the image captured and stored is yet to be processed. The face recognition is carried out using the deep neural network concepts. Few known faces are stored in the repository. This helps in identifying the person. Figure 8 shows a representation of CNN used in the recognition of human face.

5 Conclusion

The current paper shows a modest implementation of a home security system. ZigBee modules (which normally run using batteries) consume less power and are used as a means of communication at location where the proximity sensor is continuously sensing and out of reach of power supply. Wi-Fi is placed inside the house where there is a scope to connect to power supply. Wi-Fi module consumes relatively high power as compared to ZigBee. The Arduino senses the approaching module and using the ZigBee module sends the information to the module placed in-house. For the person at home, an alert about an approaching object is obtained in the form of LED glow and also a ringing buzzer. An email alert is also sent to the registered user about the approaching object. This alert can also be accompanied by a picture of the person. Currently, the image of the object is captured and stored. Image capture and identification is carried out as another project which is yet to be integrated with the current project. Deep neural network is used for identification of face. The project can be enhanced by not setting alert for every object approaching the gate, but only for person or vehicle approaching the gate. Only one single instance of the frame needs to be sent. Automated gate opening from person who has access right to home can also be implemented.

The project provides a secure environment for individual who are at home far away from the entrance. The buzzer provides alert to the person who is blind and flashing light for the deaf. If there is a disabled individual at home, an email alert to the registered ID or mobile phone is an added advantage. Capturing photograph and sending the identity of the person will be an added advantage after integrating the module.

References

1. Dahmen, J., Cook, D.J., Wang, X., et al.: Smart secure homes: a survey of smart home technologies that sense, assess, and respond to security threats. J. Reliable Intell. Environ. **3**(2), 83–98 (2017)
2. Lobaccaro, G., Carlucci, S., Löfström E.: A review of systems and technologies for smart homes and smart grids. J. Energ. 1–33 (2016)
3. Pirbhulal, S., et al.: A novel secure IoT-based smart home automation system using a wireless sensor network. J. Sens. **17**(1) (2017)
4. Kang, W.M., Moon, Y.S., Park, J.H.: Human-centric Comput. Inf. Sci. **7**(6) (2017)
5. Yang, H., Lee, W., Lee, H.: IoT smart home adoption: the importance of proper level automation. J. Sens. 1–11 (2018)
6. Sarah, J.: Darby smart technology in the home: time for more clarity. Build. Res. Inf. **46**(1), 140–147 (2018)
7. Li, M., Gu, W., Chen, W., et al.: Smart home: architecture, technologies and systems. In: 8th International Congress of Information and Communication Technology (ICICT-2018) Science Direct Procedia Computer Science, vol. 131, pp. 393–400 (2018)
8. Batalla, J.M., Gonciarz, F.: Deployment of smart home management system at the edge: mechanisms and protocols. Neural Comput. Appl. **31**, 1301–1315 (2019)
9. Arduino UNO Arduino UNO R3, Pin Diagram, Specification and Applications. https://www.elprocus.com/what-is-arduino-uno-r3-pin-diagram-specification-and-applications/
10. Xbee S2C Module Pinout, Features and Datasheet-Components 101. https://components101.com/wireless/xbee-s2c-module-pinout-datasheet

Implementing Internet of Things in Small Businesses of Bahrain

Abdulla Abdulla, Ali Albash, Hassan Ashoor, Allam Hamdan, Anjum Razzaque, Abdulmuttaleb Musleh, and Muneer Al Mubarak

Abstract Internet of things (IoT) is one of the most advanced technologies that gradually occurs in organizations through a platform for exchanging information from one point to another, using interconnected devices. Almost all sorts of firms use of IoT. Although, most of the small firms still fall shy of implementing due to ample reasons. This study explores ways it is vital to implement IoT in Bahraini firms where most of the business is affected thru the implementation of IoT. This study focuses on factors that aid in implementing IoT in small businesses, by identifying how IoT technology is implemented in small businesses of Bahrain.

Keywords IoT · SMEs · Bahrain

1 Introduction

Internet of things (IoT) is a network of connected systems or devices that are embedded with sensors and electronic network that enables a platform to exchange data from one point to the other. IoT is one of the most advanced technologies that can be used in business as it helps in improving the overall performance of the business while reducing the possibilities of human error [4]. Many companies, small or large, are adopting the use of IoT technology within their business systems to handle the complex situations in a lucid way. Implementation of IoT is the next step forward for most of the businesses and consists of complex processes [7]. Among all the IoT developed countries Bahrain is the finest example where IoT devices are growing in a rapid manner using IoT-enabled devices in their concept of business system [10]. Thus, it is evident that the use of IoT technology is gradually taking up the business sectors in a rapid way.

The following research is carried out to explain the impact of IoT technology in small business determining its positive or negative impact and providing ways by which this technology can be implemented within the business sector. Various

A. Abdulla · A. Albash · H. Ashoor · A. Hamdan (✉) · A. Razzaque · A. Musleh · M. Al Mubarak
MIS Department, Globalits, Manama, Bahrain
e-mail: allamh3@hotmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_49

scholars shared ideas pertaining to the impact of IoT and their various implementations on small business sectors. The implementation of IoT in small business helps them in evolving in the global marketplace and stand out within the competition between the small and large business sectors [17]. Various scholars reflected their views on the adaption of IoT in Bahrain. Several IoT projects have contributed to the development of the country as reflected in many case studies while other reflected disadvantages when applying IoT in Bahrain. Therefore, this research identifies the ways for implementing IoT in the small businesses of Bahrain. This study also aims to comprehend the core concept of IoT and its implementation in the case analysis of Bahrain. This would help understand the aspects of the business sector and how it affects other business sectors. The main objectives of this study are to discover how IoT benefits small businesses of Bahrain when implementing IoT, and pinpoint strategies for improving the implementation of IoT within the small businesses.

Bahrain is one of the smaller economies in the GCC, but it is one of the more diversified, with its well-developed financial services and manufacturing sectors. Despite its diversified economy, Bahrain faces pressures in recent years as a result of the 2014–2015 drop in oil prices. However, the 2018 aid package provided by the Gulf Corporation Council (GCC) countries, with an accompanying fiscal-adjustment plan, underpinned by the support of the dwindling oil price followed by the recent oil and gas discovery within Bahrain; offers hope for some form of a turnaround by 2019. The interest of Bahrain in small and microbusinesses was inspired from the economic policy aimed at diversifying the sources of income, reducing dependence on oil as the main source of income, and increasing foreign economic, service and production within the country. The problem is to realize the various reasons behind which small business organizations run behind to adopt ways to implement IoT in their business processes with their special implication on the financial or structural disabilities that they face while implementing IoT. It is often observed that IoT technology is adopted by mostly large business organizations as compared to small enterprises [9]. The main reason that contributes to this factor is the financial disadvantages that the small companies face while implementing IoT within their business process. Various reasons other than financial factor contribute to this cause. The lack of an appropriate infrastructure within small businesses, along with the absence of the appropriate strategies, limits small businesses to shy away from implementing IoT. For any business, budget should be suitably planned for a successful implementation of IoT for successful business operations. For small businesses, a financial problem along with lack of an appropriate infrastructure was identified as the key reason for lack of an implementation of IoT within business operations. Barriers exist within and outside small businesses which also limit the implementation of IoT.

The following sections include literature review, research methodology, data analysis, findings, discussions, conclusion and recommendation.

2 Literature Review

2.1 IoT Implementation in Various Sectors of Bahrain

IoT connects computer networks, sensors and devices with things via the Internet. IoT can increase the performance of the business and in Bahrain; most of the companies are using this technology for reducing human errors and improving the effectiveness of their services. The major focus of this literature review is to examine the importance of the IoT in Bahrain. According to Cedeño et al. [2] IoT has the capability for controlling and managing issues and problems faced by consumers in the society and home automation is the best application of this technique where users may control devices and systems with the help of sensors. It allows consumers to reduce air pollution and manage their life in an effective way. In Bahrain, many users and firms use IoT to access computer networks from anywhere while effectively handling complex situations. Recently, Dijkman et al. [3] suggested that in Bahrain there are several applications of IoT, e.g., smart cities, home automation, cleaner air and water, smart agriculture and cutting food waste; where IoT-enabled devices manage the responsibilities of consumers in an appropriate manner and reduce issues and problems faced by consumers. In Bahrain, more than 56% of waste management companies use IoT. Also, traffic monitoring is a major pilot of any smart city where IoT devices automatically manage traffic, handle injuries and accidents occurred in Bahrain [16]

2.2 Impact of IoT on the Entrepreneurship of Bahrain

According to Fleisch et al. [4] IoT is crucial in entrepreneurship for connecting multiple devices without wired connections so the entrepreneur can comfortably communicate with employees. This makes IoT a revolution that allows entrepreneurs to work effectively with reduced issues and problems in workplaces. Enterprises require an effective technology that helps them reduce communication gaps and enhance the productivity of the business. IoT security solutions assist companies to effectively communicate with their stakeholders and consumers and manage the data analysis process. IoT hubs efficiently interconnect and manage all IoT assets to allow employees and to communicate with all their things. The IoT links various computing devices on the Internet and a platform for sending and receiving data over a network. According to Ju et al. [5], IoT is expected to be a major part of the business to facilitates managers and improve product and service performances. As per the scholars of this study, it was observed by them that IoT has the potential to enhance the efficiency of computer devices and networks. IoT-enabled devices to offer new business opportunities in various sectors such as health care, agriculture and communication. In the area of health care, it controls and manages the medical devices and systems and support nurses to check the health of patients in an effective manner [13]. For instance, in Bahrain, more than 55% of hospitals and health care communities use IoT devices for accessing real-time information and reducing the communication gap between physicians and patients: e.g., evaluating and integrating data from patient's system provides doctors to observe problems earlier and design effective treatment options for patients. Moreover, it also helps companies to maintain IT infrastructures and reduce problems and issues that impact the performance of their services and businesses. According to Khan et al. [6], the use of IoT-enabled devices is increasing very fast and FinTech is the best application of this technology in which it controls and manages activities and IT systems [12]. IoT is capable of improving communication. It globally contributes about $2 trillion worth of benefits and financial technology. The main impact of IoT on FinTech is when it evaluates all the processes of financial technologies and improves the business performance. Moreover, the IoT supports financial technology to work effectively for effective quality decision-making [12].

2.3 Impact of IoT on the International Trade Finance Decision

As per Krotov [7] international trade finance decisions and consumers authentication can be implemented with the help of IoT technology in order to improve the security of transactions. It is estimated that the use of IoT will be increased by 67% in FinTech by improving operations and services more effectively. There are ample gains of IoT in FinTech, for example, enhance consumer services, reduce risk and issues occurred in the workplace, better decision-making and enhance security and privacy of personal details and many more. Retail banking is the best example where companies can use the concept of IoT in order to monitor customer services and internal operation in an effective manner. Therefore, it is stated that the IoT uses the wireless networks and sensors in order to connect two or more devices that can help the financial industry to easily link their networks with each other and enhance their productivity. According to Tawfiqi et al. [15], Bahrain heads for the entrepreneurship initiative, where numbers of companies are moving toward innovation and government of Bahrain is looking to make Bahrain a Middle Eastern for startups. It is observed that in order to push employees toward entrepreneurship Bahrain has developed Economic Development Board that focuses on the startups in the Middle East. It has found that in Bahrain the number of startups has increased by 46% CAGR in the last three years. Mainly, Bahrain focuses on software development, health care, financial technology and cyber-security that helps companies for improving their businesses and making entrepreneurs. IoT is a network of systems, devices and things which are embedded with sensors and electronic networks and provide a platform to exchange data from one location to another. It is one of the best technologies from the perspective of business because it can improve the performance of devices and systems used in the workplace and reduce human errors

[1]. It is now used in many companies to control and handle more complex situations in an effective manner. Fleisch et al. [4] highlighted that IoT plays a crucial part in entrepreneurship where it provides a way for connecting two or more devices with each other without wired connections so entrepreneurs easily communicate with employees. IoT is a revolution which is allowing entrepreneurs to work effectively and reducing issues and problems occurred in the workplace. Enterprises require an effective technology that helps them reduce communication gaps and enhance business productivity. IoT security solutions help companies for effectively communicating with their stakeholders and consumers and manage the data analysis process. Ju et al. [5] stated that IoT helps the business improve products and services. IoT is a kind of network which sends and receives information from one location to another by using wireless connections. Within the context of entrepreneurship, IoT is more effective that has the potential to enhance the performance of employees in the workplace. Moreover, the IoT provides a way to collect and evaluate data of the consumer in an effective way which aids entrepreneurship gather accurate information from computing networks. Palattella et al. [11] identified that IoT can provide a cost-saving approach to the management team where employees can easily control and manage work performance with the help of wireless monitoring systems. Moreover, IoT-based device allows companies to make smarter and faster decisions that reduce errors and problems faced by consumers in the workplace. IoT develops a platform for entrepreneurs to control output based on the monitoring functions and systems. There are numerous business communities where IoT-enabled devices effectively impact the entrepreneurs, e.g., health care, energy, retail industry and transportation. Data analysis is a major part of the IoT technology which helps entrepreneurs obtain and evaluate business data. It was revealed that more than 55% of Bahrain consumers use IoT devices in daily lives for controlling and managing works, and more than 45% of companies use IoT systems.

However, home automation is the best application which can increase the use of IoT in Bahrain and it can help societies for improving their skills and users may connect two or more devices with the physical things and perform communication between computer networks. According to Sadeghi et al. [14], Bahrain has the potential to design and implement IoT-based systems and devices because there are numbers of companies in Bahrain uses IT systems and networks which can be handled by using this technology. In the Middle East and Bahrain, the use of information technology is increasing very fast, and IoT is more effective technology that can help Bahrain to develop smart cities and workplaces. A smart city is the best application of IoT due to which Bahrain can adopt this technology and control traffic signals automatically by using wireless networks and IoT devices. Developing and implementing IoT technology in Bahrain can change the way of communication and users may control and manage their work from any location. In transport companies, IoT can be used for developing wireless devices based on computing networks and reduce human efforts and errors in the workplace. Therefore, Bahrain has the capability of implementing IoT devices with less cost. In Bahrain, IoT can extend Internet connections and produce a new environment where consumers can send and receive data from any location and control physical devices. Moreover, IoT connects two or more devices with each other and reduces the communication gap between consumers.

2.4 Research Gap

Small and microbusinesses pressure huge economic and social impacts on the economic strategies of countries, decision makers, financial experts and the business sector representatives. They are solutions to weak economies and social activity and they contribute to improving the economy by enhancing the standard of societal living. They play a leading role in reducing unemployment concerns of most countries. Entering the world of finance and business is a suitable option for the youth as it solves the problem of unemployment while helping the economy grows and prospers. The development of exports attracts more businesses and makes the country better. Such an institution is flexible and can quickly adapt to economic changes. It is also capable of facing crises because they are pivoted with specifications that are not available to other large enterprises; for instance, they are more capable of switching from one production sector to another that meets the requirements of the market with ease of entry and exit with very minimal losses. It was revealed that different authors have different point of views about the implementation of IoT in small business sectors of Bahrain with respect to their impacts on the specific fields. By combining the literature and critiquing ideas on this topic, a major research gap was identified, i.e., that most of the IoT research cites the IoT implementation effect within small businesses of Bahrain. However, none of the past literature has yet focused on how IoT aids small business and microbusiness. Furthermore, no previous studies mentioned how the social Internet can be utilized to support things within small businesses to implement IoT. Thus, there remains insufficient evidence of any holistic view pertaining to the implementation of IoT within small businesses. Also, there seems a scant proposed structure effective for eliminating/reducing the problems that small business face when implementing IoT.

3 Research Methodology

There are various applicable methods on the implementation of IoT in small businesses. It is imperative to identify the appropriate research methodology that goes exactly with the study objectives. Identifying the proper methodology for the research is an important part of any research as it helps scholars to conduct further investigations on the topic and discovers more details on the topic [8]. Using an appropriate methodology helps eliminating the possible errors that are present within the study by human or any other causes. This study focuses mainly on identifying the appropriate reason behind each rationale that has been formed to choose techniques in order to evaluate the impact of IoT implementation on small business sectors in Bahrain.

3.1 Population and Sampling Size

Quantitative method is applied in this study's deductive approach to gather literature pertaining to the usage of IoT in small businesses. The Instagram measured the business performance. An indicator is used through the Instagram pages to attain data on financial statement for certain business. This type of information is not published publicly. For this research, data was gathered on unknown SMEs population of companies in Bahrain. A sample of 123 was chosen on the availability of data. According to the Ministry of Commerce and Industry [10], the concept of small business refers to the type of institutions that varies from state to state. The concept of small business varies according to the economic sectors in which it operates based on the economic, industrial and social development state of the business as experienced by this country or with the difference in the definition of specific institutions based on the. The criteria that are adopted are as follows: Small enterprises are defined as the type of enterprises consisting of several workers ranging from 50 to 11 workers. The invested range of capital is from BD 20,000 to BD 50,000 having an annual turnover of BD 100,000–BD 1,000,000 Million. For this research, data is gathered on unknown population of companies in Bahrain.

3.2 Measuring Variables

The dependent variable is the performance, measured through proxy as cannot be obtained directly such as the net profit of business. The independent variable is the food delivering system for Bahrain country is considered. The Talabat and Hunger station including the logistic delivery where IoT is used as a part of the project. As it is one of the embedded devices, hence, things remain embedded with sensors and electronic networks providing a platform to exchange data from one location to another. The dummy variable number 1 is used which means that the business is using IoT and as 0 meaning there is no use of IoT.

4 Data Analysis, Findings and Discussions

4.1 Descriptive Analysis

Descriptive statistics were employed to offer a clear demonstration of the target population of this study. Several small firms were not using IoT, determined to be 42, and the organizations which use IoT within their operations are 81 out of 123 (sample size of this study). Thus, for the small and medium-size enterprises that use of IoT is 66%; and who do not use IoT is 34%.

Table 1 Using IoT by SMEs in Bahrain

	Frequency'0		Frequency'1		Total	
	No.	%	No.	%	No.	%
Using IoT by SMEs	42	34	81	66	123	100

Table 2 Descriptive analysis

Variable	Descriptive Statistics			
	Mean	Maximum	Minimum	Std. deviation
Number of posts	871.93	11,179	9	1,616
Number of followers	14,650	203,000	156	27,618
Average of comments	44.951	902	0	157

The descriptive statistics (Table 2) were acquired when conducting descriptive analysis depicting that the maximum number posts that are made by the organization with the use of IoT tend to be about 11,179 with the minimum number of posts to be 9. Based on these statistics, the mean, that is, obtained as 871.93 and the standard deviation as 1616.448. Similarly, for the number of followers, the maximum size extends to up to 203,000 while the minimum number ranges to up to 156. For this, the obtained mean is 14,650.20 and the calculated standard deviation is 27,618.421.

4.2 Correlations Matrix

As depicted in Table 3 (correlation analysis), the relation between the IoT and small businesses is expressed thru three proxies: number of posts, number of followers and the average number of comments. The analysis is conducted such that the relation between these proxies is determined if measuring between −1 and +1. Pearson

Table 3 Spearman correlations

	Proxy 1	Proxy 2	Proxy 3	Using IoT by SMEs in Bahrain
Performance proxy 1: number of posts	1			
Performance proxy 2: number of followers	0.568**	1		
Performance proxy 3: average of comments	0.008	0.016	1	
Using IoT by SMEs in Bahrain	0.084	0.017	0.142	1

** indicates 1% level

correlation revealed that three proxies are interrelated as the values include 0.084, 0.017 and 0.142 proxies of number of posts, number of followers and the average number of comments, respectively.

Table 3 also showed an insignificant relationship between using IoT by SMEs in Bahrain and SMEs performance proxies. Therefore, not supporting this study's hypothesis which claims a positive and statistically significant relationship between the use of IoT by SMEs in Bahrain and the performance of SMEs. Thus, the above analysis indicates that there exists no strong relation between IoT and small business sectors: possibly since the cost of implementation: among the several reasons there can be a certain reason for which IoT is not implemented within small businesses. The cost of implemented IoT can be one of the factors for which SMEs steps back toward implementing IoT. The implementation of IoT requires high-cost implementation and hence leads to expenses without revenue. Security issues: With the implementation of IoT there exist several security issues with that of the connected devices across several organizations. The reason for which small businesses steps back toward the implementation of IoT as they lack the infrastructure to provide security to the cyber issues that are faced with the implementation of IoT. Unaware of the benefits of IoT: There are still various organizations which remain unaware of the benefits of implementing IoT within their business processes and hence lack the capability of implementing IoT within their business process.

5 Conclusion and Recommendation

Two research questions are employed to identify the ways by which IoT can benefit the small business sectors. In literature review, it was identified various benefits that Bahrain has obtained while implementing IoT within their business sectors. The researchers have used various opportunities like posts taken from Instagram profile of the small enterprises, their followers, average rate of comments and the logistic food delivery system. In order to obtain the answer to the second objective, the authors have developed significant questions based on the ways by which Bahrain can implement IoT within the business sectors. Linking computing devices, implementing cost-saving approach to control and manage work, implementing wireless monitoring systems and enabling home automation systems are ways to implement IoT within small and medium enterprise systems. In order to obtain the answer for the third objective, which is based on identifying the strategies so as to improve the further implementation of IoT within small business sectors, the literature review has provided various strategies such as implementing effective strategy to change the way of communication so as to reduce the communication gap, implementing strategy to improve the skills required to implement IoT within business organizations. The utilization of the implementation of IoT within business organizations benefited many sectors in Bahrain such as hospitality, and banking which ultimately led to improve the whole economy in the country.

Based on the analysis of the information gathered during the study, the analyst also developed various recommendations for eliminating the barriers that are often faced while implementing IoT within business sectors, and especially within the SMEs sectors. Thinking about staffing: as the Internet of things is a disruptive technology, hence it needs more efficient staff to get the desired outcome from business sectors. Taking security seriously: The most important concern in case of IoT technology is security that most of the large business sectors are aware of. But it is a serious concern for most of the small enterprises as it tends to attract individuals to make it riskier. IoT platforms within business organizations also attract black hat people to launch DDoS attacks. Implementing effective strategies for better project management, better scheduling of services and many more.

References

1. Ahmad, A.M.B., Masri, R.B., Lee, C.A.: Future challenges to the entrepreneur in relation to the changing business environment. Glob. Bus. Manag. Res. **11**(2), 12–18 (2019)
2. Cedeño, J.V., Papinniemi, J., Hannola, L., Donoghue, I.: Developing smart services by the internet of things in the manufacturing business. LogForum **14**(1), 14–16 (2018)
3. Dijkman, R.M., Sprenkels, B., Peeters, T., Janssen, A.: Business models for the internet of things. Int. J. Inf. Manage. **35**(6), 672–678 (2015)
4. Fleisch, E., Weinberger, M., Wortmann, F.: Business models and the internet of things. In: Interoperability and Open-Source Solutions for the Internet of Things (pp. 6–10). Springer, Cham (2015)
5. Ju, J., Kim, M.S., Ahn, J.H.: Prototyping business models for IoT service. Procedia Comput. Sci. **91**, 882–890 (2016)
6. Khan, R., Khan, S. U., Zaheer, R., Khan, S.: Future internet: the internet of things architecture, possible applications and key challenges. In: 2012 10th International Conference on Frontiers of Information Technology (pp. 257–260) (2012)
7. Krotov, V.: The internet of things and new business opportunities. Bus. Horiz. **60**(6), 831–841 (2017)
8. Kumar, R.: Research Methodology: A Step-by-Step Guide for Beginners. Sage Publications Limited (2019)
9. Lee, I., Lee, K.: The internet of things (IoT): applications, investments, and challenges for enterprises. Bus. Horiz. **58**(4), 431–440 (2015)
10. Ministry of Commerce and Industry: Small and microenterprises in the Kingdom of Bahrain, Official website, No. 117, Issue 2973 (5), (2010)
11. Palattella M. R. et al.: Internet of Things in the 5G Era: enablers, architecture, and business models. IEEE J. Sel. Areas Commun. **34**(3), 510–527 (2016)
12. Razzaque, A., Hamdan, A.: Peak-end rule promotes social capital for knowledge management in thru social internet of things. In: Hassanien, A.E., Bhatnagar, R., Khalifa, N.E., Taha, M.H. (eds.) Toward Social Internet of Things (SIoT): Enabling Technologies, Architectures and Applications. Springer Nature, Switzerland AG (2020)
13. Razzaque, A., Hamdan, A., Cummings, R.: The propensity to use FinTech: input from bankers in the Kingdom of Bahrain. J. Inform. Knowl. Manage. (2019)
14. Sadeghi, A.R., Wachsmann, C., Waidner, M.: Security and privacy challenges in the industrial internet of things. In: 2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC) (pp. 1–6) (2015)
15. Tawfiqi, D., Hamdan, A., Masfaka, M., Muwafaq, B.: The role of Islamic banks in the entrepreneurial success. Evid. Bahrain **22**(2), 1–22 (2018)

16. Westerlund, M., Leminen, S., Rajahonka, M.: Designing business models for the Internet of Things. Technol. Innov. Manage. Rev. **4**(7), 5–14 (2014)
17. Whitmore, A., Agarwal, A., Da Xu, L.: The Internet of Things—a survey of topics and trends. Inform. Syst. Front. **17**(2), 261–274 (2015)

A Comparative Study of Model-Free Reinforcement Learning Approaches

Anant Moudgalya, Ayman Shafi, and B. Amulya Arun

Abstract This study explores and compares three model-free learning methods, namely, deep Q-networks (DQN), dueling deep Q-networks (DDQN) and state-action-reward-state-action (SARSA), while detailing the mathematical principles behind each method. These methods were chosen as to bring out the contrast between off-policy (DQN) and on-policy (SARSA) learners. The DDQN method was included as it is a modification of DQN. The results of these methods and their performance on the classic problem, CartPole were compared. Post-training, testing results for each of the models were as follows: DQN obtained an average per episode reward of 496.36; its variant and improvement, DDQN obtained a perfect score of 500 and SARSA obtained a score of 438.28. To conclude, the theoretical inferences were decisively reaffirmed with observations based on descriptive plots of training and testing results.

Keywords Reinforcement learning · Q-learning · DQN · DDQN · SARSA

1 Introduction

Reinforcement learning (RL) refers to mapping scenarios to actions, with the purpose of increasing a reward [1]. The RL problem involves a goal-directed agent which interacts with an unknown environment, by choosing actions to be performed on the environment, resulting in a change of state. Each new state has an associated reward, which is used by the agent to make decisions, in order to maximize accrued reward scores. RL is distinctly different from other forms of learning in that the agent starts with little or no awareness of the environment but learns through conversance with it. One of the main challenges faced in RL is maintaining a balance between exploration and exploitation. The agent needs to be able to determine when to exploit what it already knows and when to explore other options to make significantly better action choices in the future [1]. This form of learning based on actions

A. Moudgalya (✉) · A. Shafi · B. A. Arun
PESIT South Campus, Bengaluru, Karnataka, India
e-mail: anant9798@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_50

and rewards/punishments originates from behavioral psychology and its analysis in humans and animals.

Deep learning architecture when combined with principles of RL results in what is called deep reinforcement learning (deep RL). This combines the features of RL with the capability of deep architecture to progressively extract higher-level features from the input data. Recently, Google DeepMind developed the deep Q-network (DQN), a novel deep neural network architecture that has been shown to be proficient in learning (and surpassing) human-level control strategies on a collection of different Atari 2600 games [2].

In this report, attempts have been made to understand how deep RL can be applied and how to overcome the challenges faced. DQN [3] and its variants have been used in an attempt to compare results and performance statistics. Results of these algorithms are used on the environment, "CartPole-v1" of the OpenAI Gym as a common measure of performance. An implementation of DQN is studied, and the architecture is modified to form a dueling DQN (DDQN) [4], then state-action-reward-state-action (SARSA) algorithm is implemented. A comparative study of these architectures is performed and the experimental results are analyzed.

2 Related Work

2.1 Reinforcement Learning

A reinforcement learning (RL) task deals with training an agent in an environment. The agent performs actions to transition between different scenarios of the environment, referred to as states. Actions and yield rewards, which could be positive, negative or zero, the purpose of an agent is to maximize the total reward it accrues over an episode. Hence, the agent is reinforced to perform certain actions by providing it with favorable, mathematically significant rewards, and to avoid other actions by providing unfavorable rewards. This is how an agent learns to develop a policy/strategy. RL has many learning methods, but only model-free learning methods have been explored as they are robust and independent of an environment description.

There are two types of learning within RL based on-policy, an off-policy learner derives the optimal policy irrespective of the agents' actions, provided it explores to a good extent. If the learning agent chooses only the best action learned up to that point, it may not explore to an extent enough to obtain the optimal action on the whole. The policy value is not learned in off-policy learning because exploration steps are included in the policy. On the other hand, in on-policy learning, the policy that the agent was following, specified previously, is learned. Deep RL has been used in the past for surpassing human scores on Atari games, few of them are, Stadie et al. [5]; Mnih et al. [2]; Guo et al. [6]; Bellemare et al. [7]; Nair et al. [8]; Schaul et al. [9] and van Hasselt et al. [4]. This study incorporates knowledge gained from their experiments and puts it to good effect while performing the experiments and inferring from results in the sections that follow.

2.2 Q-Learning

Q-learning is a model-free RL algorithm, meaning, it does not require a model/description of the environment to be applied [10]. The main motive of Q-learning is to help an agent formulate a strategy, which will help determine the most suitable action to be taken from any state. Q-learning as a technique maximizes potential cumulative reward to all future states, from the current state [11]. To better understand, let us consider a state S, an action A and the associated reward for that action as R. The relationship between Q, S, A and R can be denoted as follows:

$$Q : S \times A \rightarrow R \tag{1}$$

Before learning begins, the Q-value is randomly assigned to a fixed, initial value, chosen by the programmer. At timestep t, the agent chooses action a_t from the set of valid actions, $A = \{0, 1 \ldots K\}$. Each action a_t has an associated reward r_t. For every such action performed from state S_t to state S_{t+1}, Q-learning performs an iterative value update using a weighted average of current and old state information, as follows:

$$Q(s_t, a_t) \leftarrow (1-\alpha)Q(s_t, a_t) + \alpha\Big[r_{t+1} + \gamma \max_a Q(S_{t+1}, a)\Big] \tag{2}$$

The learning rate, denoted by α in the given Eq. (2) [1], denotes the rate with which the agent learns new things about the environment. If $\alpha = 0$, then the agent essentially learns nothing, but in practice, α is generally considered to be a constant value, such as 0.1. The discount factor γ is the measure with which the agent prioritizes short- and long-term rewards, i.e., with a smaller discount factor, or with $\gamma = 0$, the agent always takes short-term rewards. Similarly, with a high discount factor, or $\gamma = 1$, the agent will make moves for long-term high rewards. Short-term rewards are when the agent picks the move that provides for instant maximum rewards, whereas long-term rewards are when the agent makes moves that compromise the highest possible reward at current state, so that it can obtain higher rewards at a later state.

Generally, the basic implementation of Q-learning is with a Q-table, which serves as storage for Q-values of state-action pairs. But with an increase in the size of state/action spaces, this approach is expensive and inefficient. Therefore, this overhead can be reduced by combining Q-learning with function approximation. A good function approximator to speed up learning in finite spaced environments is an artificial neural network.

2.3 Deep Q-Networks

Deep Q-networks (DQN) as designed by DeepMind [3] used convolutional neural networks (CNN) to filter out frames of Atari games as game memory and use the consequent neural network as function approximator.

DQNs also use a technique called experience replay, a biologically inspired process that uniformly samples experiences from memory instead of the most recent action and updates their Q-value. Experience replay and memory replay buffer help in maintaining a balance between exploration and exploitation for the agent. Nevertheless, DQNs and experience replay tend to falter in environments with sparse rewards. The solution for this problem was proposed by Nair et al. [12], the depths of which are not explored in this paper.

Since Q-values are being updated in batches from the previous experience, it allows for better efficiency and makes data distribution more uniform. As per the algorithm deep Q-learning with experience replay [3], only the last N state-action tuple values are being stored in the experience replay buffer and sample D number of tuples randomly. The algorithm can be understood in-depth as explained in [3].

2.4 Dueling Deep Q-Networks

It has been observed in DQNs that the Q-value is assigned for particular state-action pair and that it determines the quality of the state and the action that can be taken from that state. In the architecture for dueling deep Q-networks (DDQN), two separate estimators are represented, a state-value function and an action-advantage function [13]. Therefore, for state S and corresponding action a, the computation of Q-value can be divided into two different, value (V) and advantage (A) streams.

$$Q(S, a) = A(S, a) + V(s) \tag{3}$$

The sum of these two functions is what gives the correctly estimated Q-value for that particular state-action pair. Two networks can be used for these two functions and a special aggregation layer that results in the Q-value. By decoupling the estimation, the DDQN can begin to perceive which states are worthwhile without needing to learn the impact of every action at every single state (since its also calculating $V(s)$). In DQNs, the small Q-updates can cause noise and disruption in the order of actions, which can cause the policy to make drastic, inefficient decisions. But since advantage values are being calculated separately, the dueling architecture is robust to such circumstances.

2.5 SARSA

State-action-reward-state-action (SARSA) is an algorithm in RL, extensively used for learning a decision strategy inspired by the mathematics behind Markov chains. [14]. In SARSA, the Q-value is derived from the current game-state (S_1) of the agent, the game-action chosen by the agent (A_1), the reward score (R) associated with that game-action, the new game-state (S_2) entered by the agent after taking the game-action and the next game-action (A_2) chosen by the agent from its new game-state. The acronym for the quintuple, $(S_t, A_t, R_t, S_{t+1}, A_{t+1})$ is SARSA [15]. As opposed to DQNs, SARSA is an on-policy learning algorithm. It is so called because at every game-step, the agent updates the currently followed policy based on its interchange with the game environment for the corresponding action, in that game-step.

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma Q(s^j, a^j) - Q(s, a)] \tag{4}$$

The Q-value for a state-action pair is updated at each step by a value, which corresponds to the resultant error. This is adjusted by using the learning rate α. Each Q-value is a combination of both the probable reward score that will be obtained in the following game-step, for taking the game-action A in the current game-state s and the discounted/adjusted future reward score that it receives from the subsequent state-action pair, as explained in the works of Perez et al. [16].

In SARSA, the Q-value is updated taking into account the action A performed in state S. This is in contrast to Q-learning, where the action with the highest Q-value in the next state S is used to update the Q-value. The basic difference between SARSA and Q-learning is that SARSA agent chooses its action based on the same current policy and updates its corresponding Q-values, whereas the agent in Q-learning chooses the action which gives maximum Q-value for the state.

3 Experiments and Results

3.1 Model Architecture and Result Metrics

The models used in this project were taken from the keras-rl [17] repository on GitHub. The model architecture is identical for all three algorithms barring intricate differences in implementation. On the whole, as an alternative to a Q-table (which maps current state to next state of highest reward), function approximators are used. In this case, the function approximator used is an artificial neural network. Before analyzing the results of the three algorithms, the following are the meanings of a few important terms.

Reward is a term that denotes feedback obtained from the environment. Reward can be positive or negative. Normally, the aim is to maximize the reward obtained throughout the course of an episode.

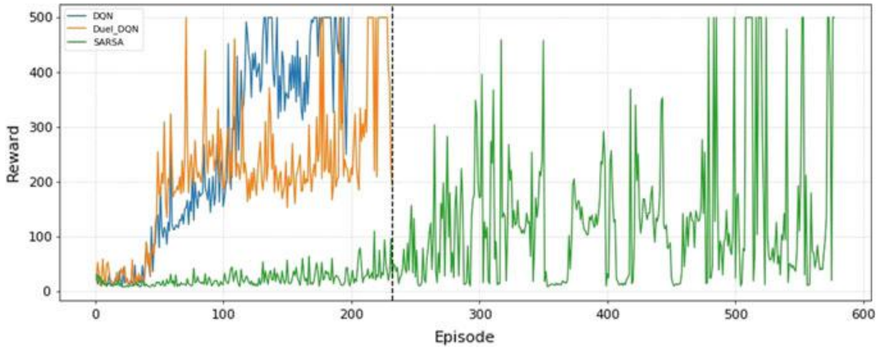


Fig. 1 Rewards per episode across all methods

Mean Q-value refers to the mean of Q-values (denoted by $Q(S, a)$, which is a computational estimate of the anticipated reward assuming that the agent is in state S and performs action a, and then plays until the end of the current episode following some policy).

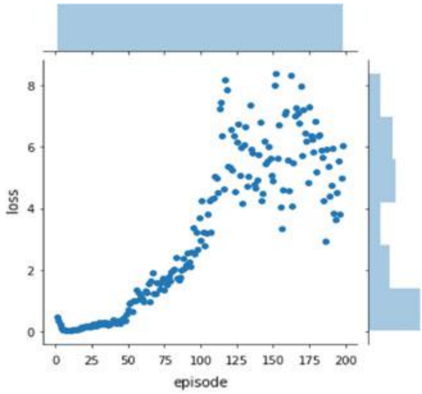
The hyperparameters of the model were set as follows, $\alpha = 0.001$, $\gamma = 0.99$. The model was trained for 50,000 game-steps as a limit, where any number of steps can occur in an episode, from initial state to terminal state. The rewards per episode during training were observed and plotted, as shown in Fig. 1.

From the graph, in Fig. 1, the validity of the mathematical functions discussed in the previous sections can be observed. It can be concluded that DQN takes the least number of episodes to conclude training and reach max reward per episode. Duel DQN takes the second most number of episodes, more than DQN because of the duel estimation architecture. It can also be seen that although Duel DQN takes longer to conclude training, it learns better, i.e., experiences episodes with bigger rewards faster than DQN, and during the end of the training period, rewards are consistently higher. SARSA, on the other hand, being an on-policy learning method, does not prioritize cumulative rewards, and hence takes longer to learn. Even during the end of the training period, rewards per episode are not consistently high.

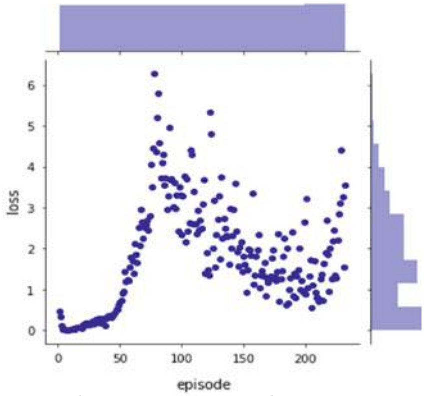
The next result metric in the observations is loss per episode during training. Training loss in most deep learning algorithms can be plotted and inferences can be made to decide upon the quality of the said algorithm. The loss plotted for each algorithm is as shown in Fig. 2.

From the scatter plots, the loss per episode can be observed during training for the first method, a DQN. Training loss is generally high in off-policy RL algorithms as the agent must explore the environment first and exploit its past experiences eventually.

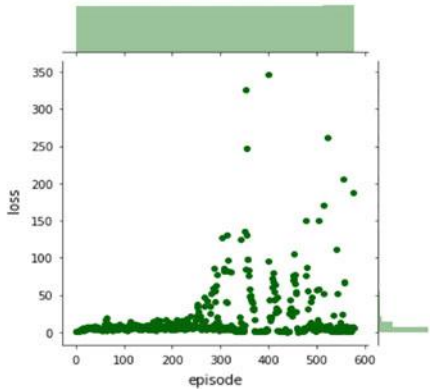
For DQN (Fig. 2a), it can be observed that training loss is high as the agent explores the environment, gradually forming the policy. Also, from the plot, it is concluded that the range of values of training loss is large and that the algorithm incurs heavy losses at some points during training. Once a policy is formed, loss



(a) Loss Per Episode-DQN



(b) Loss per Episode-DDQN



(c) Loss per Episode-SARSA

Fig. 2 Loss per episode graphs

per episode decreases as the policy chooses the optimal *Q*-values and prioritizes cumulative rewards.

Next, the graph for loss per episode for DDQN is observed (Fig. 2b) which is a variant of the DQN. It is already noted that it uses the dueling network architecture and that it takes more number of episodes than DQN during training.

It is observed that the loss pattern for DDQN is similar to that of a DQN, the only difference being the policy is formed faster and there is a sharper decrease in loss after the formation of the aforementioned policy. It is also observed that the range in training loss values is lesser compared to DQN.

The variation in training loss for SARSA can be observed from the plot in (Fig. 2c). The general pattern of training loss is very low this is because SARSA is an on-policy learning method. Therefore, unlike DQNs and DDQNs, SARSA does not incur losses during the policy-formation phase, as it follows a policy from the first stage. Loss incurred may become very high (occurrence of 350 in Fig. 2c) as the agent may

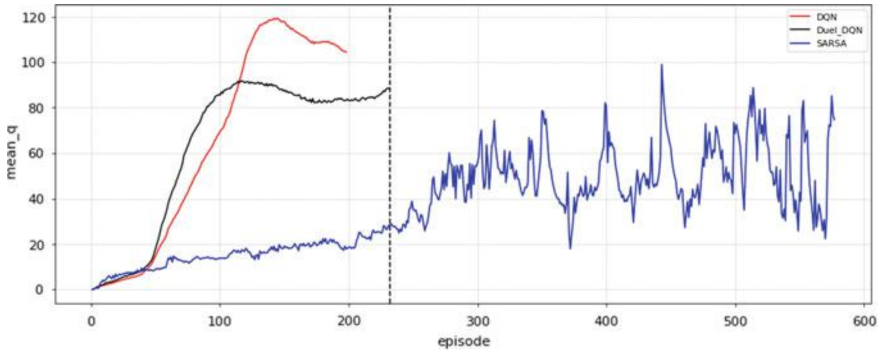


Fig. 3 Mean Q-values across all methods

sometimes not obey the policy in every episode, and therefore suffers great loss for that episode (Fig. 3).

Next, the mean Q-value per episode is observed during training for all three algorithms. It is observed that the behavior of Q-values aligns with the mathematical equations as described above. Hence, from the graph, it is inferred that Q-values for a DQN increases during exploration, as the Q-values for state-action pairs are updated. It is also noted that the Q-value stabilizes, and also decreases after a point, which is at the completion of exploration, as the actions are chosen according to the maximum Q-value for that state in each particular episode.

Duel DQN behavior is also similar to DQN, except it again confirms that the exploration phase is completed quicker as the stabilization on mean Q-value per episode occurs earlier than DQN, and the updates to Q-values are minimal after completion of exploration.

Mean Q-values in SARSA on the other hand are gradually increasing in nature which reaffirms the fact that it is follows on-policy learning. The updates to Q-value follow the previously described policy, and that it updates Q-values based on current reward and does not take cumulative rewards into consideration.

3.2 *Test Results*

After 50,000 timesteps of training, the model is trained on 150 episodes, consisting of 500 timesteps each, on the same environment, i.e., CartPole. The rewards per episode results were plotted in a graph. It can be observed as shown in Fig. 4. Training and testing score averages can be referred to in Table 1. After calculations, it is seen that the average score obtained for DQNs was 496.36, which is close to almost a perfect score on each episode.

The dashed vertical line at the end of the graph is indicative of the dueling DQN's test scores. The average reward per episode obtained was 500.0. Astonishingly,

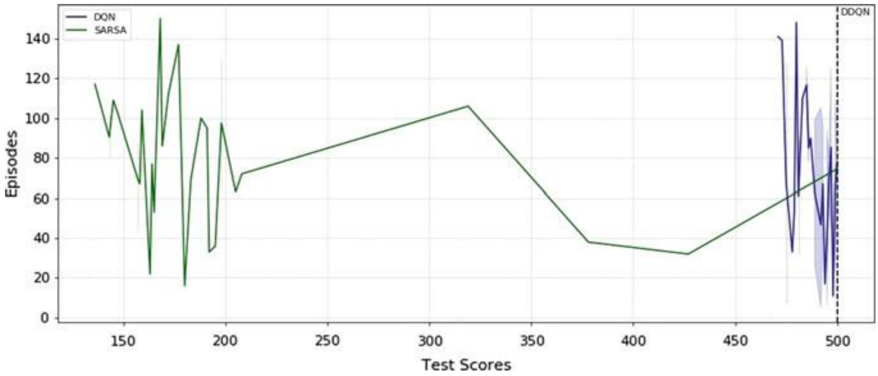


Fig. 4 Test results for DQN, DDQN and SARSA over 150 episodes

Table 1 Result metrics

Algorithm	Reward	Loss	Mean Q	Test reward
DQN	250.08	3.25	64.86	496.36
DDQN	214.31	1.81	63.29	500.0
SARSA	86.56	17.19	35.81	438.28

DDQNs gave a perfect score on all 150 test episodes, conclusively proving that it is invariably an improvement upon the DQN model, beating its near-perfect score to an actual perfect score.

On the contrary, SARSA, being an on-policy learner, produces erratic rewards per episode during testing. Similar to the training phase, the Q-values are updated according to the policy and not in an optimal manner, and cumulative rewards are not prioritized, therefore, a lot of changes in the reward score obtained in each test episode can be observed. Despite this, after 50,000 timesteps of training, the model does produce good results in a number of episodes. The calculated average reward per episode for SARSA is 438.28. While it may seem tawdry in comparison with the DQN and DDQN, it is still a substantially good score per episode.

Therefore, it is observed that the agent does well in all three environments producing a sufficiently good score, proving that these algorithms can indeed perform as well, if not better, than human level in any environment when trained for an adequate amount of time with fine-tuned hyperparameters. Although these results are observed for the CartPole environment, these results will translate correspondingly with other environments, albeit the probability of achieving perfect scores with expanding state and action spaces may vary.

Finally, we can observe that with these performance characteristics, we can look to their practical applications. RL is already being used to provide value; in robotics to design and automate control principles, in chemistry to optimize chemical reactions such as in [18] and in financial trading, to make smarter risk-averse trades, as described in [19].

4 Conclusion

This paper consists of detailed observations on the performance of three model-free learning algorithms, namely deep Q-networks, dueling deep Q-networks and state-action-reward-state-action (SARSA). Multiple inferences are produced from these observations, mainly confirming the validity and performance of an agent using these three algorithms on the selected environment, CartPole. Based on the experimental results, it is conclusively proven that off-policy learning is more effective than on-policy learning and that duel DQNs are a definite improvement on its primitive form in DQNs. Knowing that these algorithms are model-free, it only engages one's curiosity about its innumerable applications in different environments. We look forward to learning more and hope to make a significant contribution to the field.

References

1. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction, 2nd edn. The MIT Press (2014, 2015)
2. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg1, S., Hassabis, D.: Human-level control through deep reinforcement learning. Nature **518** (7540), 529–533 (2015)
3. Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M.: Playing Atari with deep reinforcement learning (2013)
4. van Hasselt, H., Guez, A., Silver, D.: Deep reinforcement learning with double q-learning (2015)
5. Stadie, B.C., Levine, S., Abbeel, P.: Incentivizing exploration in reinforcement learning with deep predictive models (2015)
6. Guo, X., Singh, S., Lee, H., Lewis, R.L., Wang, X.: Deep learning for real-time Atari game play using offline monte-carlo tree search planning, pp. 3338–3346. NIPS (2014)
7. Bellemare, M.G., Ostrovski, G., Guez, A., Thomas, P.S., Munos, R.: Increasing the action gap: new operators for reinforcement learning. AAAI (2016)
8. Nair, A., Srinivasan, P., Blackwell, S., Alcicek, C., Fearon, R., Maria, A.D., Panneershelvam, V., Suleyman, M., Beattie, C., Petersen, S., Legg, S., Mnih, V., Kavukcuoglu, K., Silver, D.: Massively parallel methods for deep reinforcement learning (2015)
9. Schaul, T., Quan, J., Antonoglou, I., Silver, D.: Prioritized experience replay. ICLR (2016)
10. Watkins, C.: Learning from delayed rewards (1989)
11. Melo, F.S.: Convergence of q-learning: a simple proof (2001)
12. Nair, A., McGrew, B., Andrychowicz, M., Zaremba, W., Abbeel, P.: Overcoming exploration in reinforcement learning with demonstrations. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 6292–6299. IEEE (2018)
13. Wang, Z., Schaul, T., Hessel, M., van Hasselt, H., Lanctot, M., de Freitas, N.: Dueling network architectures for deep reinforcement learning (2016)

14. Rummery, G., Niranjan, M.: Online q-learning using connectionist systems (1994)
15. Marco, W., Jurgen, S.: Fast online q(λ). Mach. Learn. **33**(1), 105–115 (1998)
16. Perez, J., Germain-Renaud, C., K´egl, B., Loomis, C.: Grid differentiated services: a reinforcement learning approach. In: 2008 Eighth IEEE International Symposium on Cluster Computing and the Grid (CCGRID), pp. 287–294. IEEE, (2008)
17. Plappert, M.: keras-rl (2016). https://github.com/keras-rl/keras-rl
18. Zhou, Z., Li, X., Zare, R.N.: Optimizing chemical reactions with deep reinforcement learning. ACS Central Sci **3**(12), 1337–1344 (2017)
19. PIVA, A.: Dynamic trading strategy for a risk-averse investor via the q-learning algorithm (2019)

Location-Aware Security System for Smart Cities Using IoT

V. R. Azhaguramyaa and K. Srinivasan

Abstract Digitization is a key to provide smart solutions to the various infrastructure, economic and security issues in the modern world. Fortunately, many of the new technologies have made this theme successful. Cities which provide such smart solutions to these issues are termed as smart cities. The smart cities are abundantly provided with science and technological resources. One of the biggest nightmares of such cities is the challenge faced while providing security to its people and its resources. This paper provides a smart solution to tackle such security issues economically using RFID technology. In recent years, there has been a sharp rise in the applications using RFID because of its simplicity and intangibility with the object. The main idea behind this work is the subject may not be in multiple places at the same time. This system enables us to keep only the right persons in the right location. The system can provide increased governance on precious resources.

Keywords Security · Smart city · RFID · IoT · Cloud

1 Introduction

A smart city is an urban development mission which is taken to integrate bundles of information and communication technology (ICT) solutions for various security and economical issues in efficient and secured manner [1]. It helps to manage the assets of the city including, but not limited to, schools, libraries, hospitals, banks, transportation schemes and other community-level services [1]. The main aim of the smart city is to provide a quality life with enhanced technology development and

V. R. Azhaguramyaa (✉)
Department of Computer Science and Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, India
e-mail: azhaguramyaa@skcet.ac.in

K. Srinivasan
Department of Electronics and Communication Engineering, Sri Krishna College of Technology, Coimbatore, India
e-mail: k.srinivasan@skct.edu.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_51

to meet the needs of the residents. The information and communication technology allows the administrators to make direct contact with the local community and can monitor the happenings in and around the city [1]. This allows bringing out how it is being evolved and how to enable better life quality. By using the data collectors like sensors integrated with real-time monitoring systems with the help of IoT, useful information from residents and devices are collected. Then, the information is processed and analyzed [1]. The gained knowledge can be used as key valuable resources for tackling the problems in the future.

ICT provides enhanced quality and performance of services to reduce the consumption of valuable resources and provides healthy contact between the people and their government. Large number of smart city applications is developed only in the motto to provide a good level of response to the challenge with the previously gained experience [1]. Almost many of the sectors in smart cities are undergoing such transition by making use of this technology.

Providing security to the citizens especially the children and women is an important responsibility of the government. Smart cities are also abundant with valuable resources and information. Hence, it is a mandate one to keep them safe from the harmful social elements [1]. Document [2] clearly explains the various security issues in smart cities.

Though security measures like CCTV are providing good service, but it is not up to the mark in specifying the valuable information over the incident. We can provide a high level of security by using this digital connectivity concept in an altered manner. RFID is one of the hottest topics in such security providing mechanisms [3]. It has been successfully satisfied with many of the requirements for establishing safe environments. In this paper, a security mechanism has been proposed which uses RFID as its key technology.

2 Literature Review

RFID stands for radio-frequency identification [4]. The prime reason for taking this RFID is due to its low cost. It enables identification from distance unlike bar codes and QR codes [4].

This system consists of small tag that is embedded with antenna for the purpose of transmission and reception. The tag is capable of carrying 2,000 bytes of data or less [4]. The tag serves much similarly as the strips attached to the e-cards and ATM cards, etc. The other scanning mechanisms, like QR code and bar code, require complex scanning mechanisms. But the RFID readers are costly and functional efficient [4].

The QR codes and bar codes can be scanned up to some centimeters. The document clearly [4] explains that RFID reader can scan more than 20 feet without human intervention and no need to position the tag relatively to the scanner.

In spite of this RFID have lot and lot of advantages like blocking particular blocks of the unique code, killing the unwanted tags and making tag sleep at a particular time. But still the RFIDs have no common standards in the industry.

Fig. 1 RFID scanners

That is the only reason RFID mechanism has not been publicly well known. The RFID tags are of two types, active and passive tags [4].

The active tags have their own power source and the passive tags are not. The active tag uses battery source to energize the system [5]. The passive tags use the energy emitted by the reader and reflect the information that carry in return. The former have limited life span, but the later is not [4].

Basically, the RFID system consists of three components a scanning antenna, a transceiver and transponder. The antenna emits radio-frequency signals of short range. Usually, the ranges are 128 kHz, 13.6 MHz and 915 MHz, and it provides the RFID tag with the necessary energy to communicate [4].

The scanners can be affixed to a surface, but handheld scanners are also available in the markets. When an object holding RFID gets passed through the scanner, it immediately makes the tag wake and transmits the information to the scanner which is illustrated in Fig. 1.

The RFID tags may be programmable. It returns information even the environments are harsh like moisture, chemical fields, etc. Some of the security techniques for authentication used in this RFID field is selective tag blocking, hash locks, random hash locks, hash chain and re-encryption [6]. Some of the major advantages include:

1. The tag need not to be on the surface of the object
2. The read time is typically less than 100 ms
3. Large numbers of tags can be read at once rather than item by item.

Ultra-high frequency can be used for scanning since it can scan up to 12 m [6]. Due to its simplicity and efficiency, a large number of security applications have been developed by using this technology. Document [7] vividly explains about the track and trace movement of the people inside the smart city. The work tells about the usage of RFID in finding out the location of the people. The RFID systems have been installed at the street ways, and the scanners scan the people who are walking on the lane [7]. From this, they got the information of who were all passing through that particular area. The details read by the scanners which are immediately sent to the cloud, and the statistics are used for the future use [7]. The system proposed in document [7] just reads and allows the person to pass through the scanners. By taking this methodology as a key, our system has been developed for ensuring the security to the citizens of the smart city.

The government of India has taken much effort to provide a unique id to their citizens named "AADHAR." All details of the people are stored for their identification purpose. We can provide esteemed level of security to the inmates of the country and modern smart cities by adding one more attribute called RFID tag to that card [6]. The citizens can be provided with a permanent unique RFID. They can carry the id wherever they go inside the city. By using this unique id, they get scanned and their locations can be obtained in case of any issues [8]. Already the important places in cities like government organizations, malls, cinema theaters, etc. have this type of security at their entrances and exits. This system of RFID scanning security mechanism has already been successfully implemented in various fields like:

2.1 E-passport governance
2.2 Smart transportation
2.3 Toll collection booths
2.4 Hospitals
2.5 Banks.

3 Design and Implementation Methodology

The basic working principle of this mechanism is much similar to the others in which we are going to scan and retrieve data. As of security concern we can depict a permanent unique RFID in Aadhar card of each every individual in the city. Here, the permanent id will act as a primary key, and it is unique [6, 9].

3.1 Assumptions

1. Smart city environment has been considered (24*7 power supply, Internet facility)
2. Tags have required information.
3. Server's computational ability is higher than the reader.
4. The communication over server and reader is secured.
5. The communication over tag and reader is secured.

3.2 Design

The system works as a network of sensors. There are three different types of scanning accessibilities provided to the sensor. Main scanners are directly connected to the government's main server which contains complete details of the card holder. Local scanners are connected to the particular location's local server. It can only send information to the government server if required and not in return. The local scanners can have the write access to the card, and hence, it can provide local id equivalent to the primary id. Similarly, it can have the access to destroy the local id, and already provided local sub-scanners are connected to the cloud of the local scanners only. They cannot talk to the main server. It can update the details to the local database.

3.3 Implementation

The RFID tag usually consists of three different layers. The primary layer consists of the primary id1 which is scannable only by the main scanners installed at main parts of the city.

This will make the government's chief server updated with the location of scanning. Possibly, we can install four to five main scanners inside the city. The main scanners cannot have the access to modify the details of the card. The secondary layer consists of primary id 2 which is scannable by the local scanners. These scanners have the write access to the card and can generate an equivalent local id to that primary id. The local id is added to the tertiary layer of the card. The local id provided is viable till it is scanned by the provisioned sensor to destroy it. The local sub-scanners can have the access to the tertiary layer only and can read the local id. This arrangement has been clearly explained in Fig. 2.

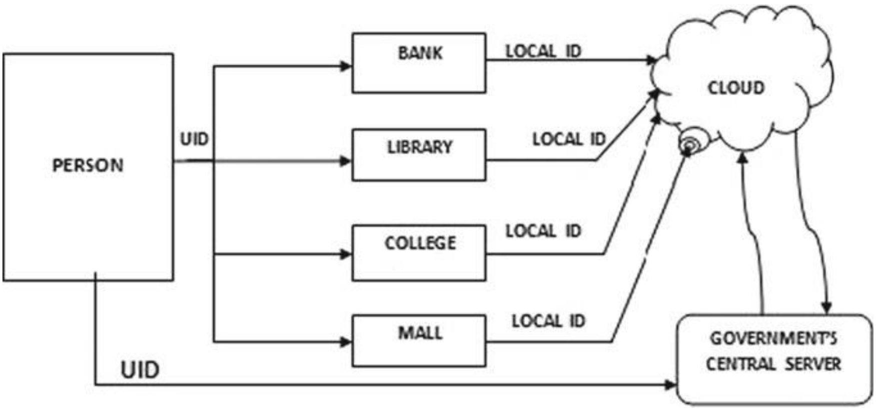


Fig. 2 System architecture

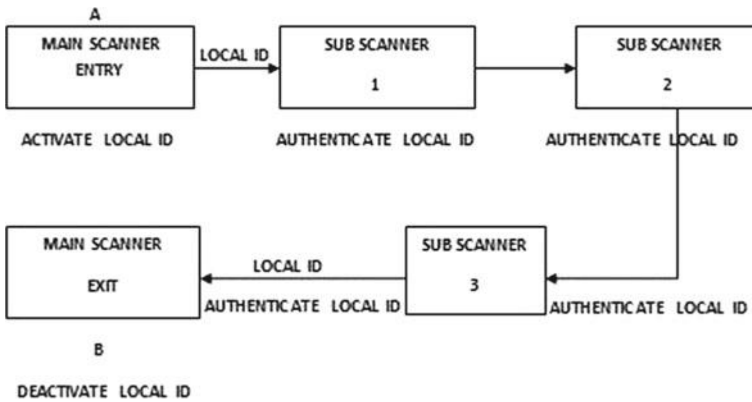


Fig. 3 Local server mechanism

For example, if a person goes to a bank in the city, the permanent id is scanned and a local id is provided in connection to the primary id. The local id is generated by the local server of the bank. The entry time of the person with that id is noted. If he leaves the bank the local id connected with the primary id is destroyed then the exit time is also noted. The local id can be added and destroyed at the predefined entry and exits of the bank installed with local scanners.

All details from local scanners are updated to the cloud. From this a person may not roam more than one place since, only after the deactivation of the local id generated by the bank the new local id can be connected to the primary id. Similarly, if a person enters into the IT company, a local id is generated for him. If he does not leave the campus even after the working hours, the local id generated is not deactivated. Hence we can know that he is still inside. The mechanism has been explained in Fig. 3.

Similarly, if any suspicious activity is done by a person the local id can be connected to the primary id and all the details regarding the primary id is maintained by the government so we easily found the details of the person. All the local servers (i.e., the local id generated by the important places like banks, malls, hospitals) are connected to the centralized server maintained by the government.

The local scanners only know about the primary id, but all the details are highly secured under the government. In case of requirement, all the local servers are connected and the data from various resources. Hence, we can link the details obtained and analyzed. It is much helpful for the police department to find the guilty. By using this project, we can increase the probability of catching the person. The detailed working mechanism has been explained in the flowchart (Fig. 4).

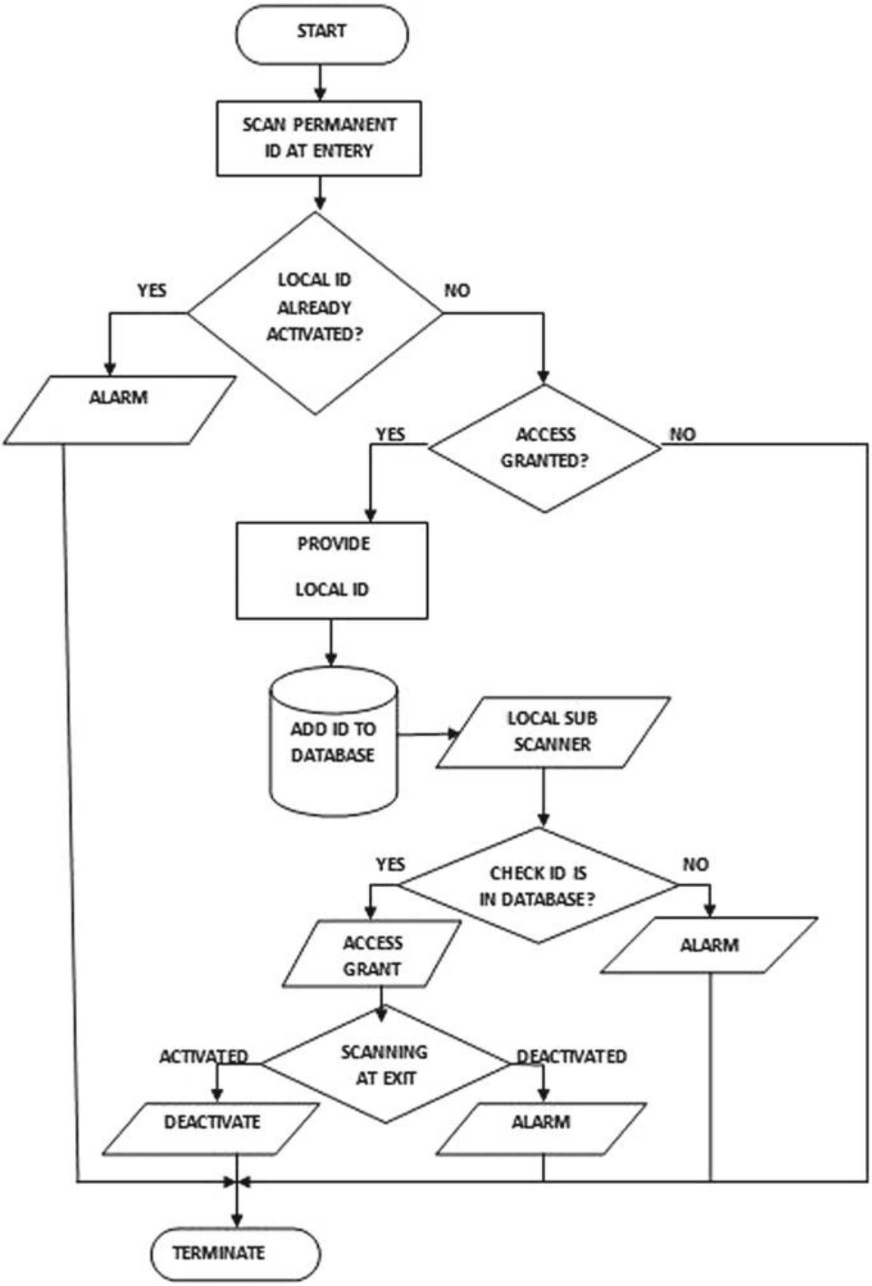


Fig. 4 Working mechanism

4 Application Areas of the System

4.1 Entry Permission Through Location Sensing

By using the RFID tags, we can restrict a person from entering into particular area by blocking the secondary layer in the card. If he does not have the access to a site that he wants to enter, the particular card number has been excluded and the local id may not be given to him.

4.2 Protection for Children

The Aadhar is provided only to the children of age above 6 years, for them we can provide the unique id along with birth certificate as it is mandatory. The printed RFID tags can be attached to the id cards of the school going children [9].

The ids can be stitched on their uniforms. The id is scanned when they board on the school bus.

Similarly, the id is scanned at the school entrance. Again, the id is scanned when they leave the campus. At each and every scanning, an alert message can be sent to their parents [8]. If the children come out of the school without the knowledge of the parent the tag is scanned and it alarms their parents. If the local id created is not deactivated at the evening after school hours, we can take necessary steps to find the child. By using this mechanism, we can track the movements of the children [8]. We can provide them with multiple numbers of copies. Since there is no need to wear this tag, we can keep it inside their bag also.

Similarly, the newborns in hospitals are more prone to abuse or kidnap. We can use RFID tags for the identification of the children [10].

4.3 Protection for Women

We can scan the tags with various scanners installed inside the city. If they are kidnapped or any harmful events occur, we can match the scanning time along with CCTV footages and get some useful clue of that crime. We can localize them easily by finding the location between the previously scanned location and the very next location to be scanned [9].

4.4 Protection at Social Gatherings

We can install the RFID scanners at people gathering places like parks. The count of scanning over various places inside the park is taken into account [11]. If a person roams abnormally an intimation message can be send to the guards immediately and one more important thing is that we can use this RFID mechanism extensively at places where the mobile phones are strictly prohibited. In places like temples, we can use this mechanism to find the lost person in midst of huge crowd [11]. The last scanned location provides their location and also the tag provides the details of the person. Hence, this technology helps a lot during the big temple festivals and social gatherings, etc.

4.5 Protection for Assets

We can add tags over the assets like jewelery, cars, etc., and the scanned location helps to find the lost assets [2].

4.6 Smart Transport System

We can fix RFID scanners at the entrance and exits so that we can pay using that id also [8]. It is impossible to travel in the bus without buying tickets. After each stage, the passengers list is displayed; hence, the person traveling even after their stage can be caught [8].

4.7 Monitoring Traffic

We can monitor the vehicles on the road by using this technology. By matching the CCTV footages and scanned id, we identify the person who caused the accident. We can identify the person who violated traffic rules also. The over speed vehicle can also be noted and tracked [8].

4.8 Attendance Management

We can monitor the student's presence inside the campus. While they enter into the college, the local id is created and stored [12]. The id will remain till they leave the campus.

Table 1. Comparative Analysis

Existing system	Proposed system	Future work
• If tag is present allows the person without any prohibition • Restrict the person only when that person does not have the tag	• Even though a person has a tag, this system can decide who needs to be in and who needs to be out • Find out the person who illegally came out from the previous location	• Still could not find out the id card maculation • To avoid that need to incorporate the proposed system with various other security mechanisms like Bio-scanners and CCTV cameras to improve security

5 Comparative Analyses

The existing works clearly explains about the scanning mechanism done by the RFID. All of them can scan the tag and allows the subject without any prohibition [7]. We can restrict the subject only when he does not have the tag. But by using this system, we can decide who needs to be in and who needs to be out. Due to the local id provision we can find out the person who illegally got out from the previous location. Suspicious roaming can also be identified by setting the maximum number of entries for an individual in a particular area. The main drawback of this system is, we can't find out the ID card maculation. We can incorporate this system with various other security mechanisms like Bio scanners, CCTV cameras to overcome this issue. Otherwise we can improvise this system by utilizing the future technologies [13]. Detailed analysis presented in Table 1.

6 Privacy Concerns and Drawbacks

Applications of Internet of things can bring convenience to people, but if it cannot ensure the protection of personal privacy, private information may be leaked at any time. So the security of it cannot be ignored. Once the signal of IoT is stolen or interrupted, it will directly affect the security of the entire information of IoT. With the widely spreading of IoT, it will provide more extensive information; the risk of exposure of such information will increase. If it has a good solution for security issues, it will largely restrict its development. Thus, above all the problems of IoT, security problem is particularly important [14]. But in this method the unique id is only read by the local scanners installed in various places of the city. They can create one local id related to the physical ID and store it in cloud. The government only has the authority to access and relate the details of the local id and primary id. And the government also does this thing only when there is an issue. Hence it doesn't affect the privacy also [14]. The one more issue with this system is tag identification when there is more than one tag under scanning area [11]. We can provide solution to this problem by using various Tag resolution techniques available [11, 15].

7 Conclusion

Each and every science experiments has two sides, i.e., both positive and negative side. It benefits can be achieved depend on which side we use. This project helps to enhance the smart city's security in considerable manner. But the only thing is, we need to carry a card wherever we go and it makes us to feel like someone is tracking. Carrying card is not a big deal since we carry driving license, ATM card everywhere. But we need to come to a mentality that there is no need to track a good citizen and should understand that it only bothers the harmful elements of the society. Definitely, this system will produce a large impact on the society and helps to reduce the possibility of crimes.

References

1. Jing, Q., Vasilakos, A.V., Wan, J., Lu, J., Qiu, D.: Security of the internet of things: perspectives and challenges, 2481–2501 (2014)
2. Ijaz, S., Shah, M.A., Khan, A., Ahmed, M.: Smart cities: a survey on security concerns. (IJACSA) Int. J. Adv. Comput. Sci. Appl. **7**(2) (2016)
3. Dixit, V., Verma, H.K., Singh, A.K.: Comparision of various security protocols in RFID. Int. J. Comput. Appl. (0975–8887) **24**(7) (2011)
4. Kaur, M., Sandhu, M., Mohan, N., Sandhu, P.S.: RFID Technology Principles, Advantages, Limitations & Its Applications. Int. J. Comput. Electr. Eng. 31793–8163(1) (2011)
5. RFID: a technical overview and its application to the enterprises. IT Prof. (2005)
6. Nie, X., Zhong, X.: Security in the internet of things based on RFID: issues and current counter measures. In: Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)
7. Mora-Mora, H., Gilart-Iglesias, V., Gil, D., Sirvent-Llamas, A.: A computational architecture based on RFID sensors for traceability in smart cities. Sensors, 13591–13626 (2015). https://doi.org/10.3390/s150613591, ISSN 1424-8220
8. Prasad, A.K., Saxena, N., Xiang, T.: Location-aware and safer cards: enhancing RFID security and privacy via location sensing. In: Proceedings of the Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks, pp. 51–62 (2012)
9. Kumar, N.M., Goel, S., Mallick, P.K.: Smart cities in India: features, policies, current status, and challenges. In: Technologies for Smart-City Energy Security and Power (ICSESP) (2018)
10. Chien, H.Y., Yang, C.C., Wu, T.C., Lee, C.F.: Two RFID-based solutions to enhance inpatient medication safety. J. Med. Syst. **35**(3), 369–375 (2011)
11. Zheng, Y., Li, M.: Fast tag searching protocol for large-scale RFID systems. Trans. Netw. 21(3) (2013)
12. Arulogun, O.T., Olatunbosun, A., Fakolujo, O.A., Olaniyi, O.M.: RFID-based students attendance management system. Int. J. Sci. Eng. Res. 4(2) (2013)
13. Kazak, A.N., Shamayeva, N.: Separate aspects of smart cities security. In: IEEE International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS) (2018)
14. Liang, X., Shetty, S., Tosh, D.: Exploring the attack surfaces in blockchain enabled smart cities. In: 2018 IEEE International Smart Cities Conference (ISC2) (2018)
15. Pai, P., Shashikala, K.L.: Smart city services—challenges and approach. In: International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) (2019)

An Assessment Study of Gait Biometric Recognition Using Machine Learning

Abdullah Mohmand, Munish Sabharwal, and Isao Echizen

Abstract The biometric gait recognition system has attracted much attention in the computer vision community because of its discreet recognition advantage in relatively distant locations. The present study provides a comprehensive overview of recent developments in approaches to detecting walking. The survey focuses on three main themes of a general walk-detection system, namely the appearance of the image of walking, the reduction in the dimensionality of features, and the classification of walking, and in addition, a review of the available records or datasets for the public to process the gait reorganization using different methods and calcifications. The final discussions describe a series of research challenges and give promising pointers for the future on the ground.

Keywords Biometrics · Gait biometric recognition system · Classification techniques · Machine learning

1 Introduction

With the growing demand for visual surveillance systems, remote identification of people has recently become an area of great interest. The growing number of threats is pushing society today to protect itself from cruel crimes such as terrorism. This scenario requires reliable biometric systems for effective human identification, even remotely. Detecting gait essentially aims to solve this problem by identifying people

A. Mohmand (✉) · M. Sabharwal
Computer Science Engineering, Chandigarh University, Mohali, Punjab, India
e-mail: abdullah.mumnd@gmail.com

M. Sabharwal
e-mail: mscheckmail@yahoo.com

I. Echizen
National Institute of Informatics, Graduate Institute of Advanced Studies, The University of Tokyo, Tokyo, Japan
e-mail: iechizen@nii.ac.jp

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_52

by their gait. Few other biometric data such as face, iris, fingerprint, and hand geometry have been studied and methodically applied in surveillance systems. Despite their generalized application, these biometric resources suffer from two major issues: (1) they do not match low-resolution images and images remotely, and (2) user collaboration is required to produce accurate results [1]. In addition, this can be hidden in most serious crime situations.

The various different types of biometric types available as on date are depicted in Fig. 1:

Expanded consideration has been paid as of late to viably distinguish people for the avoidance of terrorist assault. Numerous biometric innovations have made to recognize and confirm individual's facial investigation, fingerprints, palm, iris, walk, or combination of these properties [3].

Contrasted with other biometric strategies, walk location offers a few interesting highlights. The loveliest characteristic is his prudence, which does not require consideration and participation of the subjects watched. Likewise, human acknowledgment framework can be distinguished at a huge span without this being fundamental physical data of the subjects. This element has incredible focal points, particularly when individual data, for example, the facial picture is secret also, the acknowledgment of the methodology offers extraordinary potential for low-goals video discovery where

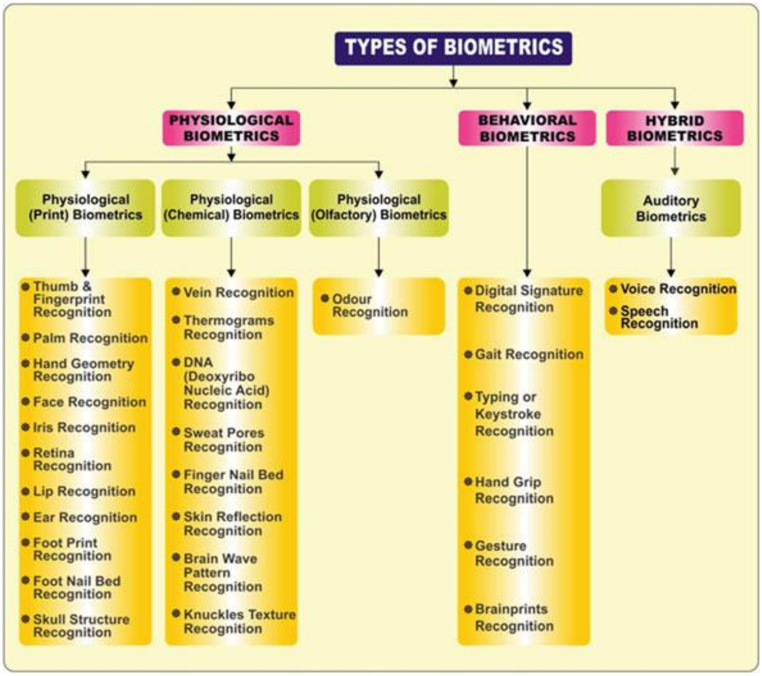


Fig. 1 Various types of biometric technologies [2]

Fig. 2 4 distances angles [8]

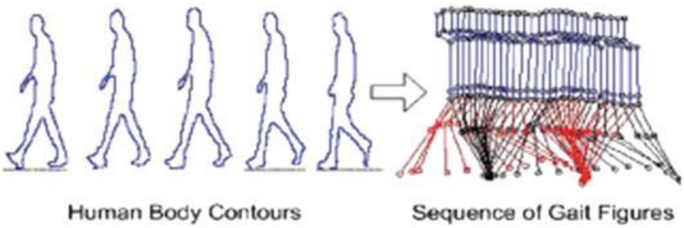


Fig. 3 4 Sequence of gait figures [30]

other biometric advances may be incorrect because of bad pixels to compare human subjects [3].

The general structure of programmed acknowledgment comprises topic acknowledgment, outline extraction, highlight extraction, determination of attributes and characterization. When moving subjects are distinguished, individuals are perceived and isolated from the foundation of the image. The most broadly utilized technique is the foundation subtraction, the attempts to separate articles from the distinction between the demonstrated foundation and the present picture [4]. The first acknowledgment of individuals in pictures and the steady division of the foundation can be considered as a stage of pre-preparing the acknowledgment of the passageway; this would go past the extent of this survey.

After individuals were removed from the background and the capacities that can be utilized for discovery extracts from these divided people. There are basically two kinds of walk methodology, to be specific model-based and model-free methods. On another page, the capacities without model more often than not utilize just double outlines and do not have to manufacture a model for individuals who run [5]. Model-based and modeless methodologies are talked about in area 2.

The highlights removed from portioned video successions are commonly not effective for order, and as a result of their huge dimensionality, it requires too many learning tests. To take care of this issue, numerous dimensionality decrease techniques have been proposed. These proposed techniques are head segment examination (PCA) [6] and direct separation investigation (LDA) [7] are often utilized. Segment 3 looks at different strategies for diminishing the dimensionality of highlights. The last advance is to rank the test arrangement dependent on the removed highlights for a specific person. The arrangement of the properties of the methodology is chiefly founded on three classes of techniques, to be specific the immediate grouping, the likeness of the worldly successions, and the model of the state space. The immediate order is by and large utilized in the wake of separating a solitary portrayal or key edges from an arrangement of walk pictures. While the comparability of time groupings is utilized to quantify the separation between two arrangements of the stride. This stochastic methodology unequivocally utilizes both the closeness data between the test and reference arrangements and the probability of event of the structures [3]. Area 4 displays these three classes of grouping techniques. As portrayed in Sect. 5, a few standard run records are accessible to the general population (Figs. 2 and 3).

Some completely different techniques model items of the chassis severally. In Lee and Grimson [9], the chassis is incontestable as fourteen inflexible components consolidated at the joints. The outcomes, above all the joint edge sign of the joints, square measure viewed as procedure parts for distinctive proof and check. You likewise get static body information obsessed on the Procrustes form investigation of fixing moving outlines which will be applied freely or in combine to enhance acknowledgment. All the additional as currently, Bulgouris [10] has divided the chassis into many segments and consolidated the result from varied items of the body into a typical separation metric. Analyzing every gathering's commitment to acknowledgment execution has improved the acknowledgment rate utilizing the foremost causative components. Also, Cunado et al. [11] halftones the traditional outlines of a mobile cycle into seven distinctive components and abridges the impacts of every part on the acknowledgment of the stride.

Objective of study:

- To review the previous research work done using various classification techniques for gait biometric recognition
- To analyze and find out the best method for gait biometric recognition.

Fig. 4 Standardized GEI [5]

2 Review of Literature

2.1 Gait Image Illustrations

2.1.1 Model-Based Techniques

This methodology gives a lot of static/dynamic body parameters by displaying or following body parts, for example, appendages, legs, arms, and thighs. The level marks got from these model parameters are utilized to recognize an individual. Clearly, model-based methodologies are outwardly invariant and scaling autonomous. These focal points are significant for handy application as it is far-fetched that the reference arrangements and test successions will be considered from a similar perspective [3]. In any case, model-based methodologies are delicate to the nature of the step successions.

Model-based essential methodologies utilize static body structure parameters as recognizing highlights. Qinghan [12] present advance structure parameters, including step and rhythm. The speed is evaluated by the periodicity of the strolling and the length of the progression is determined by the proportion of the separation voyaged and the means. Boulgouris et al. [8] Fig. 4 partings of human bodies, to be specific the separation among head and foot, the separation among head and pelvis, the separation among foot and pelvis, and also the separation among left and right foot. All the additional as these days, Nixon and Carter [30] removed nine directions of the shapes of the chassis keen about human anatomical info to make a drawing in second, as appeared.

2.1.2 Model-Free Approaches

Demonstrating methodologies concentrate on the shapes of silhouettes or the overall development of the organic structure as important displaying the overall organic structure or a section of the body. Modeless methodologies area unit obtuse toward silhouettes and contrasted with model-based methodologies has the face of low

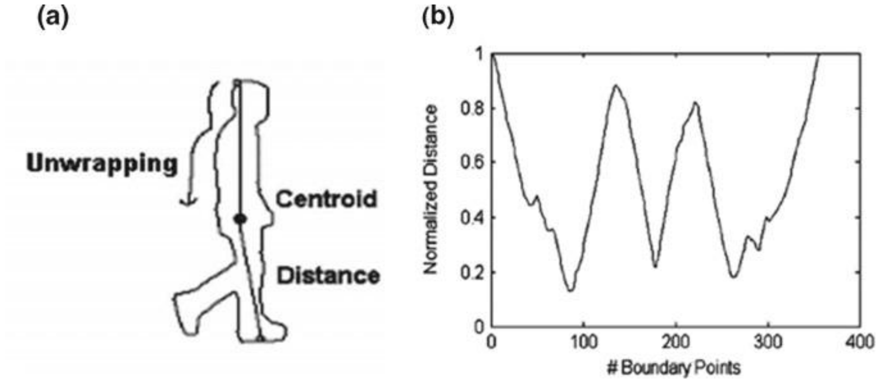


Fig. 5 **a** Counterclockwise **b** normalized distance

Fig. 6. Frame from the USF dataset [35]

procedure sweat. yet they are unremarkably not hearty as method as vision and scale.

The examination by Tao et al. [35] utilizes the outlines themselves as highlights that square measure scaled and adjusted before use. Whereas the progress signature at intervals the bottom calculation might be a grouping of passageways. The seven photos on the left in each column square measure outlines of running successions, the image on the extreme right is that the relating cluster vitality image. GEI changes over spatial-fleeting data into a solitary second step example throughout a walk cycle, on these lines maintaining a strategic distance from match qualities in time groupings. GEI is generally vigorous against commotion by averaging photos of a step cycle. All the same, it loses the dynamic choice between the progressive photos. Nixon and Carter [30] apply GEI wave disintegration to infrared walk identification. The dimension is characterized as a result of the extent separation between the furthest left peel and conjointly the furthest right component of the form. The dimension of the external sort may not be solid because of the standard of the outlines. Be that as a result of it would, the element itself would be extra eligible for poor data.

Afterward, BenAbdelkader et al. [13] partners the total summary and conjointly the dimension of the outline of the external type with the trail. Wang et al. [1] extract the second layout of the outline into a 1D signal by utilizing the house between the pixels on the framework and conjointly the attentiveness of the shape (see Fig. 4) outlines. Dadashi et al. apply a wave modification to those 1D sign to induce eliminate the wave parcel molecule coefficients as a walk signature rather than reckoning a separation between every component on the shape and conjointly the attentiveness of gravity. Piccardi [14] split the outline into rakish divisions and reason the traditional separation between the forefront pixels and conjointly the attentiveness of gravity in each edge [*fr1].

Some altogether totally different calculations be a part of significance to the examination of the state of outlines. Wagg and Nixon [10] applied the investigation of the type of out of doors layer on the classes of the figure and removed a conventional type of covering from a rendezvous of outlines as a progress signature. Alotaibi and Mahmood [15] plays out an inert gas modification of the twofold outlines to amass a model of walk successions.

Programmed step acknowledgment is new rising exploration field with just a couple inquired about systems. It has the benefit of being inconspicuous on the grounds that body invading gear is not expected to catch walk data. From an observation point of view, stride acknowledgment is an alluring methodology since it might be performed a way off, clandestinely. In this paper, we present a stride acknowledgment system that recognizes individual's dependent on static body parameters recouped during the strolling activity over various perspectives. The expectation is that in light of the fact that these parameters are legitimately identified with the three-dimensional structure of the individual they will be less delicate to blunder presented by variety in view edge. Likewise, rather than detailing percent right (or acknowledgment rates) in a restricted database of subjects, we infer a normal disarray metric that enables us to foresee how well a given element vector will channel character over an enormous populace.

2.2 Dimensionality Reduction of Features

The spatiality decrease of the extraction feature from the step coaching dataset is often additional noteworthy than that of the preparation data, transferal concerning the frustration of customary order calculations. This can be referred to as a test issue. Consequently, a part decrease calculation is anticipated to get rid of valuable and instructive highlights for grouping. Head half investigation (PCA) and straight discriminant examination (LDA) [16] are methods for alteration standard but so much reaching highlights. Wang et al. [1] PCA applies time-fluctuating separation sign got from the define image arrangement to diminish the spatiality of the capability house. Wang et al. [7] conjointly performs PCA and LDA to break down and develop the topological structure and lower the spatiality of the sensible house. Charalambous and Bharath [5] combined PCA and multiple discriminant analysis (MDA) to technique the gait energy image (GEI).

PCA-based scheme holds just persons extractions that add most to the correction which will not be ideal for characterization. Makihara et al. [17] builds up a typical GTDA to accumulate biased knowledge on Gabor highlights and to utilize the LDA for grouping. All the additional as lately, letter and Yang et al. [28] used DLA to decrease the spatiality of organically intended highlights, whereas Hu [18] uses a two-arrange ACP and DLA to bring the form qualities. A two-dimensional LPP is defined by Kim and Paik [19] to enhance the clear character of separated qualities obsessed on a lively energy image (AEI). Whereas the overwhelming majority of the on top of methodologies focus on decreasing the spatiality of highlights; Guo and President of the USA [14] choose a set of the quality of the methodology by augmenting the common knowledge concerning the quality of the methodology.

PCA, LDA, LLP, DLA are used for dimensional reduction.

2.3 Gait Classification

2.3.1 Direct Classification

The grouping techniques for the immediate methodology do not take into account the transient information of the step successions. They rely upon the one portrayal or the key edges off from a succession of walk photos. The nearest neighbor classifier K decides the category of check properties as indicated by the amount of k next learning models. The foremost well-known category, recorded underneath the k shut getting ready models, is chosen because of the category for the check work. Vankoski and Dias [20] take away the key edges of a run cycle to form a model and subsequently stick with it composition the nearest neighbor keen about the model scores. k-closest neighboring guideline is delineated by Makihara et al. [17] data on the repeat of the hip development for arrangement (Fig. 5).

Likewise, a couple of creators utilize distinctive discriminant classifiers. The support vector machine (SVM) is delineated by Sarkar et al. [21] for wavelet-deteriorated properties from the trail energy image (GEI). SVM is taken into account as a summed up traditional classifier and could be a normalized learning technique instead of associate degree administered knowledge classifier. Utilize an electrical device support vector (TSVM) machine to play out a semi-administered grouping of the following mark removed by wave bundles. The TSVM considers large highlights and viably analyzes the association structures of the equipment attributes (Table 1).

2.3.2 Matching of Temporal Sequences

The methodology could be a distinctive human movement. The mark of the methodology usually contains a grouping of highlights of the methodology, with the guide of that the closeness of two worldly successions of the step is characterized. Many creators utilize the mixture separation during a run cycle licitly because of the separation between two groupings. Alotaibi and Mahmood [15] drags the check arrangement to the reference grouping to make decision things of the bottom separation [3]. Tao et al. [35] uses the amount of the number of pixels at the crossing purpose and therefore the association of two define footage to measure the similitude between detector

Table 1 Comparative chart of gait biometric

Author	Title	Data set	Classifier	Result
Ashvini et al. [1]	Human Identification By Walk Recognition Using Relevance Vector Machine	CASIA	SVM, RVM	The classification they have used SVM (91%) and RVM. Comparing both, RMV(98%) results to be better than SVM classifier
Pradeep Kumar et al. [3]	Multimodal Gait Recognition with Inertial Sensor Data and Video using Evolutionary Algorithm	Generated Dataset	SVM, CNN, GWO	An average accuracy of 91.3% has been recorded using the GWO optimizer on all walk sequences
Zifeng Wu et al. [15]	A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs	CISIA-B Dataset	CNN	The average recognition rate can reach 94.1%, much better than the previous best result (less than 65%)
Alexandra Vieira et al. [22]	Dynamic Human Gait VGRF Reference Profile Generation via Extreme Learning Machine	Generated Dataset	BNN, ELM, MSVR	The BNN presents training and testing times clearly larger than the MSVR and ELM. Considering the real-time constraints are important, the BNN is discarded from this discussion

define footage and exhibition define footage. The check successions square measure divided into many subsequences relying upon the term of the cycle. In lightweight of the comparison of the images, they figure the connections between each subsequence and therefore the arrangements of the total show and choose the center greatest relationship esteems as likeness for wholeheartedness.

In any case, the direct mixture separation is signally not acceptable for the estimation of step groupings since it is accepted that the check arrangement and therefore the reference succession have AN indistinguishable strolling amount [3]. This is not frequently things much speaking. Dynamic time warp (DTW) could be a valuable methodology to change double cross sign of assorted lengths. Jiwen and Zhang [6] uses dynamic time corners to quantify the likeness between two groupings of separation waves. The DTW parameters square measure controlled by the progression return and therefore the stage distinction. Han and Bhanu [23] apply the DTW strategy to time standardize two methods that square measure movement like projections in chance area work (SoPF). All the additional as these days, Liang et al. [24] improve the DTW keen about mythical being form separations to determine the separations between the form groupings.

Moving is AN intermittent action, which means that return investigation of spatial-transient step sign is a beguiling methodology [3]. Alotaibi and Mahmood [15] isolates the walk style into seven locales and uses circles to suit each space. They apply the Fourier amendment to the time sign of those circles and concentrate sufficiency and stage elements for characterization. Sarkar et al. [21] matches the Fourier descriptors and therefore the major Fourier descriptors as classifiers and infers that the key Fourier descriptors of the human forms surpass the Fourier descriptors.

2.3.3 Matching of Temporal Sequences

This model is used effectively in discourse acknowledgment and signal acknowledgment [6, 35]. Gee speaks to varied periods of a procedure as shrouded states. They expect that this state is compact clearly by the past state which it is freed from the condition of the story. The possibilities of perception and alter probability area unit determined utilizing learning input data. The foremost elevated back probability subject is chosen because of the perceived outcome. Gee-based mostly methodologies area unit ordinarily fascinating over completely different methods since they abuse each the compare of structures among take a look at and reference arrangements and also the likelihood of the event and action of structures throughout a time of idleness [35].

2.4 Gait Datasets

The achievement and improvement of another application will depend intensely on the information accessible to help the test results. The primary worry here is the uniqueness and practicability, which for the most part results from the quantity of subjects in the database and the imaging conditions. On the off chance that the primary reason for the database is to show essential practicability, the information can be determined in a research center situation. Nonetheless, if the fundamental objective is to demonstrate this can be accomplished through PC vision, the information can be determined by shooting outside of subjects. 10 years back, when the PC had low power, stockpiling, and source. Some prevalent freely accessible the rigging records are depicted beneath (Table 2).

Table 2 Human gait databases [4]

Database	Sub	Seq	Scene description	Covariate factors
Covariate SOTON	12	12,730	Indoor conditions	Footwear, clothing, walking speed, viewpoint carrying conditions
Large SOTON	118	10,442	Indoor, outdoor, and treadmill	Viewpoint
Gait challenge	122	1,870	Outdoor environment	Viewpoint, surface, footwear, time, and carrying conditions
CMU	25	600	Indoor and treadmill	Walking speed, viewpoint, surface, and carrying conditions
GATECH	15	168	Outdoor	Viewpoint
MIT	24	194	Indoors	Time
CASIA A B	20 124	240 4092	Indoors and outdoors	Viewpoint, footwear, clothing, carrying luggage, and walking speed

2.4.1 Gait Datasets

The dataset was gathered at the (USC) University of South Sunshine State (USC) and it contains (1870) successions from (122) subjects. Each individual strolled around AN oval before cameras. There are up to five covariates for each individual: two distinct shoes types; with or while not a folder case; grass surface or solid surface: left perspective and right perspective; and two various time moments [7]. A model edge from the USF stride dataset is appeared in Fig. 6.

2.4.2 Mobo Dataset

This dataset was produced by the AI (Institute of Carnegie philanthropist University). This dataset includes 25 peoples in 3D space. Our different walk patterns: slow walk, fast walk, incline walk, and walking with a ball. All subjects are captured using six high-resolution color cameras distributed evenly around the treadmill [19].

2.4.3 Southampton Dataset

The record of the Southampton approach comprises of two information, the less amount of record into the database and the huge amount of records into the database. The less amount of record incorporates twelve subjects who drive an enclosed path at various paces. Every individual was caught with different shoes, garments, and packs. The subjects in the huge database were taped outer, internal, yet additionally from six distinct perspectives [4].

2.4.4 CASIA Gait Dataset

The gait biometric dataset is given by the Chinese Academy of Sciences of the Institute of Automation. There are three records in the CASIA gait record (i.e., record A, record B, and record C). Record B is an enormous multi-see record containing one hundred and twenty-four subjects in eleven perspectives. One hundred and fifty-three subjects went under four unique conditions: strolling ordinarily, strolling gradually, strolling energetically, and strolling regularly with a sack [20].

The primary worry here incorporates the uniqueness and common sense which are generally uncovered by the quantity of subjects in the database and the imaging conditions individually. On the off chance that the basic role of the database is to prove essential common sense, at that point information can be determined in a research facility situation however in the event that the fundamental point is to demonstrate that it very well may be accomplished by PC vision then the information can be inferred by taping subjects outside. 10 years sooner when the PC had little power and memory and the wellspring of 25 videos were simple camcorder based, which brought about edges being digitized exclusively there were truth be told, almost no extension for stride acknowledgment.

Additionally, the studies by Munish, first [25], facilitated the researcher in overall preparation of literature review and planning for the overall research and the second [26] assisted in analysis.

3 Conclusions

This document gave a far-reaching diagram of key achievement systems and ongoing advancements in the distinguishing proof and ID of the methodology. Three primary issues of acknowledgment, to be specific the introduction of the picture, the decrease of the dimensionality of the attributes, and the characterization of the strolling, are talked about. The highlights used to portray the incitement can be separated into two principle gatherings: model-based highlights and non-format highlights. Model-based highlights are removed by means of demonstrating or human body following segments, while without model methodologies underscore outlines or entire body developments. Model-based highlights are naturally more inflexible and less versatile than highlights without layouts. Nonetheless, model-based methodologies must catch excellent walk groupings and more calculation time. Interestingly, model-less methodologies are less delicate to outlines and progressively proficient in processing. Lessening the dimensionality of highlights is critical to make the arrangement increasingly proficient and spare important calculation time to fulfill the needs of constant applications. Straight and nonlinear dimensional decrease strategies are usually utilized in rigging acknowledgment. Clearly, direct dimensionality decrease techniques, for example, PCA, may not be ideal for the arrangement of steps, and non-straight strategies would be prevalent for this situation.

With respect to plan of the gangue classifier, direct arrangement strategies and those dependent on estimating the likeness of grouping successions are regularly utilized in the writing. Direct step characterization strategies lose fleeting varieties in stride designs or overlook the time request of walk successions, in spite of the fact that they ordinarily have high computational proficiency. Conversely, comparability-based techniques utilize fleeting and dynamic data over picture successions. They would consequently be more qualified for characterizing the step. Dissimilar to coordinate characterization strategies and time grouping comparability techniques dependent on separation measurements, HMM-based techniques model the periods of a shrouded state approach. Gee-based strategies are commonly desirable over different techniques since they utilize both comparability data and the likelihood of event of shapes.

Albeit much research has been done, the acknowledgment of the way to deal with individual recognizable proof is still a long way from useful. The promising bearings for future research are described below.

1. None of the research papers have used their classification methods on more than one dataset.
2. None of research paper have used combination of SVM and CNN for classification of gait recognition.

References

1. Wang, L., Hu, W.M., Tan, T.N.: Recent developments in human motion analysis. Pattern Recognit. **36**(3), 585–601 (2016)
2. Munish, S., Sohan, G.: The summation of potential biometric types and technologies for authentication in e-banking. Int. J. Sci. Rev. Res. Eng. Technol. (IJSRRET) **1**(2), 83–92 (2016)
3. Kumar, P., Mukherjee, S., Saini, R.: Multimodal gait recognition with inertial sensor data and video using evolutionary algorithm. Ski. Res. Technol. **18**, 389–396 (2019)
4. Parise, A.L.: Human Identification by walk recognition using relevance vector machine. IEEE **21**(6), 1675–1684 (2017)
5. Charalambous, C.C., Bharath, A.A.: A data augmentation methodology for training machine/deep learning gait recognition algorithms. IEEE J. **21**(6), 1685–1693 (2016)
6. Jiwen, L., Zhang, E.: Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion. Pattern Recognit. Lett. **28**, 2401–2411 (2016)
7. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette analysis-based gait recognition for human identification. IEEE Trans. Pattern Anal. Mach. Intell. **25**(12) (2015)
8. Boulgouris, N.V., Hatzinakos, D., Plataniotis, K.N.: Gait recognition: a challenging signal processing technology for biometric identification. IEEE Signal Process. Mag. **22**, 78–90 (2005)
9. Lee, L., Grimson, W.E.L.: Gait analysis for recognition and classification. In: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 148–155 (2002)

10. Wagg, D.K., Nixon, M.S.: On automated model-based extraction and analysis of gait. In: Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 11–16 (2004)
11. Cunado, D., Nixon, M.S., Carter, J.N.: Automatic extraction and description of human gait models for recognition purposes. Comput. Vis. ImageUnderstanding **90**, 1–41 (2003)
12. Qinghan, X.: Technology review-biometrics-technology, application, challenge, and computational intelligence solutions. IEEE Comput. Intell. Mag. **2**, 5–25 (2007)
13. BenAbdelkader, C., Cutler, R., Davis, L.: Motion-based recognition of people in Eigen Gait space. In: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 267–272 (2002)
14. Piccardi, M.: Background subtraction techniques: a review. In: 2004 IEEE International Conference on Systems, Man and Cybernetics, vol. 4, pp. 3099–3104 (2004)
15. Alotaibi, M., Mahmood, A.: Improved gait recognition based on specialized deep convolutional neural network. IEEE **36**(4), 994–1004 (2017)
16. Zhou, X., Bhanu, B.: Feature fusion of side face and gait for video based human identification. Pattern Recognit. Part Special issue: Feature Generation and Machine Learning for Robust Multimodal Biometrics, **41**(3), 778–795 (2012)
17. Makihara, Y., Kimura, T., Okura, F., Mitsugami, I., Niwa, M.: An automatic gait data collection system in conjunction with an experience-based long-run exhibition. IEEE **36**(3), 849–858 (2016)
18. Hu, W.M., Tan, T.N., Wang, L., Maybank, S.: A survey of visual surveillance of object motion and behaviours. IEEE Trans. Syst. Man Cybern., Part C **34**(3), 334–352 (2014)
19. Kim, D., Paik, J.: Gait recognition using active shape model and motion prediction. IET Comput. Vis. **4**, 25–36 (2010)
20. Vankoski, S., Dias, S.: Clinical Motion Analysis, The Child's Doctor, Journal of Children's Memorial Hospital, Chicago, Spring (2015)
21. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, K.W.: The human ID gait challenge problem: datasets, performance, and analysis. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 162–177 (2005)
22. Liu, W., Zhang, C., Ma, H., Li, S.: Learning efficient spatial-temporal gait features with deep learning for human identification. Sci. World J. (2018)
23. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**, 316–322 (2006)
24. Liang, W., Tieniu, T., Huazhong, N., Weiming, H.: Silhouette analysis-based gait recognition for human identification. IEEE Trans. Pattern Anal. Mach. Intell. **25**, 1505–1518 (2003)
25. Aravind, S., Amit, R., Rama, C.: A hidden Markov model based framework for recognition of humans from gait sequences. In: Proceedings of the 2003 IEEE International Conference on Image Processing (2003)
26. Sabharwal, M.: Contemporary research: intricacies and aiding software tools based on expected characteristics. AIMA J. Manag. Res. **10**(2/4), 1–16 (2016). ISSN:0974-497
27. Conniea, T., Goha, K.O.M., Teohb, A.B.J.: Multi-view gait recognition using a doubly-kernel approach on the grassmann manifold. IEEE Trans. Med. Imaging **36**(9), 1876–1886 (2016)
28. Yang, X., Zhou, Y., Zhang, T., Shu, G., Yang, J.: Gait recognition based on dynamic region analysis. Signal Process. **88**(9), 23502356 (2016)
29. Gavrila, D.: The visual analysis of human movement: a survey. Comput. Vis. Image Underst. **73**(1), 82–98 (2013)
30. Nixon, M.S., Carter, J.N.: Automatic recognition by gait. Proc. IEEE **94**, 2013–2024 (2006)
31. Sabharwal, M.: The use of soft computing technique of decision tree in selection of appropriate statistical test for hypothesis testing. In: International Conference on Soft Computing: Theories and Applications (SoCTA 2016), Amity University, Jaipur, India, 28–30 Dec 2016
32. Liu, J., Zheng, N., Xiong, L.: Silhouette quality quantification for gait sequence analysis and recognition. Signal Process. **89**(7), 1417–1427 (July 2009)
33. Wang, L., Hu, W.M., Tan, T.N.: Recent developments in human motion analysis. Pattern Recognit. **36**(3), 585–601 (2003)

34. Jiwen, L., Zhang, E.: Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion. Pattern Recognit. Lett. **28**, 2401–2411 (2007)
35. Tao, D., Li, X., Xindong, W., Maybank, S.J.: General tensor discriminant analysis and gabor features for gait recognition. IEEE Trans. Pattern Anal. Mach. Intell. **29**(10), 1700–1715 (October 2007)
36. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette analysis-based gait recognition for human identification. IEEE Trans. Pattern Anal. Mach. Intell. **25**(12) (2003)
37. Vankoski, S., Dias, L.: Clinical Motion Analysis, The Child's Doctor, Journal of Children's Memorial Hospital, Chicago, Spring (1999)
38. Hu, W.M., Tan, T.N., Wang, L., Maybank, S.: A survey of visual surveillance of object motion and behaviours. IEEE Trans. Syst. Man Cybern., Part C **34**(3), 334–352 (2004)
39. Gavrila, D.: The visual analysis of human movement: a survey. Comput. Vis. Image Underst. **73**(1), 82–98 (1999)
40. Zhou, X., Bhanu, B.: Feature fusion of side face and gait for video based human identification. Pattern Recognit. Part Special issue: Feature Generation and Machine Learning for Robust Multimodal Biometrics **41**(3), 778–795 (2008)
41. Chen, C., Liang, J., Zhao, H., Hu, H., Tian, J.: Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recognit. Lett. **30**(11), 977–984, 1 Aug 2009
42. Lee, S., Liu, Y., Collins, R.: Shape variation-based frieze pattern for robust gait recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR '07, pp. 1–8, 17–22 June 2007
43. Lee, C.-S., Elgammal, A.: Gait Style and gait content: bilinear models for gait recognition using gait re-sampling. In: Proceedingsof the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 147–152 (2004)
44. Collins, R.T., Bross, R., Shi, J.: Silhouette–based Human Identification from Body Shape and Gait. In: Proceedings of the IEEE International Conference Automatic Face and Gesture Recognition, Washington DC, pp. 351–356 (2002)
45. Han, J., Bhanu, B.: Statistical feature fusion for gait–based human recognition. In: Proceedings of the IEEE International Conference Computer Vision and Pattern Recognition, Washington, DC, vol. 2, pp. 842–847 (2004)
46. Kale, A., Sundaresan, A., Rajagopalan, A.N., Cuntoor, N.P., RoyChowdhury, A.K., Kruger, V., Chellappa, R.: Identification of humans using gait. IEEE Trans. Image Process. **13**(9), 1163–1173 (2004)
47. Lee, L., Grimson, W.E.L.: Gait analysis for recognition and classification. In: Proceedings of the IEEE International Conference Automatic Face and Gesture Recognition, pp. 155–162. Washington, DC (2002)
48. Liu, Y., Collins, R.T., Tsin, Y.: Gait sequence analysis using frieze patterns. In: Proceedings of Europe Conference Computer Vision, vol. 2, pp. 657–671 (2002)

Miscellaneous and Other Trends

A Study on Risk Management Practices in Online Banking in Bahrain

Aysha Saud Al-Ajmi, Gagan Kukreja, and Shahram Sarkani

Abstract One of the most significant technological advances that has influenced banking practices in recent years is online banking services. Those services creates the risks that banks should be aware of, and therefore should implement very strict measures of risk management, which made Basel come up with a framework of principles to successfully mitigate and deal with online banking risks. Hence, the Researchers conducted this research to study risk management in online banking in the Kingdom of Bahrain in terms of the framework provided by Basel and investigate the extent to which they applying these principles, as well as discover the perception of banks on risks associated with online banking. The Researchers conducted a semi-structure interview as a pilot study with three risks managers in various banks in Bahrain to be provided with an in-depth picture about risk management related to online banking. Moreover, a survey questionnaire was developed and distributed to nine commercial banks in Bahrain to investigate and determine their application of the 14 principles of online banking risk management by Basel. Results show that banks in Kingdom of Bahrain are working well in terms of the principles, as they have a very positive understanding of risk management and risks imposed by online banking.

Keywords Risk management · Online banking · Basel · Security controls · Bahrain

A. S. Al-Ajmi (✉) · S. Sarkani
The George Washington University, Washington, USA
e-mail: allamh3@hotmail.com

A. S. Al-Ajmi
Bahrain Defense Force, Manama, Bahrain

G. Kukreja
Ahlia University, Manama, Bahrain
e-mail: gkukreja@ahlia.edu.bh

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_53

1 Introduction

One of the most significant technological advances that has influenced banking practices in recent years is online banking services. Basel Committee on Banking Supervision [1] defines e-banking as: "the provision of retail and small value banking products and services through electronic channels as well as large vale electronic payments and other wholesale banking services delivered electronically". These services according to Driga and Isac [2] became the main mean for costumers to deal with banks resulting in e-banking unquestionably overcoming the traditional banking. However, this has created new exposures to different kind of risks to banks, and customers. As a result, banks are expanding their risk management practices in order to deter, detect, monitor and mitigate risks posed by online banking. Basel [1] also issued 14 principles of risk management for online banking to assist banking organizations to expand their risk management strategies and procedures to insure the counter-productivity of the online banking activities. These 14 principles fall under three categories of issues: Board and management oversight, security controls, and legal and reputational risk management.

The aim of this research is to study and investigate to what degree risk managers of Bahraini banks understand and apply risk management based on the 14 principles identified by Basel, and to discover the perception of risk managers in Bahraini banks related to the risks posed by online banking, in order to provide a step towards crafting risk management strategy. The researchers firmly believe the research findings of this research will be useful for banks of Bahrain in crafting risk management strategy related to online banking. The study will address How do banks in Bahrain perceive risks associated with online banking? To what extent are banks applying risk management in online banking based on the 14 principles prescribed by Basel committee?

2 Literature Review

Al-Soufi [3] defines internet banking as delivering banking services over the internet. On the other hand, AlSoufi and Ali [4] noted that mobile banking and online banking are both forms of electronic banking or e-banking, but they only differ on the channel of banking service delivery. Whereas online banking uses computers connected to the internet to transact, mobile banking uses wireless devices to do transactions. Sethi [5] defines e-banking as the use of technology to perform banking transactions electronically without visiting a mortar and brick institution. Therefore, mobile banking and online/Internet banking are all forms of e-banking. Thus, this paper will use online banking, mobile banking and internet banking interchangeably.

2.1 Opportunities and Challenges of Online Banking

Cheng et al. [6] found out that the rapid adoption of online banking, which is part of e-commerce, by the banking sectors is attributable to the fact that their products can easily be virtualized. The products of the banking industry are of more priority than the place where banking services are being offered [6]. Banks and other financial institutions generate their revenue by selling credit cards and giving out loans as well as account access fees. Therefore, another reason why the banking sector rapidly adopts online banking is the fact that it can generate more revenue by easily selling those products plus fees of frequent account access [6].

Despite the rapid growth of the banking sector due to online banking, banks and customers have serious concerns about the security of using the internet as a delivery channel. Banks are currently actively promoting their fundamental abilities such as advising clients through the internet, but due to the relative novelty, accessing customer accounts through the internet poses a major security risk, which is a very big challenge. A study that explored the quality of online banking in the Kingdom of Bahrain discovered that most customers perceive the benefits and risks of online banking [3].

A study by Labrecque et al. [7] found out that most customers today are more and more looking for services they can access from a single-entry point. Banks have also taken this consumer behavior as a competition avenue [7]. Banks tend to gain a competitive advantage by providing services to their customers at a single-entry point. This has, in turn, motivated banks to move obstinately in launching coalitions and joint ventures to uphold their privilege to online banking. A study by Alwan and Al-Zubi in [8] on Jordanian banking sector found out that, owing to the significant changes in the availability of spare time among individuals, most people have changed their consumer behavior in banking. A Bangladesh-based study by Sattar Titu [9] discovered that most of consumers prefer online banking due to its flexibility, mobility, and independence of time and place. This is the reason banks are aggressively moving in seeking alliances in online banking.

Most of the governments today encourage taxpayers to submit their tax returns using electronic means result in new opportunities to internet banking. For example, in the United States, 98.6 million individual income tax returns were filed using an e-file or electronic means [10]. This is owed to the rapid technology adoption and diffusion in the United States. Even though there is no data about the number of electronic tax return filings in the Kingdom of Bahrain as there is no income tax or other taxes, but government started value added tax from January 1, 2019 which will boost e-banking further. Another major challenge facing online banks is regulatory barriers [11]. Without proper regulations, online banking might cause damage to its stakeholders. Regulations are prerequisite because of the enormous acceptance of online banking across the world [11]. Nonetheless, the internet as an independent entity has resulted in the continuous emergence of new rivals in the market from all directions [11]. Besides security concerns by online consumers, several recent studies suggest that the security measures for internet banking are likely to reduce its acceptability because most of them cause significant inconveniences to users [12–14]. Anti-hacking security features, particularly, lower the acceptance rate of online banking [14]. Maditinos et al. [13] suggest that there is a need to differentiate between theoretical security and adequate security. Effective anti-hacking security features often lower the usability and acceptability of internet banking [13]. Theoretical security is often perceived as more robust and crucial to the usability of e-banking systems, but when practically applied, it is, of course, capable but slightly lower than theoretical security and lowers the usability and acceptability of online banking [13].

2.2 The Principles of Risk Management of Online Banking Provided by Basel

A deeper insight into what each of the 14 principles means will help in determining the incongruity or consistence in risk management practices in internet banking in the Kingdom of Bahrain. Risk management refers to the process of identifying all risks and putting them under control to keep all processes in good working order [15]. The 14 principles are categorized into three. The first category A is 'Board and Management Oversight' (Principle 1–3). Under this category, the Basel guarantees the Board and Management of any banking institution the authority to determine whether the bank to offer e-banking transaction services or not. The Board's decision-making about e-banking plans should be aligned with the banking establishment's broader business strategy—corporate strategic goals [1]. If the e-banking plans are well suited to the organizational strategic goals, the Board should go ahead and carry out a careful risk analysis before implementing the plans. Afterward, correct risk mitigation strategies should be put in place and implemented. Continuous monitoring of the e-banking plans should resume after implementation to assess if the project is exactingly corresponding with corporate strategy and yielding desired results. In order to ensure this, Basel proposed the following:

Principle 1 requires the Board and Management to come up with an effective management oversight over e-banking risks plus risks that e-banking might cause on the traditional banking system.

Principle 2 is all about security control process whereby the Board of Directors (BoD) is required to evaluate and commend the core features of the bank's security control process.

Principle 3 touches on third-party dependencies and outsourcing relationships supporting e-banking. It commends the BoD to come up with a comprehensive and constant due meticulousness and monitoring process for managing these third parties and relationships.

In Category B, 'Security Controls' (Principle 4–10), the Basel acknowledged that e-banking can pose enhanced security challenges that might not only need BoD to ensure that correct security control processes are in place, but also special management attention is given. The Basel outlined the following principles to ensure this:

Principle 4 is about proper authentication and identification of customers with whom the institution is conducting business. This should take into account hacking and corruption which can result in deletion of customers' important information from the database, displacement of customers, and insecurity of e-banking sessions.

Principle 5 emphasizes the importance of accountability of customers' transactions by using proper transaction authentication methods and promoting non-repudiation.

Principle 6 is about segregation of duties within e-banking systems, databases, and applications, which is intended to reduce the risk of fraud.

Principle 7 outlines how proper authorization within e-banking systems, databases, and applications can be conducted. This is related to Principle 6 because properly controlled authorization and access privileges are the only ways of ensuring appropriate segregation of duties.

Principle 8 elaborates how the data integrity of e-banking systems, databases, and applications can be protected optimally. This includes ensuring data in storage or in transit is not altered with by an unauthorized party.

Principle 9 touches on the criticality of ensuring that all audit trails for all transactions done through e-banking, especially through the internet exist.

Principle 10 closes this category by elaborating why it is important to preserve the confidentiality of all the information in the e-banking system. The measures taken should be proportionate to the sensitivity of the information stored or in transit in the e-banking system.

In category C, 'Legal and Reputational Risk Management' (Principle 11–14), the Basel concluded that the e-banks should observe the same laws and regulations to those of traditional banks. The overall responsibility given to banks includes e-banking consumer information protection, informing e-banking users, and business availability. To ensure this, the following principles are paramount:

Principle 11 requires e-banks to display adequate information on their websites so that users can make informed decisions when conducting transactions.

Principle 12 requires e-banks to uphold the privacy of the user under the laws and regulations of a jurisdiction under which the bank is operating.

Principle 13 outlines that e-banking systems should always be available to customer by coming up with appropriate measures that will ensure banks have planning processes that have effective capacity, business continuity and contingency.

Principle 14 requires banks to have appropriate and effective incident response plans because with the novelty of e-banking, unusual and unexpected incidents can arise any time.

Overall, the Basel Committee gave more details about each principle, and the above descriptions are just highlighting for each principle. The Committee has recommended best e-banking practices for each principle that was not thoroughly covered in this paper. Besides, detailed appendices and references were also provided by the Committee. Banks with e-banking plans should consult this publication as it contains what it means to be risk-free regarding internet banking.

3 Research Methodology

This study used both qualitative and quantitative methods. As for quantitative method, data were collected through survey questionnaire to address the degree to which the banks in Bahrain applied the 14 principles. The questionnaire included 14 close-ended questions aimed at evaluating the application of each principle. However, qualitative data were obtained through semi-structured interview with 3 participants from different commercial banks. The purpose of conducting these interviews was to get a better perspective to the idea of risk management and online banking. Thus, information obtained through interviews gave value and depth to the data gathered by the survey. Data were obtained from survey questionnaires that have a response rate of 93.85%, which has been answered by 122 risk managers or other online banking staff, after that, data were statistically analyzed by SPSS version 25 to discover how various banks in Kingdom of Bahrain apply the risk management principles suggested. A total of 130 survey questionnaires were distributed in 9 commercial banks of Bahrain, 122 questionnaires were received. However, only 119 questionnaires met the criteria and were used for this study. The 3 remained were not used due to incompletion of the questionnaire.

4 Data Analysis and Interpretation of Results

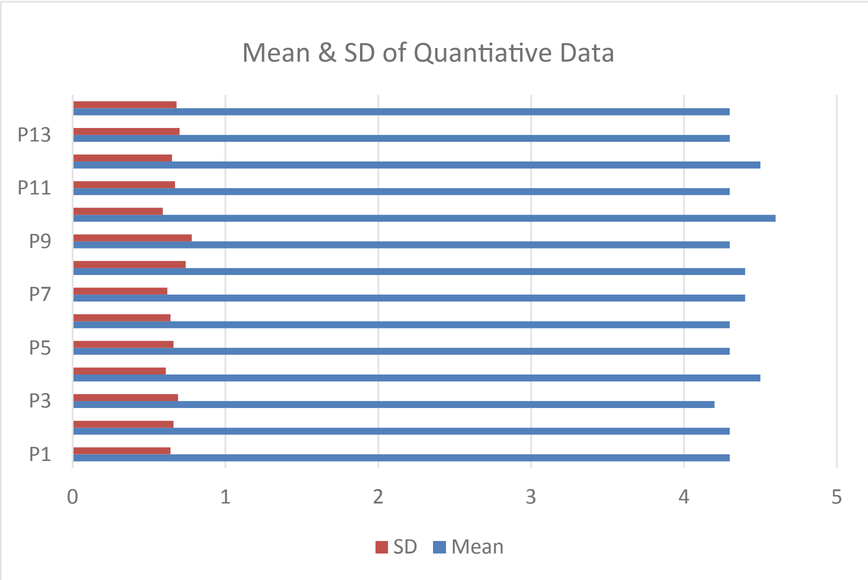
4.1 Quantitative Data Analysis

The extent to which banks applied risk management principles in online banking proposed by Basel was explored through three main categories: Board and Management Oversight, Security Control, and Legal and Reputational Risk. Participants in the survey were asked to respond according to their level of agreement to the principles. The results of agreements are shown in Table 1.

Likert scale 1 is strongly disagree, 2 disagree, 3 is neutral, 4 is agree and 5 strongly disagree.

Table 1 Quantitative data

	Response (%)					Mean	Std. deviation
	1	2	3	4	5		
Category A. Board and Management Oversight:							
Principle 1	0	0.8	6.7	52.1	40.3	4.3	0.637
Principle 2	0	1.7	5.9	52.1	40.3	4.3	0.661
Principle 3	0	0.8	12.6	51.3	35.3	4.2	0.687
Category B. Security Controls:							
Principle 4	0	0.8	3.4	41.2	54.6	4.5	0.609
Principle 5	0	0.8	8.4	49.6	41.2	4.3	0.661
Principle 6	0	0	10	51.3	38.7	4.3	0.64
Principle 7	0	0	6.7	46.2	47.1	4.4	0.615
Principle 8	0	1.7	10.1	37.8	50.4	4.4	0.735
Principle 9	0.8	0.8	12.6	41.2	44.5	4.3	0.78
Principle 10	0	0.8	2.5	35.3	61.3	4.6	0.59
Category C. Legal and Reputational Risk Management:							
Principle 11	0	0.8	9.2	50.4	39.5	4.3	0.666
Principle 12	0	0.8	5.9	39.5	53.8	4.5	0.648
Principle 13	0	1.7	8.4	45.4	44.5	4.3	0.702
Principle 14	0	0.8	10.9	49.6	38.7	4.3	0.682



Respondents perceived that principle 10 highly complied with mean value 4.6 and principle 3 lowest complied with mean value 4.2. The responses in principle 9 has high variation with 0.78 SD and in principle 4 with 0.61 SD.

4.2 Qualitative Data Analysis

The qualitative data collected using semi-structured interviews as a pilot study to further investigate the perception and application of risk management in Kingdom of Bahrain. A similar study was done by Abdou et al. [16] to determine the perception of risk management for online banking in United Kingdom banks.

Two risk managers and one online banking specialist participated in this interview as the interview consisted of 5 questions that each participant responded to willingly.

Question 1: What is the main benefit of providing online banking services to the customers?

All participants agree that accessibility of online banking services anytime anywhere is the main benefit of online banking.

Question 2: What are the main online banking risks that the bank experience?

Participant A states that hacking is one serious risk that exposes confidential data of its customers for manipulation. Another common risk is phishing, as phishers may try to build a website like the bank's website and try to drag the customers to access it and steal their credentials. He admitted that he himself faced such problem.

Participant B states that the main risk is losing the confidentiality of information that the customer confided the bank with.

However, participant C while he agrees that the main risk is hacking, he adds that the reputation of the bank is highly at risk when it comes to online banking, because if risk management isn't doing their work right, this may put the bank at legal risk when customers private information goes to public. Participant A adds that the risk of system failure also affects the quality of services.

Question 3: How does the bank deal with risks related to online banking?

Participant A states that the bank must apply intense technical security measures on its platforms in order to safeguard its systems and ensure that they are effectively protected, as they prefer to mitigate the risks rather than dealing with them. Participant B states that they are continuously updating their systems to be in line with the latest developments in the market to prevent viruses and hackers, and to maintain the availability of their online banking systems. Participant C however emphasize on the importance of authentication of the customers and the integrity of the transactions as the main part to succeed in managing risks arising from online banking. Participants A and C argues that because phishing happens from the side of the customer, the bank must provide awareness for its customers for such risk and how to deal with it.

Question 4: What are the risks that you believe would highly affect customers using such service?

All participants agree that hacking the computers of the customers and stealing their credentials might be highly affecting customers using such service, as they might get in the risk of being stolen.

Question 5: Do you believe your bank applied risk management of online banking successfully?

All participants agreed that the application of risk management in their banks is successful, and that very strict measures are applied in order to protect the privacy and confidentiality of information.

Overall, researchers learned from the interview that all participants perceive risks of online banking as a security risk, and risk management as maintaining the privacy of information.

5 Conclusion

Realizing the significance of risk management in online banking, this study sought to investigate the degree to which banks in Bahrain are applying the risk management principles provided by Basel; and how they (the banks) perceive risk as far as online banking is concerned, which is in line with the findings by Al-Soufi and Ali [4] and Sethi [5]. The researchers in the current study keenly deliberated the responses and found that risk manager have a positive perception of risks associated with online banking. However, mean values of survey questionnaire is 60.89, which means that they are positively utilizing Basel's principles framework. The results also show that the risk managers believe that risk management for online banking in Bahrain is successful and they are operating with the principles identified in Basel.

Regarding the relevance of each principle regarding online banking in Bahrain, the study revealed that the banks to work more on principle 3, principle 9, and principle 13. Regarding categories of principles of risk management, researcher found that category B is highly implemented than category A and category C. Where the results showed that the mean value of category B was 4.39, while the mean value of category A was 4.28, and the mean value of category C was 4.33. Unlike the generalized approach employed in Wissem [15], the scores above shows that risk management in banks in Bahrain pour their attention mostly towards security control, as they perceive that these controls are mainly what keeps customers continuing to conduct their transactions online. It is significant to identify that the mean value for principle 4 and that of principle 10 are higher than the rest of the principles, where principle 4 has the mean value of 4.50, and principle 10 has the mean value of 4.57. While Nikam and Bhoite [11] identified that risk management in online banking is mainly employed as a growth strategy, these values reveal that their risk management perceive privacy and confidentiality as the main aspects of risk management. However, principle 3, principle 9, and principle 13 has the least values of 4.21, 4.28, and 4.26 respectively. Although the results are relatively good in these responses, risk management should pay more attention to these principles which is in line with the findings by Wissem [15]. In a nutshell, it can be concluded that risk managers in Bahrain's banks perceive the risks associated with online banking positively. On the same note, the results indicated that banks have succeeded in the application of the principles outlined in Basel.

References

1. Basel Committee on Banking Supervision.: Risk management principles for electronic banking and electronic money activities, Bank for International Settlements (2003)
2. Driga, I., Isac, C.: E-banking services—features, challenges and benefits. Ann. Univ. Petroşani, Econ. **14**(1), 49–58 (2014)
3. Al-Soufi, A.: Exploring the quality of internet banking services in kingdom of bahrain. Int. J. E-Educ., E-Bus., E-Manag. E-Learn. **3**(5), 410–415 (2013)
4. Al-Soufi, A., Ali, H.: Customer's perception of M-banking adption in kingdom of bahrain: an empirical assessment of an extended tam model. Int. J. Manag. Inf. Technol. **6**(1), 1–13 (2014)
5. Sethi, J.: Elements of Banking and Insurance. [S.l.]: Prentice-Hall of India (2012)
6. Cheng, T., Lam, D., Yeung, A.: Adoption of internet banking: an empirical study in Hong Kong. Decis. Support Syst. **42**(3), 1558–1572 (2006)
7. Labrecque, L., vor dem Esche, J., Mathwick, C., Novak, T., Hofacker, C.: Consumer power: evolution in the digital age. J. Interact. Mark. **27**(4), 257–269 (2013)
8. Alwan, H., Al-Zubi, A.: Determinants of internet banking adoption among customers of commercial banks: an empirical study in the jordanian banking sector. Int. J. Bus. Manag. **11**(3), 95–104 (2016)
9. Sattar Titu, M.: Online banking system-its application in some selected private commercial banks in Bangladesh. IOSR J. Bus. Manag. **9**(4), 37–44 (2013)
10. Collins, B.: Projections of Federal Tax Return Filings: Calendar Years 2011–2018, pp. 181–190. Internal Revenue Service (IRS) (2012)
11. Nikam, A.V., Bhoite, D.: A critical study of net-banking services in banks: strength and weaknesses. Int. J. Sci. Res. **2**(1), 152–156 (2012)
12. Komb, F., Korau, M., Belás, J., Korauš, A.: Electronic banking security and customer satisfaction in commercial banks. J. Secur. Sustain. Issues **5**(3), 411–422 (2016)
13. Maditinos, D., Chatzoudes, D., Sarigiannidis, L.: An examination of the critical factors affecting consumer acceptance of online banking. J. Syst. Inf. Technol. **15**(1), 97–116 (2013)
14. Twum, F., Ahenkora, K.: Internet banking security strategy: securing customer trust. J. Manag. Strat. **3**(4), 78–83 (2012)
15. Wissem, E.: Risk management: new literature review. Pol. J. Manag. Stud. **8**(1), 288–297 (2013)
16. Abdou, H., English, J., Adewunmi, P.: An investigation of risk management practices in electronic banking: the case of the UK banks. Banks Bank Syst. **9**(3), 1–24 (2014)

Deep Learning Techniques: An Overview

Amitha Mathew, P. Amudha, and S. Sivakumari

Abstract Deep learning is a class of machine learning which performs much better on unstructured data. Deep learning techniques are outperforming current machine learning techniques. It enables computational models to learn features progressively from data at multiple levels. The popularity of deep learning amplified as the amount of data available increased as well as the advancement of hardware that provides powerful computers. This article comprises the evolution of deep learning, various approaches to deep learning, architectures of deep learning, methods, and applications.

Keywords Deep learning (DL) · Recurrent neural network (RNN) · Deep belief network (DBN) · Convolutional neural network (CNN) · Generative adversarial network (GAN)

1 Introduction

Deep learning techniques which implement deep neural networks became popular due to the increase of high-performance computing facility. Deep learning achieves higher power and flexibility due to its ability to process a large number of features when it deals with unstructured data. Deep learning algorithm passes the data through several layers; each layer is capable of extracting features progressively and passes it to the next layer. Initial layers extract low-level features, and succeeding layers combine features to form a complete representation. Section 2 gives an overview of

A. Mathew (✉) · P. Amudha · S. Sivakumari
Department of Computer Science and Engineering, School of Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India
e-mail: amitha.ann@gmail.com

P. Amudha
e-mail: amudharul@gmail.com

S. Sivakumari
e-mail: sivakumari_cse@avinuty.ac.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_54

the evolution of deep learning models. Section 3 provides a brief idea about the different learning approaches, such as supervised learning, unsupervised learning, and hybrid learning. Supervised learning uses labeled data to train the neural network. In supervised learning, the network uses unlabeled data and learns the recurring patterns. Hybrid learning combines supervised and unsupervised methods to get a better result. Deep learning can be implemented using different architectures such as unsupervised pre-trained networks, convolutional neural networks, recurrent neural networks, and recursive neural networks, which are described in Sect. 4. Section 5 introduces various training methods and optimization techniques that help in achieving better results. Section 6 describes the frameworks which allow us to develop tools that offer a better programming environment. Despite the various challenges in deep learning applications, many exciting applications that may rule the world are briefed in Sect. 7.

2 Evolution of Deep Learning

First generation of artificial neural networks (ANNs) was composed of perceptrons in neural layers, which were limited in computations. The second generation calculated the error rate and backpropagated the error. Restricted Boltzmann machine overcame the limitation of backpropagation, which made the learning easier. Then, other networks are evolved eventually [15, 24]. Figure 1 illustrates a timeline showing the evolution of deep models along with the traditional model. The performance of classifiers using deep learning improves on a large scale with an increased quantity of data when compared to traditional learning methods. Figure 2 depicts the performance of traditional machine learning algorithms and deep learning algorithms [6]. The performance of traditional machine learning algorithms becomes stable when

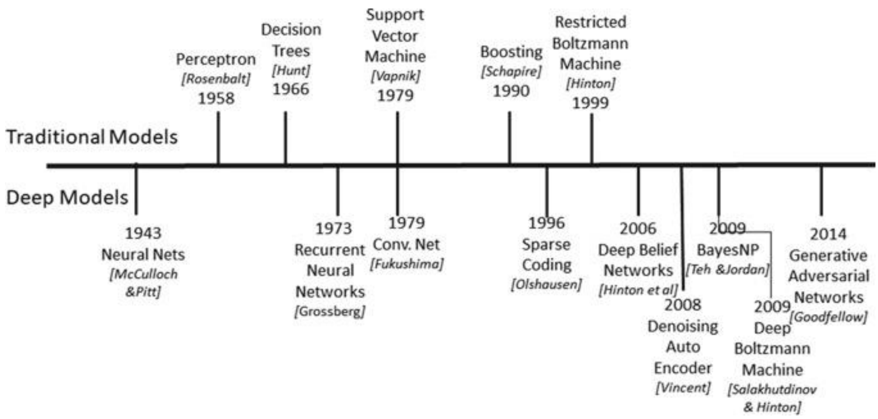
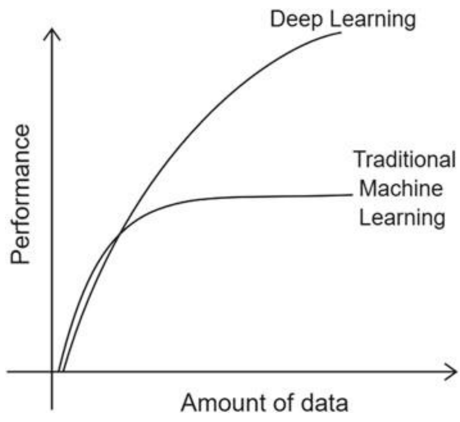


Fig. 1 Evolution of deep models

Fig. 2 Why deep learning?



it reaches the threshold of training data whereas the deep learning upturns its performance with increased amount of data. Nowadays, deep learning is used in a lot many applications such as Google's voice and image recognition, Netflix and Amazon's recommendation engines, Apple's Siri, automatic email and text replies, and chatbots.

3 Deep Learning Approaches

Deep neural networks are successful in supervised learning, unsupervised learning, reinforcement learning, as well as hybrid learning.

3.1 Supervised Learning

In supervised learning, the input variables represented as X are mapped to output variables represented as Y by using an algorithm to learn the mapping function f.

$$Y = f(X) \tag{1}$$

The aim of the learning algorithm is to approximate the mapping function to predict the output (Y) for a new input (X). The error from the predictions made during training can be used to correct the output. Learning can be stopped when all the inputs are trained to get the targeted output [11]. Regression is used for solving regression problems [18], support vector machines for classification [21], and random forest for classification as well as regression problems [20].

3.2 Unsupervised Learning

In unsupervised learning, we have the input data only and no corresponding output to map. This learning aims to learn about data by modeling the distribution in data. Algorithms can be able to discover the exciting structure present in the data. Clustering problems and association problems use unsupervised learning. The unsupervised learning algorithms such as K-means algorithm are used in clustering problems [9], and Apriori algorithm is used in association problems [10].

3.3 Reinforcement Learning

Reinforcement learning uses a system of reward and punishment to train the algorithm. In this, the algorithm or an agent learns from its environment. The agent gets rewards for correct performance and penalty for incorrect performance. For example, consider the case of a self-driving car: The agent gets a reward for driving safely to destination and penalty for going off-road. Similarly, in the case of a program for playing chess, the reward state may be winning the game and the penalty for being checkmated. The agent tries to maximize the reward and minimize the penalty. In reinforcement learning, the algorithm is not told how to perform the learning; however, it works through the problem on its own [16].

3.4 Hybrid Learning

Hybrid learning refers to architectures that make use of generative (unsupervised) as well as discriminative (supervised) components. The combination of different architectures can be used to design a hybrid deep neural network. They are used for action recognition of humans using action bank features and are expected to produce much better results [3].

4 Fundamental Deep Learning Architectures

Deep learning architectures perform better than simple ANN, even though training time of deep structures is higher than ANN. However, training time can be reduced using methods such as transfer learning and GPU computing. One of the factors which decide the success of neural networks lies in the careful design of network architecture. Some of the relevant deep learning architectures are discussed below.

4.1 Unsupervised Pre-trained Networks

In unsupervised pre-training, a model is trained unsupervised, and then the model is used for prediction. Some unsupervised pre-training architectures are discussed below [4].

Autoencoders: Autoencoders are used for the reduction of the dimension of data, novelty detection problems, as well as anomaly detection problems. In an autoencoder, the first layer is built as an encoding layer and transpose of that as a decoder. Then, train it to recreate the input using the unsupervised method. After training, fix the weights of that layer. Then, move to the subsequent layer until we pre-train all the layers of deep net. Then, go back to the original problem that we want to solve with deep net (classification/regression) and optimize it with stochastic gradient descent by starting from weights learned using pre-training. Autoencoder network consists of two parts [7]. The input is translated to a latent space representation by the encoder, which can be denoted as:

$$h = f(x) \tag{2}$$

The input is reconstructed from the latent space representation by the decoder, which can be denoted as:

$$r = g(h) \tag{3}$$

In essence, autoencoders can be described as in Eq. (4). r is the decoded output which will be similar to input x:

$$g(f(x)) = r \tag{4}$$

Deep Belief Networks: The first step for training the deep belief network is to learn features using the first layer. Then, use the activation of trained features in the next layer. Continue this until the final layer. Restricted Boltzmann machine (RBM) is used to train layers of the deep belief networks (DBNs), and the feedforward network is used for fine-tuning. DBN learns hidden pattern globally, unlike other deep nets where each layer learns complex patterns progressively [19].

Generative Adversarial Networks: Generative adversarial networks (GANs) were presented by Ian Goodfellow. It comprises generator network and discriminator network. Generator generates the content, while the discriminator validates the generated content. Generator creates natural-looking images, while the discriminator decides whether the image looks natural. GAN is considered as a minimax two-player algorithm. GANs use convolutional and feedforward neural nets [5].

4.2 Convolutional Neural Networks

Convolutional neural networks (CNNs) are used mainly for images. It assigns weights and biases to various objects in the image and differentiates one from the other.

It requires less preprocessing related to other classification algorithms. CNN uses relevant filters to capture the spatial and temporal dependencies in an image [12, 25]. The different CNN architectures include LeNet, AlexNet, VGGNet, GoogLeNet, ResNet, ZFNet. CNNs are mainly used in applications such as object detection, semantic segmentation, and captioning.

4.3 Recurrent Neural Networks

In recurrent neural networks (RNNs), outputs from the preceding states are fed as input to the current state. The hidden layers in RNN can remember information. The hidden state is updated based on the output generated in the previous state. RNN can be used for time series prediction because it can remember previous inputs also, which is called long short-term memory [2].

5 Deep Learning Methods

Some of the powerful techniques that can be applied to deep learning algorithms to reduce the training time and to optimize the model are discussed in the following section. The merits and demerits of each method are comprised in Table 1.

Backpropagation: While solving an optimization problem using a gradient-based method, backpropagation can be used to calculate the gradient of the function for each iteration [18].

Stochastic Gradient Descent: Using the convex function in gradient descent algorithms ensures finding an optimal minimum without getting trapped in a local minimum. Depending upon the values of the function and learning rate or step size, it may arrive at the optimum value in different paths and manners [14].

Learning Rate Decay: Adjusting the learning rate increases the performance and reduces the training time of stochastic gradient descent algorithms. The widely used technique is to reduce the learning rate gradually, in which we can make large changes at the beginning and then reduce the learning rate gradually in the training process. This allows fine-tuning the weights in the later stages [7].

Dropout: The overfitting problem in deep neural networks can be addressed using the dropout technique. This method is applied by randomly dropping units and their connections during training [9]. Dropout offers an effective regularization method to reduce overfitting and improve generalization error. Dropout gives an improved performance on supervised learning tasks in computer vision, computational biology, document classification, speech recognition [1].

Max Pooling: In max pooling, a filter is predefined, and this filter is then applied across the nonoverlapping subregions of the input taking the max of the values

Table 1 Comparison of deep learning methods

Method	Description	Merits	Demerits
Backpropagation	Used in optimization problem	For calculation of gradient	Sensitive to noisy data
Stochastic gradient descent	To find optimal minimum in optimization problems	Avoids trapping in local minimum	Longer convergence time, computationally expensive
Learning rate decay	Reduce learning rate gradually	Increases performance and reduces training time	Computationally expensive
Dropout	Drops out units/ connection during training	Avoids overfitting	Increases number of iterations required to converge
Max pooling	Applies a max filter	Reduces dimension and computational cost	Considers only the maximum element which may lead to unacceptable result in some cases
Batch normalization	Batch-wise normalization of input to a layer	Reduces covariant shift, increases stability of the network, network trains faster, allows higher learning rates	Computational overhead during training
Skip-gram	Used in word embedding algorithms	Can work on any raw text and requires less memory	Softmax function is computationally expensive, and training time is high
Transfer learning	Knowledge of first model is transferred to second problem	Enhances performance, rapid progress in training of second problem	Works with similar problems only

contained in the window as the output. Dimensionality, as well as the computational cost to learn several parameters, can be reduced using max pooling [23].

Batch Normalization: Batch normalization reduces covariate shift, thereby accelerating deep neural network. It normalizes the inputs to a layer, for each mini-batch, when the weights are updated during the training. Normalization stabilizes learning and reduces the training epochs. The stability of a neural network can be increased by normalizing the output from the previous activation layer [8].

Skip-gram: Word embedding algorithms can be modeled using skip-gram. In the skip-gram model, two vocabulary terms share a similar context; then, those terms are identical. For example, the sentences "cats are mammals" and "dogs are mammals" are meaningful sentences which share the same meaning "are mammals." Skip-gram can be implemented by considering a context window containing n terms, train the neural network by skipping one of the terms, and then use the model to predict skipped term [13].

Transfer Learning: In transfer learning, a model trained on a particular task is exploited on another related task. The knowledge obtained while solving a particular problem can be transferred to another network, which is to be trained on a related problem. This allows for rapid progress and enhanced performance while solving the second problem [17].

6 Deep Learning Frameworks

A deep learning framework helps in modeling a network more rapidly without going into details of underlying algorithms. Some deep learning frameworks are discussed below and are summarized in Table 2.

TensorFlow TensorFlow, developed by Google Brain, supports languages such as Python, C++, and R. It enables us to deploy our deep learning models in CPUs as well as GPUs [22].

Keras Keras is an API, written in Python and run on top of TensorFlow. It enables fast experimentation. It supports both CNNs and RNNs and runs on CPUs and GPUs [22].

PyTorch PyTorch can be used for building deep neural networks as well as executing tensor computations. PyTorch is a Python-based package that provides tensor computations. PyTorch delivers a framework to create computational graphs [22].

Caffe Yangqing Jia developed Caffe, and it is open source as well. Caffe stands out from other frameworks in its speed of processing as well as learning from images. Caffe Model Zoo framework facilitates us to access pre-trained models, which enable us to solve various problems effortlessly [22].

Deeplearning4j Deeplearnig4j is implemented in Java, and hence, it is more efficient when compared to Python. The ND4J tensor library used by Deeplearning4j

Table 2 Comparison of deep learning frameworks

Deep learning framework	Release year	Language written in	CUDA supported	Pre-trained models
TensorFlow	2015	C++ and Python	Yes	Yes
Keras	2015	Python	Yes	Yes
PyTorch	2016	Python and C	Yes	Yes
Caffe	2013	C++	Yes	Yes
Deeplearning4j	2014	C++ and Java	Yes	Yes

provides the capability to work with multi-dimensional arrays or tensors. This framework supports CPUs and GPUs. Deeplearnig4j works with images, csv as well as plaintext [22].

7 Applications of Deep Learning

Deep learning networks can be used in a variety of applications such as self-driving cars, natural language processing, Google's virtual assistant, visual recognition, fraud detection, health care, detecting developmental delay in children, adding sound to silent movies, automatic machine translation, text to image translation, image to image synthesis, automatic image recognition, image colorization, earthquake prediction, market rate forecasting, news aggregation, and fraud news detection.

8 Conclusion

Deep learning is continuously evolving faster; still, there are a number of problems to deal with and can be solved using deep learning. Even though a full understanding of the working of deep learning is still a mystery, we can make machines smarter using deep learning, sometimes even smarter than human. Now, the aim is to develop deep learning models that work with mobile to make the applications smarter and more intelligent. Let deep learning be more devoted to the betterment of humanity, thus making our domain a better place to live.

References

1. Achille, A., Soatto, S.: Information dropout: learning optimal representations through noisy computation. IEEE Trans. Pattern Anal. Mach. Intell. **40**(12), 2897–2905 (2018). https://doi.org/10.1109/TPAMI.2017.2784440
2. Bianchi, F.M., Maiorino, E., Kampffmeyer, M.C., Rizzi, A., Jenssen, R.: An overview and comparative analysis of recurrent neural networks for short term load forecasting (2017). arXiv:1705.04378
3. Deng, L., Dong, Y., et al.: Deep learning: methods and applications. Found. Trends® Signal Process. **7**(3–4), 197–387 (2014). https://doi.org/10.1007/978-981-13-3459-7_3
4. Erhan, D., Bengio, Y., Courville, A., Manzagol, P.-A., Vincent, P., Bengio, S.: Why does unsupervised pre-training help deep learning? J. Mach. Learn. Res. **11**, 625–660 (2010)
5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems, pp. 2672–2680 (2014)
6. Goyal, P., Pandey, S., Jain, K.: Introduction to natural language processing and deep learning. In: Deep Learning for Natural Language Processing, pp. 1–74. Springer, Berlin (2018). https://doi.org/10.1007/978-1-4842-3685-7_1
7. Hubens, N.: Deep inside: autoencoders—towards data science (2018)

8. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift (2015). arXiv:1502.03167
9. Jain, A.K.: Data clustering: 50 years beyond k-means. Pattern Recognit. Lett. **31**(8), 651–666 (2010). https://doi.org/10.1016/j.patrec.2009.09.011
10. Kotsiantis, S., Kanellopoulos, D.: Association rules mining: a recent overview. GESTS Int. Trans. Comput. Sci. Eng. **32**(1), 71–82 (2006)
11. Kotsiantis, S.B., Zaharakis, I., Pintelas, P.: Supervised machine learning: a review of classification techniques. Emerg. Artif. Intell. Appl. Comput. Eng. **160**, 3–24 (2007)
12. Le, G.V., et al.: A tutorial on deep learning part 2: autoencoders, convolutional neural networks and recurrent neural networks. Google Brain 1–20 (2015)
13. Liu, C., Li, Y., Fei, H., Li, P.: Deep skip-gram networks for text classification. In: Proceedings of the 2019 SIAM International Conference on Data Mining, pp. 145–153. SIAM (2019)
14. Lorraine, J., Duvenaud, D.: Stochastic hyperparameter optimization through hypernetworks (2018). arXiv:1802.09419
15. Miikkulainen, R., Liang, J., Meyerson, E., Rawal, A., Fink, D., Francon, O., Raju, B., Shahrzad, H., Navruzyan, A., Duffy, N., et al.: Evolving deep neural networks. In: Artificial Intelligence in the Age of Neural Networks and Brain Computing, pp. 293–312. Elsevier (2019). https://doi.org/10.1016/B978-0-12-815480-9.00015-3
16. Mnih, V., Badia, A.P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., Kavukcuoglu, K.: Asynchronous methods for deep reinforcement learning. In: International Conference on Machine Learning, pp. 1928–1937 (2016)
17. Sinno Jialin Pan and Qiang Yang: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2009). https://doi.org/10.1109/TKDE.2009.191
18. Panigrahi, A., Chen, Y., Kuo, C.C.J.: Analysis on gradient propagation in batch normalized residual networks (2018). arXiv:1812.00342
19. Salakhutdinov, R., Hinton, G.: Semantic hashing. Int. J. Approx. Reason. **50**(7), 969–978 (2009). https://doi.org/10.1016/j.ijar.2008.11.006
20. Scholkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press (2001)
21. Seber, G.A.F., Lee, A.J.: Linear Regression Analysis, vol. 329. Wiley (2012)
22. Sharma, P.: Top 5 deep learning frameworks, their applications, and comparisons! (2019)
23. Takahashi, T.: Statistical max pooling with deep learning, 3 July 2018. US Patent 10,013,644
24. Wang, H., Raj, B.: On the origin of deep learning (2017). arXiv:1702.07800
25. Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K.: Convolutional neural networks: an overview and application in radiology. Insights Imaging **9**(4), 611–629 (2018). https://doi.org/10.1007/s13244-018-0639-9

A Multilayer Deep Learning Framework for Auto-content Tagging

Shikha Jain, Ajit Kumar Jain, and Shashi Pal Singh

Abstract With the rapid and recent development of Internet of things (IoT), Bigdata, the most fundamental challenge is to explore the large volume of data from heterogeneous data sources (logs, audio, video, reports, news, social media, etc.). The tools and techniques that are being used in successful transforming of structured data into business intelligence simply do not work when it comes to unstructured data. An estimated 80% of data held by firms today are unstructured. We have powerful computations tools such as graphics processor units (GPUs) for complex calculations and Hadoop, spark for Bigdata management still technology is lacking in complete utilization of available data. Deep learning is a powerful tool to resolve the issues related to data, such as automatic extraction of complex data and representation of large volume unsupervised and uncategorized raw data. But extracted data do not have any standard format. In the survey of Bigdata mining system and IoT systems, it has been found that a layer is needed in Bigdata mining framework that can work on structured, unstructured and semi-structured data and can integrate the complete raw data. The proposed framework is a multilayer deep learning model that performed the clustering of unstructured huge text data followed by classification layer, CNN, to extract features and classify the raw data clusters in parallel manner. Conversion of unstructured data in standard format by automates content tagging helps decision-making systems to easily identify the related data and increase the data utilization.

Keywords Deep learning · Convolutional neural network (CNN) · Unstructured data (UD) · Text analytics · Bigdata

S. Jain (✉) · A. K. Jain
Banasthali Vidyapith, Banasthali, Rajasthan, India
e-mail: shikha150994@gmail.com

A. K. Jain
e-mail: ajitjain_2k@yahoo.co.in

S. P. Singh
AAI, Center for Development of Advanced Computing, Pune, India
e-mail: shashis@cdac.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_55

1 Introduction

Exponential growth of digital data brings opportunities for various fields such as medical, agricultural systems, expert systems, educational services, and commercial industry. It is reported by The National Security Agency that about 1.8 PB data is generating on the Internet every day. Google is the source of approximately 20,000 TB data and 3.6 TB data is generating by Flicker [1]. Data can be of any type like text, video, audio, images and stream data. Unfortunately, 85% of data exists in unstructured format [2]. According to a Forrester Research study, firms are currently able to analyze only 12% of available data. Information either is not organized in pre-defined manner or does not have well-defined data model [3]. Valuable information and knowledge are hidden in the large volume of data. Most of the industrial as well as research domains are interested in knowledge extraction from Bigdata so that it can be used in decision-making systems [4]. ML frameworks and traditional data mining concepts are very successful to produce higher accuracy result with structured data. But when it comes to the Bigdata which is unstructured and very complex to understand, only 20% of data is ready for processing. Rest 80% of data cannot be neglected as it can have crucial hidden information that can be very helpful for knowledge discovery.

Major issues that are existing are data volume, data variety and data velocity. Mostly data that is generating is real-time data, also it is huge in size, and it is not feasible to completely understand such a large data in less time. Other challenges and problems are that data is coming from heterogeneous sources that are very complex and unstructured in nature. ML algorithms are working well with structured data and semi-structured data but when unstructured Bigdata is using, they are lacking in accuracy as well as they need time to process that much amount of data. Many tools and techniques are available to work with unstructured data so that at least some information can be extracted. It is not easy to convert the unstructured data to structured data completely as lots of rules needed to be design and it can make the data preprocessing stage very time consuming. It can have a reverse effect on cost and time that will be gone in rule formation and data conversion [5].

Text mining has gained popularity in last few years, more often in the commercial rather than the academic area, probably because more than 80% business context including tweets, blogs, wikis and surveys. Text analytics also known as knowledge discovery in text (KDT) is an extension of data mining that tries to find textual patterns from large non-structured sources, as opposed to data stored in relational databases [6].

Conventional learning methods that have shallow architecture are not suitable to access such large text data, and deep learning is stand in need of it [7]. It calls attention to machine learning algorithms and techniques that are supervised/unsupervised strategies to access high-dimensional data and automatically learn hierarchical data representations.

The idea behind this proposed framework is the categorization of large amount of text data and assignment of tags to increase the data utilizations. Bigdata mining algorithms need good quality data. The proposed multilayer deep learning model classifies the unstructured data clusters so that tagged data volumes can be used in their respective domain areas.

2 Bigdata Mining

Bigdata has four characteristics, volume, velocity, variety and veracity. Large-scale data processing leads to more time complexity. Also, it is generating from heterogeneous resources and highly unstructured in nature. It is very difficult to maintain the quality of data and filtering the noise and irrelevant data. Many organizations are working on real-time data and they are spending most of the time in data preprocessing only. Hadoop/Spark have been introduced as the solutions of these issues still data mining algorithms are unable to extract the complete knowledge due to lack of standard format of data. This target can be achieved with classification and clustering techniques [5]. Conventional architectures and data mining algorithms are lacking in accuracy and they need large training time which is not acceptable in Bigdata scenario. In the era of deep learning and Bigdata, the urgent need for data analysis has revolutionized machine learning, statics and data mining. It can be seen in the growing number of researches, textbooks and applications that are dedicated to data analysis and data mining [4]. Deep learning concept and scaling-up machine learning algorithms have recently become an important topic for research that can deal with large datasets and automate the process of knowledge discovery. Researchers are adopting two approaches to address these problems:

(1) Practical approach: Working on robust platforms and frameworks like Hadoop/Spark and using machine learning and deep learning libraries with improved versions of data mining algorithms.

(2) Algorithms enhancement: Generating new versions of traditional ML algorithms to make them suitable for data analytics either by creating parallel versions or generating extended algorithms capable to support various machine learning and distributed frameworks.

Numerous traditional algorithms are well known for classification and clustering such as support vector machine (SVM), k-nearest neighbor (kNN), decision tree and k-means [8]. Also, a great number of powerful task-specific algorithms show higher accuracy on text data. These include conditional random algorithm for pattern recognition, Grivan-Newman community clustering algorithm used in text data clustering, latent Dirichlet allocation (LDA) used unsupervised learning to identify words and phrases, naïve Bayes classifier is very popular for text categorization, pointwise mutual information algorithm based on semantic orientation of phrases and porter stemmer algorithm reduces the word to its stem [2].

Recent versions of these algorithms are within easy reach to manage large heterogeneous dataset. Center of attention is high-dimensional data, core problem of machine learning. It implements data distribution to multiple machines with parallel execution of nearest neighbor algorithm to achieve fast nearest neighbor search

[9, 10]. An advanced kNN model implemented k-Tree that learns different optimal k-values for test samples. The key objective of this model is to improve classification performance. To find optimal k-value is a challenging task and not suitable for large number of samples [11, 12].

A map-reduce framework (MR-kNN) come up with an idea to split dataset into cluster and then apply the classification algorithm in parallel [13]. An extended version of MR-kNN implements spark for time reduction with high accuracy [14]. To deal with the distributed framework is somehow difficult as it required robust infrastructure and high computation power both.

A different set of work interested in scaling-up SVM as it shows very high classification accuracy [15]. An approach applied divide and conquer strategy. Clustering to partition the SVM kernel and then solve sub-problems independently and efficiently [8]. A research paper [16] presented quantum SVM that gain the logarithmic complexity of linear regression models. A generalized approach based on split and conquer method has been suggested in [8, 17]. Another challenge exists in clustering of large data introduced by [8] while scaling clustering techniques. Hence, neuro-fuzzy approaches can be used to utilize the data [17].

3 Deep Learning-Based Natural Language Processing

For decades, ML approach targeting natural language processing (NLP) problems based on shallow networks. Neural networks based on dense vector representation have been produced superior results on various NLP tasks such as text summarization, information retrieval, sentiment analysis, speech recognition and semantic matching. This trend is sparked by the success of word embedding and deep learning. Reference [7] concerns about basic principles for applying neural network NLP tutorial manner.

1. Word embedding: This concept based on the distributed representation of words and phrases with their compositionality [18]. Two effective models have been introduced [19] for word representation in high-dimensional vector space. CBOW is a fully connected network with one hidden layer used to predict target word, whereas Skip-gram models are different from CBOW as they predict surrounding context with the help of a small window of surrounding words. Both have their own limitation as CBOW are unable to identify phrases. Solution of this is, more recent methods like n-gram embedding from unlabeled data, widely used in text categorization. Learning vector representations are unsupervised method and generated labels used to train neural networks in NLP applications. Problem with Skip-gram introduced in [19], this concept is unsuited in sentiment analysis-based applications. Deep neural networks are demanded to overcome with these limitations.

2. Convolutional neural network: These networks hinge on sentence centric approach in NLP tasks. Convolutional networks have a need of lookup table to convert words into the corresponding vectors followed by feature extraction. CNN layers operate kernels to extract the specific pattern of n-gram phrase/sentence which proceeds toward max-pooling layer. Deep convolutional neural networks are effective for name-entity recognition, parts of speech (POS) tagging, word prediction as each word depends on its neighborhood [7]. After training for a task, the randomly initialized CNN became a specific n-gram feature detector [20]. However, CNN is unfitted to model long-distance dependencies and preserving sequential order standing as the main issue.

3. Recurrent neural network: These networks are efficacious to process sequential information and extensively used in language modeling, machine translation, multimodal sentiment analysis, etc. [21]. Various RNN models are available as simple RNN, long short-term memory (LSTM) and gated recurrent unit (GRU). RNN faces the issue of vanishing gradient when it comes to long sentences. LSTM and GRU are more efficient than simple RNN [22].

4. Recursive neural network: It represents sentence structure by parsing tree. Each non-terminal node in parsing tree is determined by its children representation. Based on recursive neural network, a phrase-level sentiment analysis framework has been putted in place [7]. Deep reinforced models and deep unsupervised learning are better than supervised learning. But they desired to carefully manage the state and action which in the end may restrict the learning capability of model. Also, recursive neural network needs more training than other neural networks.

4 Methodology

The proposed framework use word embedding to convert corpus words into high-dimensional sequential space. Word embedding based on semantic meaning of words so that similar words can be clustered together after conversion.

First phase is preprocessing that includes stop words removal, tokenization, creating different objects for input and output variable, and splitting the dataset into training and testing samples.

After preprocessing, the next phase is word embedding. If dataset is the collection of sentences then each sentence can be presented as S_i and corresponding vector set is $V_{i,}$ where $Si = \{w_1, w_2, w_3, \ldots, w_j\}$ and $Vi = \{v_1, v_2, \ldots, v_j\}$. In this equation, v_j ε R for all i, j ε N.

We have trained a news dataset having maximum sentence length 32 and post-sentence padding with word2vec tool, Fig. 1 visualizes that vector representation into 16-dimensional space.

After word embedding, in the third phase, these vectors will be feeded into the neural network for classification. Shallow neural networks are not efficient for large database as they lack in accuracy. Convolutional neural network has been used in the proposed framework and it is very good for binary classification. The idea is to use deep CNN to categorize news data into related category. In practical scenario, it is difficult to find related information. It needs long preprocessing and filtering to extract the required information. Auto-content tagging is very useful to find correct input data for decision-making systems. Word-vectors are input for the convolutional layers followed by a concatenation of vectors using bi-gram, tri-gram and n-gram

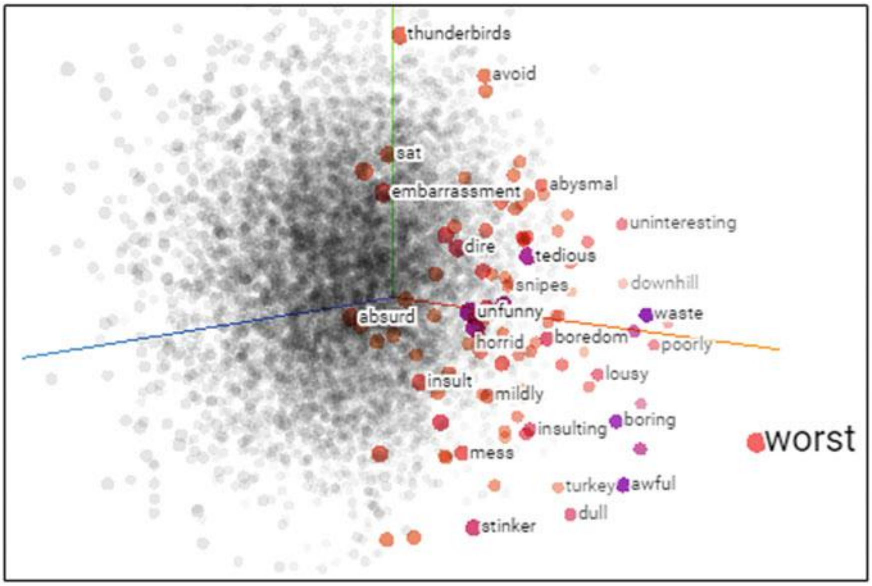


Fig. 1 Word representation in 16-dimensional vector space

combinations. Each layer output proceeds to max-pooling to extract high impact vector combinations. This strategy overcomes the CNN drawback of working with phrases. Every possible combination will be evaluated by CNN layers. At the end, final selected vectors will go the classifier as input and it generates the most suitable tag/category for the given sentence vectors. Figure 2 shows the complete working and architecture of multilayer deep framework.

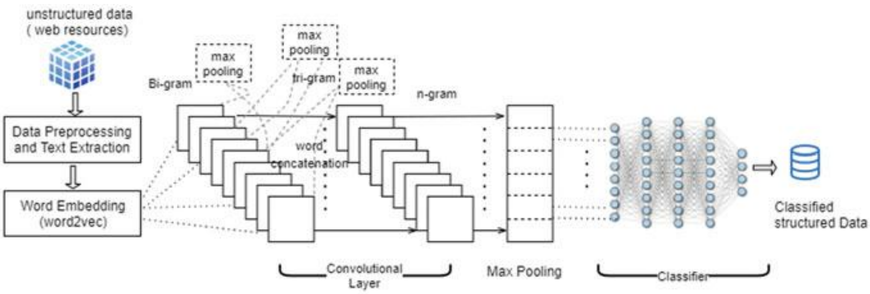


Fig. 2 A multilayer deep neural network architecture for text classification

5 Experimental Setup and Result Analysis

CNN outperform in binary text classification than any other neural networks. A news article dataset has been trained with deep neural networks and succeeds to achieve the validation accuracy level. Also, the loss is very low as compare to RNN and LSTM networks as shown in Fig. 3. An Intel i5 processor with 2 GHz power has

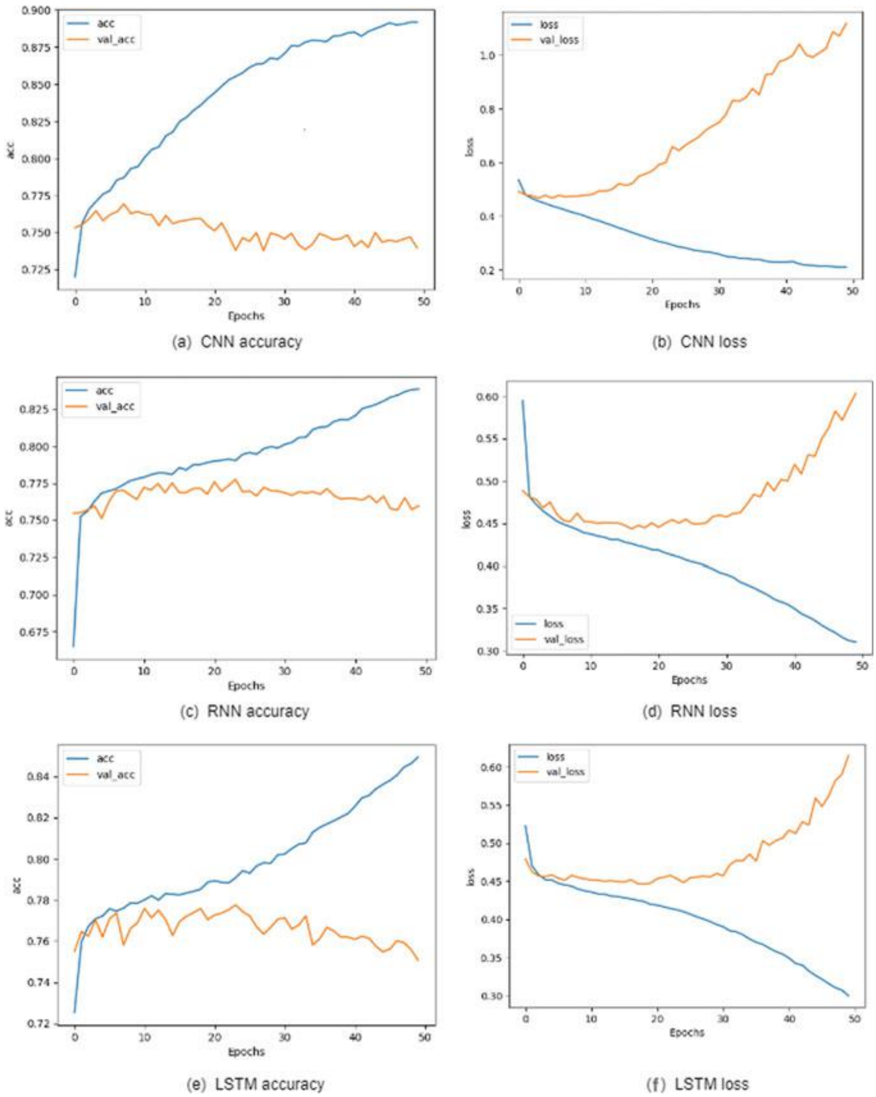


Fig. 3 Comparative analysis of convolutional neural network (CNN) with recurrent neural network (RNN) and long short-term memory (LSTM) variant of RNN for binary text classification

been used to train 3 GB wiki corpus for word embedding and news article dataset of size 2 MB classified with CNN, RNN and LSTM.

Metric parameters used to compile a model and judge its performance such as loss function and accuracy. Cross-entropy measures the performance of model as it predicts the divergence probability from actual label. Binary cross-entropy is used as loss function to measure the given classification model which output is a probability value between 0 and 1.

Cross-entropy $f(x)$ calculated as:

$$f(x) = -\sum_{n=1}^{N}\left(y_{o,n} + \log\left(p_{o,c}\right)\right) \tag{1}$$

where N—number of classes/tags, p—predicted probability and y—binary indictor (0 or 1) if class n is correct classification of observation o

Hence, binary cross-entropy $f'(x)$ is:

$$f^{'(x)} = -(y\log(p)) + ((1-y)\log(1-p)) \tag{2}$$

Accuracy metric computes the frequency with which y_{true} matches y_{pred}..Function need to take both values as argument and return a single tensor value. This return value or frequency is ultimately known as binary accuracy.

where y_{true}—true labels, y_{pred}—predictions.

Accuracy of CNN in 50 epochs is far better than expectations and it has taken less training time than LSTM. RNN is a bad practice for classification of large document as it generates result with low accuracy and needs more training time (Table 1).

Table 1 Comparative result analysis of CNN, RNN and LSTM for text classification

	Accuracy			Loss		
Epochs (50)	CNN	RNN	LSTM	CNN	RNN	LSTM
Round 1	0.889	0.849	0.867	0.191	0.325	0.284
Round 2	0.875	0.833	0.866	0.178	0.343	0.287
Round 3	0.885	0.854	0.842	0.157	0.291	0.297
Round 4	0.853	0.843	0.836	0.198	0.358	0.311
Round 5	0.899	0.841	0.877	0.199	0.298	0.292
Average	0.8802	0.844	0.8576	0.1846	0.323	0.2942

6 Application Areas of Proposed Framework

1. Marketing: Data is a hot topic in marketing literature. Many technologies are available to analyze the available data for sale predictions, understand customer views, etc. There exist a gap between theoretical researches and applications as capitalization of UD comes under "dark analytics" [23]. The aim of proposed work is to fill that gap and apply ML to underdeveloped areas that are UD search. It can be implemented in advance marketing researches.

2. Bigdata mining: Bigdata is highly unstructured in nature and it is proliferating rapidly. To analyze raw data is not easy and due to the lack of any standard format, accuracy of analytics tools could show inaccurate results [1]. The proposed framework classifies the raw data into various categories and assigns a tag to them. After this, it will be easy to utilize the data of specific domain instead of working with complete irrelevant data.

3. Decision support systems (DSS): The rapid emergence of text analytics leads to various research of new tools and technologies. Today, we have different DSS for different domains and they are hungry of quality data. These systems show highly accurate results when working with structured and classified data. But on the other hand, their predictions may fail with raw data [4]. Hence, a model required to present the raw data into its related category.

7 Future Work

Convolutional networks are very good for binary classification and short sentences but when it comes to large dataset, LSTM are better for NLP tasks. Deep convolutional networks are very efficient for polarity decisions and sentiment analysis but unsuited for long sentence learning. To classify the text in multiple classes, stand in need of more appropriate network that can learn long representations. LSTM has a deep architecture that required more computation power like GPUs [24]. So, future direction to address this issue is to create new learning frameworks and computing infrastructures with LSTM and implement that for large dataset classification with GPU unit.

8 Conclusion

It has been shown that CNN can be implemented on NLP task such as text classification and categorization. But to categorize the content in multiple classes is beyond the scope of CNN. LSTMs or deep CNN is more suitable for that. Deep CNN is a fusion model of deep neural architecture and classifier that can work in step wise manner even if corpus is large and training time is also less as compared to others. Only breakpoint where it is lacking is with long sentences understanding and multilevel classification. Recurrent neural networks and long short-term memory unit both need large corpus for training and high computation power. Hence, the proposed framework can be efficiently used in content tagging if sentences are of small length. Output of this framework will be the structured text dataset with assigned categories. This dataset can be used in business intelligence as well as in decision support system as input. This process is time-saving activity as it does not need to pre-process the data again and filtering the related structured data from the structured one.

References

1. Oussous, A., Benjelloun, F.Z., Lahcen, A.A., Belfkih, S.: Big data technologies: a survey. J. King Saud Univ.–Comput. Inf. Sci. 30, 431–448 (2018)
2. Salminen, J., Yoganathan, V., Corporan, J., Jansen, B.J., Jung, S.G.: Machine learning approach to auto-tagging online content for content marketing efficiency: a comparative analysis between methods and content type. J. Bus. Res. 101, 203–217 (2019). https://linkinghub.elsevier.com/retrieve/pii/S0148296319302607
3. The Mapping Process of Unstructured Data to Structured Data. In: 3rd International Conference on Research and Innovation in Information Systems–2013 (ICRIIS'13)
4. Zhanga, Q., Yang, L.T., Chenc, Z., Lic, P.: A survey on deep learning for big data. https://doi.org/10.1016/j.inffus.2017.10.006
5. Chen, X.W., Lin, X.: Big data deep learning: challenges and perspectives. Digital Object Identifier https://doi.org/10.1109/access.2014.2325029
6. Sandoval A.M., Redondo, T.: Text analytics: the convergence of big data and artificial intelligence (2016). ISSN 1989-1660. https://www.ijimai.org/journal/node/940
7. Larochelle, H., Bengio, Y., Louradour, J., Lamblin, P.: Exploring strategies for training deep neural networks. J. Mach. Learn. Res.1 Submitted 12 July 2009
8. Bolon-Canedo, V., Remeseiro, B., Sechidis, K.: Algorithmic challenges in big data analytics. In: ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26–28 April 2017, i6doc.com publ., ISBN 978-287587039-1. Available from: http://www.i6doc.com/en/
9. Maillo, J., Luengo, J., Garc´ıa, S., Herrera, F., Triguero, I.: Exact fuzzy k-nearest neighbor classification for big datasets. https://doi.org/10.1109/fuzz-ieee.2017.8015686

10. Hassanat, A.: Norm-based binary search trees for speeding up KNN big data classification. Computers **7**(54) (2018). https://doi.org/10.3390/computers7040054
11. Deng, Z., Zhu, X., Cheng, D., Zong, M., Zhang, S.: Efficient kNN classification algorithm for bigdata. https://www.sciencedirect.com/science/article/pii/S0925231216001132?via%3Dihub
12. Al-Jarrah, O.Y., Yoo, P.D., Muhaidat, S., Karagiannidis, G.K., Taha, K.: Efficient machine learning for big data: a review. Elsevier (2015) https://www.journals.elsevier.com/big-data-research
13. Maillo, J., Triguero, I., Herrera, F.: A MapReduce-based k-nearest neighbor approach for big data classification. Published in: 2015 IEEE Trustcom/BigDataSE/ISPA, https://doi.org/10.1109/trustcom.2015.577, Electronic ISBN: 978-1-4673-7952-6, USB ISBN: 978-1-4673-7951-9
14. Maillo, J., Ramírez, S., Triguero, I., Herrera, F.: kNN-IS: an iterative spark-based design of the k-nearest neighbors classifier for big data. J. Maillo et al./Knowl.-Based Syst. 117, 3–15 (2017)
15. Muja, M., Lowe, D.G.: Scalable nearest neighbor algorithms for high dimensional data. IEEE Trans. Pattern Anal. Mach. Intell. **36**(11) (2014)
16. Chen, L., Xu, Z., Wang, H., Liu, S.: An ordered clustering algorithm based on K-means and the PROMETHEE method. https://doi.org/10.1007/s13042-016-0617-9
17. Gallego, A.J., Calvo-Zaragoza, J., Valero-Mas, J.J., Rico-Juan, J.R.: Clustering-based k-nearest neighbor classification for large-scale data with neural codes representation. https://doi.org/10.1016/j.patcog.2017.09.038
18. Minmin Chen, M.: Efficient vector representation for documents through corruption. Criteo Research Palo Alto, Ca 94301, USA, Under Review As A Conference Paper At Iclr (2017)
19. Mikolov, T., Sutskever, I., Chen, K.: Distributed representations of words and phrases and their compositionality. Google Inc
20. Majumder, N., Poria, S., Gelbukh, A., Cambria, E.: Deep learning-based document modeling for personality detection from text. In: 2017 IEEE, IEEE Intelligent Systems Published by the IEEE Computer Society. 1541-1672/17/$33.00
21. Palangi, H., Deng, L., Shen, Y., Gao, J., He, X., Chen, J., Song, X., Ward, R.: Deep sentence embedding using long short-term memory networks: analysis and application to information retrieval. IEEE/ACM Trans. Audio, Speech, Lang. Process. https://doi.org/10.1109/taslp.2016.2520371
22. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. Google
23. Briggs, B., Hodgetts, C.: An overview. Wall Str. J. Tech trends 2017. Retrieved on 5 Sept 2017. http://deloitte.wsj.com/cio/2017/02/08/tech-trends-2017-an-overview/
24. Schlegel, D.: Deep machine learning on Gpu. University of Heidelber-Ziti, 12 Jan 2015

Case-Based Reasoning (CBR)-Based Anemia Severity Detection System (ASDS) Using Machine Learning Algorithm

Sumana De and Baisakhi Chakraborty

Abstract Anemia is a very common disease around the world. Based on the patient's signs, symptoms and hemoglobin amount of blood, the severity of anemia can be detected. This paper proposes an Anemia Severity Detection System (ASDS). The ASDS is designed to help the doctors or medical experts to detect the severity of anemia in a patient. It is developed based on Case-Based Reasoning (CBR) methodology and machine learning algorithm. CBR is one of the popular artificial intelligence techniques that uses past experiences to derive results for new cases, and it works in a cycle that includes four activities: Retrieve, Reuse, Revise and Retain. As a machine learning algorithm, K-Nearest Neighbor algorithm is used to detect whether the anemia is mild or severe. The dataset which helps to develop ASDS is from the Website of IUScholarWorks. As a screening tool, the graphical user interface (GUI) of ASDS has been developed with the help of Python, so that doctors or medical experts can access the system through the GUI. To improve accuracy and reduce training time, data preprocessing technique and feature selection technique have been applied on the dataset. Experimental results show that after using both techniques, K-Nearest Neighbor model gives highest 92% accuracy, 84% precision and 90% recall results.

Keywords Anemia Severity Detection System (ASDS) · CBR · K-Nearest Neighbor algorithm · User query · Diagnosis feedback

1 Introduction

Nowadays, anemia becomes a very common disease in the world. It is a condition in which hemoglobin amount and the number of healthy RBC in a person's blood drop below normal. When the number of healthy RBC in the blood decreases than normal,

S. De (✉) · B. Chakraborty
Department of Computer Science and Engineering, National Institute of Technology, Durgapur, India
e-mail: sumanade@gmail.com

B. Chakraborty
e-mail: baisakhichak@yahoo.co.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_56

then the person does not have enough healthy RBC to carry adequate oxygen to the body's tissues, in consequence of which the person feels fatigue and faces many other signs and symptoms. There can be many types of anemia, some caused by sudden blood loss, some caused by vitamin and mineral deficiency, some caused by heredity or some diseases that affect red blood cell production or destruction. Among these types of anemia, some of them are considered as mild anemia or not severe anemia and some are severe anemia. Based on the patient's signs, symptoms and hemoglobin amount of blood, the severity of anemia is measured. According to World Health Organization, anemia can be classified as mild anemia, moderate anemia or somewhat serious anemia and severe anemia, as referred to in [1]. For pregnant women, mild anemia when the hemoglobin level is in between 9.0 and 10.9 g/dL, moderate anemia when the hemoglobin level is in between 7.0 and 8.9 g/dL, and severe anemia when the hemoglobin level is less than 7.0 g/dL. Treatments for different types of anemia are different. The patients with mild anemia were given oral iron therapy. As it is very important to know how severe patient's anemia is, this paper proposes an Anemia Severity Detection System (ASDS). The ASDS is designed to help the doctors or medical experts to detect the severity of anemia in a patient. It is developed based on Case-Based Reasoning (CBR) methodology and machine learning algorithm. CBR is one of the popular artificial intelligence techniques that uses past experiences to derive results for new cases, and it works in a cycle that includes four activities, Retrieve, Reuse, Revise and Retain, as referred to in [2]. In this paper, the CBR cycle works as follows:

- Retrieve: Whenever ASDS accepts new inputted values, then it considers it as a new case. Next, ASDS uses machine learning algorithm to calculate similar cases and retrieves the solution or diagnosis result from casebase.
- Reuse: The new case is solved reusing retrieved solution or diagnosis result.
- Revise: Administrator of ASDS takes the responsibility to test applicability in the real world of the new proposed solution with new case.
- Retain: The new case is to be stored in the casebase with that new solution for a future problem solution.

Case is an encapsulation of problem description and solution. Problem description means case attributes with their values. And solution means the solution of that particular problem description. CaseBase (CB) is one type of database that stores such cases. In ASDS, the CB holds anemia cases. Here, each case's problem descriptions hold case attributes (patient age, symptoms, hemoglobin level, etc.) with their values. And solutions hold the target attribute with two values: mild anemia and severe anemia. The dataset that helps to develop ASDS is taken from the Website https://scholarworks.iu.edu/dspace/handle/2022/21181. It contains anemia survey of 200 patients with 25 attributes (e.g., rural/town, hemoglobin, age). To improve accuracy and reduce training time, data preprocessing technique and feature selection technique have been applied to the dataset. After using both techniques, cases are made and are stored into CB. In the ASDS, CB contains a total of 200 anemia cases with 12 case attributes and one target attribute or solution.

To evaluate solution, K-Nearest Neighbor algorithm has been used. K-Nearest Neighbor is a supervised machine learning algorithm. In the CB, 70% of cases are used to train the K-Nearest Neighbor algorithm and rest of 30% of cases are applied to the K-Nearest Neighbor model to evaluate accuracy, precision and recall result. Experimental results show that K-Nearest Neighbor model gives highest 92% accuracy, 84% precision and 90% recall results. As a screening tool, the graphical user interface (GUI) of ASDS has been developed with the help of Python. So, doctors or medical experts can submit patient's detail in ASDS, and the system can provide feedback through the GUI.

The remainder of the paper is organized as follows: Section 2 focuses on related previous works, Sect. 3 depicts detail of proposed ASDS model, and Sect. 4 shows the CB in ASDS. Section 5 focuses on methodology and system implementation, Sect. 6 depicts the accuracy comparison, Sect. 7 shows the simulation result, and at last, Sect. 8 ends up with conclusion and future work.

2 Related Previous Works

Expert systems are used in medical fields of diagnosis. It is also well applied in anemia diagnosis to help medical practitioners in diagnosis of anemia and provide corrective treatments. The research paper [3] proposed a fuzzy expert system to detect anemia on the basis of symptoms as well as clinical test. Developing the expert system using fuzzy logic helps engineers in detection of the diseases. The paper [4] aims to develop an intelligent expert system based on fuzzy logic and case-based reasoning for classification of hematology malignancy. The research paper [5] proposed a non-invasive automatic detection of anemia using image processing and pattern recognition techniques. The technique is developed to detect anemia by a digital photograph of a face image exposing conjunctiva as anemia can be identified by conjunctiva pallor. In [6], a fuzzy logic is approached to propose a model to diagnose some cases of anemia or being aware of high probability of its occurrence. A fuzzy-based algorithm consists of if-else fuzzy rule sets to enable experts to interpret. [7] In this paper, a genetic algorithm (GA) has been used for optimizing the parameters of the membership functions (MFs) of the proposed FES for diagnosis of microcytic anemia (IDA and BTT). Fuzzy expert system (FES) is a method employed for diagnosis of different diseases. In paper [8], a system that was designed for determination child anemia is the two-input and single-output neuro-fuzzy network, based on Takagi–Sugeno (TS)-type fuzzy model with triangular membership functions, product inference rule and a weighted average defuzzifier. The paper [9] constructs medical expert systems for the diagnosis of anemia and shows the stages of building a medical knowledge base and examples of writing logical rules. Some of the research papers use machine learning algorithm to diagnose anemia. The research paper [10] analyzes the pallor sites of conjunctiva and tongue for multi-class classification of patients with high pallor. To predict post-transfusion PCV in anemic dogs, different machine learning algorithms such as linear regression, XGBoost and Support

Vector Regression algorithms were used in the research paper [11]. In this paper [12], an image processing procedure had been developed to identify the existence of sickle cells. Machine learning algorithms such as k-nearest neighbor (KNN), support vector machine (SVM) and extreme learning machine (ELM) classifiers are used to test images. Using deep learning methods, an automatic cell recognition system had been developed in [13]. Angiodysplasia Detection and Localization Using Deep Convolutional Neural Networks had been proposed in [14]. The paper [15] shows diagnosis of pediatric patients with β-thalassemia minor, iron deficiency anemia or the co-occurrence of these ailments using multi-label classification technique.

3 Proposed ASDS Model

This section of the paper discusses the architectural model of Anemia Severity Detection System (ASDS). The architectural components of ASDS model are shown in Fig. 1.

The detail about each components of this model is discussed in following:

1. User. Doctors or medical experts are considered as user. For the inquiry about severity of anemia, user can post data of patient's detail to the system's user interface or data entry module.

2. User Interface or Data Entry Module. This module is responsible for interconnecting user or administrator with the system. Through this module, user provides query to the system and system feedbacks result to the user. The administrator can access the casebase also through this module.

3. Data Processing Module. This module is responsible to take the user posted data and then convert their data types into computable data types. Next, it provides that data to the Computational Module. Data Processing Module is also responsible for taking the diagnosis result from the CBR Module. Then, it checks the diagnosis result. If it finds '1', then it shows a message to the user that "Patient is suffering

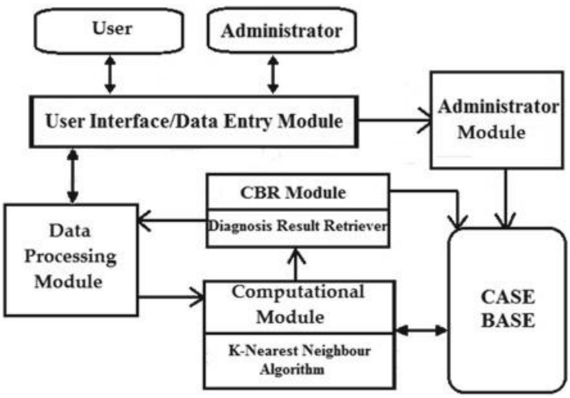


Fig. 1 Architectural components of ASDS model

from Severe Anemia", and if it finds '0', then it shows "Patient is suffering from Mild Anemia".

4. Computational Module. After getting the data from the Data Processing Module, Computational Module considers the set of data as a new case. Now, it is the responsibility of Computational Module to retrieve the most similar cases in the casebase and classify the new case belongs to "mild anemia class" or "severe anemia class". It uses K-Nearest Neighbor algorithm to calculate the similarity and classification calculation.

5. CBR Module. This module retrieves the diagnosis result associated with the new case and provides the diagnosis result to the Data Processing Module. Then, it stores the new case along with that diagnosis result to the casebase.

6. Administrator. Administrator should be doctors or medical experts, who have depth in knowledge of anemia. Only the administrator is responsible to access the casebase. He may modify the casebase by adding, deleting and updating cases.

7. Administrator Module. Administrator Module assists the administrators for administering ASDS.

8. CaseBase. CaseBase contains a total of 200 cases. Each case contains 12 attributes with their associated values and 1 target attribute which contain the diagnosis result. The detail about CB is shown in Sect. 4.

4 CaseBase (CB) in ASDS

The CaseBase (CB) of ASDS is one type of database that stores anemia cases. The case attributes are patient age, symptoms, hemoglobin level, etc. And solution is the target attribute that holds two values: mild anemia and severe anemia. Table 1 shows an example of cases in the CB of ASDS. The values of attributes are already converted from nominal to numerical, e.g., Rural = 0, Town = 1; For anemia type: Severe = 1, Mild = 0, etc.

The original dataset contains anemia survey of 200 patients with 25 attributes (e.g., rural/town, hemoglobin, age). To improve accuracy and reduce training time, data preprocessing technique and feature selection technique have been applied to the dataset. After using both techniques, cases are made and are stored into CB. In the ASDS, CB contains a total of 200 anemia cases with 12 case attributes and one target attribute or solution. The number of cases may increase after the utilization of four activities of CBR cycle.

Table 1 An example of four cases in the CB of ASDS

CASE	Rural or town	Hb (1st obs)	Iron supplement	Hb (1st reading)	Pill taken time	Care quality	Symptom1	Symptom2	Symptom3	Who is the patient	Age	Reproductive status	Anemia type
1	1	16.1	0	14.5	2	2	1	3	9	2	28	2	1
2	0	16.4	0	14.5	2	2	1	3	9	2	30	2	0
3	0	16.4	1	14.9	0.3	2	1	3	9	2	31	3	1
4	0	16.1	0	14.5	2	2	4	3	9	2	24	2	0

5 Methodology and System Implementation

This paper presents three consecutive steps to implement ASDS and diagnose the severity of anemia in a patient through the system. Steps are shown in Fig. 2.

Detail Processing of Step 1:

(a) Collect Dataset. The dataset that helps to develop ASDS is collected from the Website https://scholarworks.iu.edu/dspace/handle/2022/21181. It contains anemia survey of 200 patients with 25 attributes (e.g., rural/town, hemoglobin, age).

(b) Data Preprocessing. Data preprocessing technique that is applied to the dataset before it feeds to the algorithm. To achieve better results from the applied model, it converts incomplete raw data in a useful and efficient format. In this anemia dataset, some attributes contain null values. So, data preprocessing technique replaces null values of an attribute with the calculated mean value of the attribute. This strategy is applied on each attribute that contains null value. After that, all the attribute names are renamed in short.

(c) Set The Target. To classify the anemia is severe or mild, the attribute "How serious is anemia?" is set as target attribute which is renamed as "Anemia_Type" and the values of the target attribute are stored as '1' or '0'. '1' means severe, and '0' means mild.

(d) Use Feature Selection Technique. In machine learning, feature selection is the technique of selecting subset of relevant features or attributes to build the machine learning model. The advantages of feature selection are that it reduces redundant data on dataset and reduces training time and improves accuracy. In this paper, to develop ASDS, only those features are selected from whole dataset that are really important to predict the severity of anemia. Table 2 shows the detail about dataset attributes with their values after using feature selection technique.

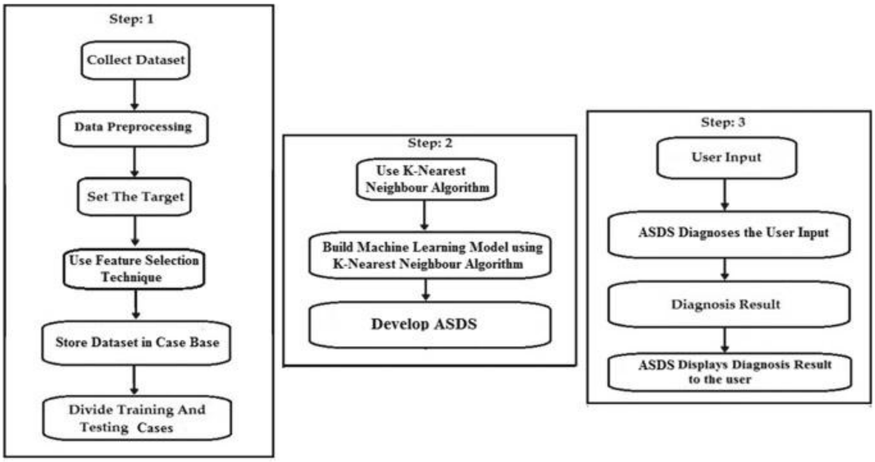


Fig. 2 Steps to implement ASDS and diagnose severity of anemia in a patient through the system

Table 2 Dataset attributes with their values after using feature selection technique

Input attributes	
Attribute	Values
Rural/Town	0–1
Hb (1st obs)	0–18.7
iron_supplement_recommended?	0–1
Hb (1st reading)	10.4–15.7
Pill taken time	0.1–3
Care quality	0–3
Symptom1	0–9
Symptom2	0–9
Symptom3	0–9
Who is the patient	1–3
Age	0–42
Reproductive status	0–4
Output attribute	
Attribute	Values
Anemia_type	0–1

(e) Store Dataset in CaseBase. CaseBase is one type of database. After applying previous steps on the original dataset, the obtained dataset is stored in the casebase. Here each of the survey is considered as each case in CB. The input attributes are considered as problem description of case, and the output or target attribute is considered as the solution of the case.

(f) Divide Training and Testing Cases: Except the target class column, the entire case dataset in CB is divided into two sets in 7:3 ratios, 70% of which is used for training the machine learning algorithm and the 30% is used to test its accuracy, precision and recall result.

Detail Processing of Step 2:

(a) Use K-Nearest Neighbor Algorithm. K-Nearest Neighbor algorithm is a supervised machine learning algorithm. After the training phase, whenever a testing sample is given to K-Nearest Neighbors classifier, then firstly it calculates distance between inputted testing sample and training samples then searches for the K-Nearest Neighbors; next it checks the majority class to which its neighbors belong and lastly assigns that majority output class for the testing sample, referred to in [16]. In machine learning, whenever a machine learning algorithm is trained with training dataset, then a machine learning model is built that can accept inputs, then analyze it statistically and lastly predicts an output more accurately. In this paper, K-Nearest Neighbor algorithm is used to build machine learning model.

(b) Build a Machine Learning Model Using K-Nearest Neighbor Algorithm. After K-Nearest Neighbor algorithm is trained with the training dataset, K-Nearest Neighbor model is built that can accept patient's data, analyze it statistically and lastly predict the severity of anemia more accurately. To check the accuracy, precision and recall results of the model, testing dataset is applied on the model. Performance results of K-Nearest Neighbor model are shown in Sect. 6.

(c) Develop ASDS. ASDS is a diagnosis application that is built using Python. To diagnose more accurately, ASDS uses K-Nearest Neighbor model.

Detail Processing of Step 3:

(a) User Input. Doctors or medical experts submit patient's health detail data to the ASDS using GUI of the system.

(b) ASDS Diagnoses User Input. After ASDS accepts a new input, using K-Nearest Neighbor model, it diagnoses whether patient has severe or mild anemia.

(c) Diagnosis Result: After the diagnosis, ASDS gets a result.

(d) ASDS Displays Diagnosis Result to the User: Finally, user gets feedback result from the system.

6 Accuracy Comparison

System's accuracy depends on the values of dataset, attributes or features of dataset and the applied machine learning model. The more accurate values of dataset with most important features increase accuracy rate of system. Also, most appropriate machine learning language should be used to build the machine learning model. In this paper, for experimentation, the dataset is applied on different machine learning algorithms. Table 3 shows their different accuracy, precision and recall results. From the table, it is clear that K-Nearest Neighbor algorithm works relatively better. Figure 3 shows that with the same dataset K-Nearest Neighbor model provides highest peak of accuracy from Neighbor = 5.

Table 3 Different accuracy, precision and recall results from different machine learning models using same casebase

Machine learning algorithm	Accuracy (%)	Precision (%)	Recall (%)
Logistic regression	90	85	88
Random forest classifier	91.7	90	93
Naïve Bayes	25	85	30
Multi-layer perceptron	90	84	92
Decision tree	78	79	77
K-Nearest Neighbor (Neighbor = 5)	**92**	**84**	**90**

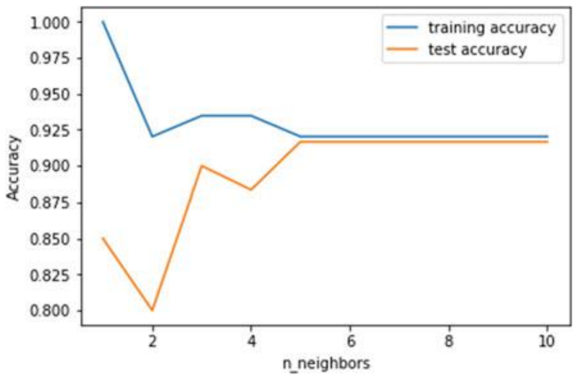


Fig. 3 Different accuracy rate of K-Nearest Neighbor with different number of neighbors

7 Simulation Result

In the Python environment, the simulation tool, Jupyter Notebook, is used to develop the GUI of the ASDS. The GUI of the ASDS is presented in Fig. 4.

8 Conclusion and Future Work

Anemia is a very common disease around the world. Based on the patient's signs, symptoms and hemoglobin amount of blood, the severity of anemia can be detected. Anemia can be mild anemia or not severe anemia. The treatments for mild and severe anemia are different. So, it is necessary to detect the severity of anemia at first to provide correct treatment. In this paper, Anemia Severity Detection System (ASDS) is designed to help the doctors or medical experts to detect the severity of anemia in a patient. ASDS is developed based on Case-Based Reasoning (CBR) methodology and machine learning algorithm. ASDS follows the four steps of CBR cycle as: **1. Retrieve**: Whenever ASDS accepts new inputted values, then it considers it as a new case. Next, ASDS uses machine learning algorithm to calculate similar cases and retrieves the solution or diagnosis result from CaseBase. **2. Reuse**: The new case is solved reusing retrieved solution or diagnosis result. **3. Revise**: Administrator of ASDS takes the responsibility to test applicability in the real world of the new proposed solution with new case. **4. Retain**: The new case is to be stored in the CaseBase with that new solution for future problem solution. As the machine learning algorithm, K-Nearest Neighbor (with Neighbor = 5) has been used. The dataset helps to develop ASDS is taken from https://scholarworks.iu.edu/dspace/handle/2022/21181. There are three steps to develop ASDS, primarily, dataset is collected from the Website, data preprocessing is done, target attribute is set, feature selection process is done, processed dataset is stored to CB, dataset is divided into training and testing dataset. In the next step, using K-Nearest Neighbor algorithm, ASDS is built. In the third step, user submits input data to ASDS and ASDS feedbacks

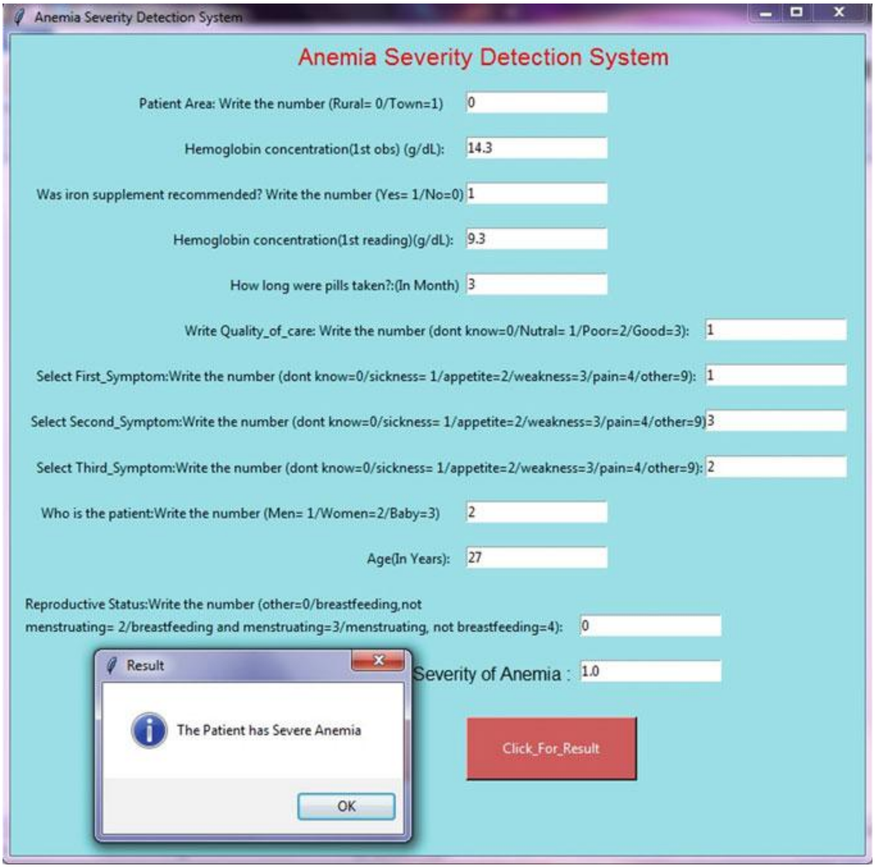


Fig. 4 According to the user inputted data, patient has severe anemia

the result to the user. To compare accuracy results, the processed dataset is applied to five different machine learning algorithms. Comparison result shows that K-Nearest Neighbor algorithm with Neighbor = 5 performs better. K-Nearest Neighbor model gives highest 92% accuracy, 84% precision and 90% recall results. As a screening tool, the graphical user interface (GUI) of ASDS has been developed with the help of Python, so that doctors or medical experts can detect the severity of anemia easily and provide right treatments for right type of anemia.

Future work lies in the enhancement of cases in the CaseBase. The survey about the severity of anemia should be large, so that the dataset can be improved in the future. The future work is also focused on incremental of accuracy, precision and recall results of the system. That means, system will be upgraded and will provide 100% accuracy, precision and recall results.

References

1. WHO: Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Vitamin and Mineral Nutrition Information System. Geneva, World Health Organization. pp. 1–6 (2011)
2. Bergmann, R., Altho, K.-D., Minor, M., Reichle, M., Bach, K.: Case-Base Reasoning-Introduction and Recent Developments. pp. 5–11 (2009)
3. Malu, S., Pal, B.L., Kumar, S.: Detection of anemia using fuzzy logic. Int. J. Comput. Appl. Technol. Res. **4**(10), 762–766 (2015)
4. Imianvan, A.A., Odigie, B.B.: Fuzzy inspired case based reasoning for hematology malignancies classification. Glob. J. Comput. Sci. Technol. (D) Neural & Artif. Intell. **15**(1), 5–14 (2015)
5. Irum, A., Akram, M.U., Ayub, S., Waseem, S., Khan, M.J.: Anemia detection using image processing. In: Proceedings of International Conference on Digital Information Processing, Electronics, and Wireless Communications (DIPEWC2016), pp. 31–36 (2016)
6. Aramideh, J., Jelodar, H.: Application of fuzzy logic for presentation of an expert fuzzy system to diagnose anemia. Indian J. Sci. Technol. **7**, 933–938 (2014)
7. Latifi, F., Zarrabi, H.: A fuzzy-genetic approach for microcytic anemia diagnosis in cyber medical systems. J. Biomed. Eng. Technol. **5**(1), 12–19 (2017)
8. Allahverdi, N.: Design of fuzzy expert systems and its applications in some medical areas. Int. J. Appl. Math., Electron. Comput. **2**, 1–8 (2013)
9. Belginova, S., Uvaliyev, I., Ismukhamedova, A.: Decision support system for diagnosing anemia. In: 2018 4th International Conference on Computer and Technology Applications, pp. 211–215 (2018)
10. Roychowdhury, S., Sun, D., Bihis, M., Ren, J. Hage, P., Rahman, H.H.: Computer aided detection of anemia-like pallor. In: 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pp. 461–464 (2017)
11. Srinilta, C., Sunhem, W., Sangunwong, P., Chanchartree, S.: Machine learning approach in predicting post-transfusion packed cell volume in anemic dogs. In: 2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST), pp. 1–4 (2018)
12. Chy, T.S., Anisur Rahaman, M.: A Comparative analysis by KNN, SVM & ELM classification to detect sickle cell anemia. In: 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), pp. 455–459 (2019)
13. Yu, W., et al.: Automatic classification of leukocytes using deep neural network. In: 2017 IEEE 12th International Conference on ASIC (ASICON), pp. 1041–1044 (2017)
14. Shvets, A.A., Iglovikov, V.I., Rakhlin, A., Kalinin, A.A.: Angiodysplasia detection and localization using deep convolutional neural networks. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 612–617 (2018)
15. Bellinger, C., Amid, A., Japkowicz, N., Victor, H.: Multi-label classification of anemia patients. In: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), pp. 825–830 (2015)
16. Jadhav, S.D., Channe, H.P.: Comparative study of K-NN, naive bayes and decision tree classification techniques. Int. J. Sci. Res. (IJSR) **5**, 1842–1845 (2016)

ECG Signal Analysis, Diagnosis and Transmission

U. B. Mahadevaswamy, Manoj R. Poojari, Murugendra Bandrad, Pavan Kalyan Kallur, and J. Jithesh

Abstract The electrocardiogram (ECG or EKG) signal plays an essential role in the field of medical science and is used in the diagnosis of various cardiovascular diseases (CVD). If the recorded ECG signal has an irregularity in the heartbeat rhythm, it is known as arrhythmia. The diagnosis is done based on ECG morphology. Heartbeat can be subdivided into five categories such as normal, supraventricular ectopic, ventricular ectopic, fusion and unknown beats. In this paper, the authors present an easy and effective way for analysing and diagnosing the nature of the arrhythmia using 1D convolutional neural network (CNN). The ECG data set was obtained from PhysioNet's MIT-BIH database. The PyTorch library was used in python in designing the CNN model which classifies the ECG test sample signal to a category. The diagnosed signal along with the patient details is sent to a cardio specialist for validation via mail and to the CVD forum android application where the case history can be maintained in the database and can be reviewed by any cardio specialist. The CNN model was trained using the data sets and achieved an average accuracy of 97.72%, and the classification achieved is presented in a confusion matrix.

U. B. Mahadevaswamy
Department of Electronics and Communication Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, India
e-mail: mahadevaswamy@sjce.ac.in

M. R. Poojari (✉) · M. Bandrad · P. K. Kallur · J. Jithesh
Department of Electronics and Communication Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, India
e-mail: manojrpoojary@gmail.com

M. Bandrad
e-mail: murugendra1997@gmail.com

P. K. Kallur
e-mail: pavankalyan.kallur@gmail.com

J. Jithesh
e-mail: jeethrao19@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_57

Keywords Arrhythmia · Convolutional neural network · CVD forum · Firebase · PyTorch

1 Introduction

According to [1], 17.9 million people die each year from cardiovascular diseases (CVDs). CVDs are the disorder of the heart and blood vessels which include arrhythmia, coronary heart disease, cerebrovascular disease, rheumatic heart disease and many other conditions. The National Heart, Lung, and Blood Institute [2] stated that arrhythmia is a type of CVD which occurs due to change in rate or irregularity of the frequency of heartbeat, heart beats too quickly, slowly or beats with an irregular pattern. There is a further classification in nature of the arrhythmia, which includes normal supraventricular ectopic, ventricular ectopic, fusion and many other types of disorder.

Electrocardiography (ECG or EKG) is a method used for analysing and diagnosis the type of CVD for treating heart disease. It was also noted in [3] that the ECG is capable of measuring tiny electrical activity of the heart over a period of interval using electrode placed at different points on the skin. This tiny electric signal contains information about the heartbeat rate and rhythm of that person. Many techniques have been designed and implemented, which are briefed out here in the literature survey of ECG arrhythmia classification. Linh et al. [4] presented a fuzzy neural network model which applied Hermite function for the feature extraction. Güler and Übeyli [5] proposed feed-forward neural networks (FFNN) as the classification method, also implemented wavelet transform for the feature extraction and used Levenberg–Marquardt algorithm for the training. Ceylan and Özbay [6] proposed principal component analysis (PCA) and wavelet transform (WT) for the feature extraction. Linh et al. [7] presented support vector machine as the classification technique which utilized Hermite function and higher-order statistics (HOS) and as the feature extraction. Song et al. [8] proposed support vector machine along with linear discriminant analysis (LDA). Polat and Gunes [9] introduced the least squares support vector machine (LS-SVM) along with principal component analysis. Melgani and Bazi [10] proposed support vector machine along with particle swarm optimization (PSO) for the classification.

Beyli [11] presented machine learning technique recurrent neural networks (RNN) classification along with eigenvector feature extraction method. Kumar and Kumaraswamy [12] presented technique of random forest tree (RFT) as the classification method. Jun et al. [13] proposed a K-NN classification method for detecting different types of ECG beats. Jun et al. [14] presented a technique called parallel K-NN classification method for faster arrhythmia classification methods. Rajpurkar et al. [15] presented 1D CNN classification method that uses more data than CNN model.

Even though many different methods are proposed for the ECG arrhythmia classification, these have one or more limitations which are good performance on only carefully selected ECG recordings without cross-validation. The quality of ECG heartbeat is reduced during the feature extraction schemes, relatively low classification performance to adopt in practice.

2 Methodology

1. ECG data set: the ECG heartbeat signals are sourced from PhysioNet's MIT-BIH arrhythmia data set, for normal case and five different arrhythmias cases and were pre-processed and segmented into many signals of a single heartbeat size. Hence, the number of samples is 109,446 with sampling frequency of 125 Hz. The data set contains a matrix in CSV file with each row as a sample in the portion of the data set and the last element of that row denoting to which class does the sample belongs to. The ECG signals are passed as the input to the CNN model.
2. The data set is trained, and the best efficient CNN model obtained is saved.
3. The obtained result is tabulated in confusion matrix.
4. A test ECG signal is diagnosed using the trained CNN model.
5. The obtained result is transmitted to a specialist and to the CVD forum for review and validation of the tested result.

3 Convolutional Neural Network

Convolutional neural network (CNN) is one of the variants of neural network which is mainly used in various classification problems. It derives its name from the type of hidden layers it consists of. The typically hidden layers in CNN are convolutional layers, pooling layers, fully connected layers and normalization layers. As noted in [16], CNN is used because it provides many advantages for the classification. It is very effective when the expectation is to derive interesting features from shorter (fixed length) segments of the overall data set. This also applies well for the analysis of data from the sensor data. Figure 1 illustrates the proposed algorithm for the CNN model.

In algorithm, there will be three convolutional layers with kernel size of three along with ReLU activation function, a hidden layer with two fully connected layers, and in addition, there is one output layer with a softmax function to hold the output of the neural network (NN). The number of layers is chosen such that high accuracy is obtained for the given data set and not overfit the model.

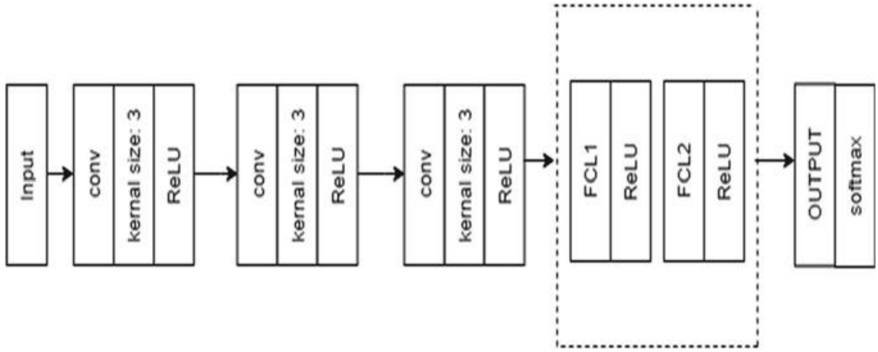


Fig. 1 Proposed algorithm for the CNN model

3.1 Convolutional Layer

A 1D convolutional layer applies convolution on the input signal as shown in Eq. (1). For the input of size (N, Cin, L), it generates the output of size (N, Cout, L_{out}). Here three convolutional layers are used, each with a kernel size of 3. The input to the first layer is (4, 1, 187) and the output size is (4, 1, 185); similarly for the next two layers, the batch size and the number of channels remain the same. The L_{out} for the second layer is 183 and for the third layer is 181.

$$\mathrm{Out}(Ni, \mathrm{Cout}j) = \mathrm{bias}(\mathrm{Cout}j) + \sum_{k=0}^{\mathrm{Cin}\text{ - }1} \mathrm{weight}(\mathrm{Cout}j, k) \times \mathrm{input}(Ni, K) \quad (1)$$

where N is batch size, C denotes the number of channels, L is the length of the sequence. Based on different parameters like padding, dilation, kernel size and stride, the length of the output varies as given by Eq. (2).

$$L_{\mathrm{out}} = \mathrm{floor}((L_{\mathrm{in}} + 2 \times \mathrm{pd} \times (k_s - 1) - 1)/\mathrm{stride} + 1) \quad (2)$$

where L_{out}, pd, and k_s represent the length of the output, padding dilation, and kernel size, respectively.

3.2 Architecture

The proposed architecture for ECG signal classification is summarized in Table 1. The architecture consists of three convolutional layers, and two fully connected layers as shown in Fig. 2. The kernel size is set to 3, and stride of size 1 is used. Rectifier

Table 1 Summary table of the proposed CNN architecture

Layers	Type	No. of Neurons	Kernel size	Stride
0–1	Convolution	185	3	1
1–2	Convolution	183	3	1
2–3	Convolution	181	3	1
3–4	FC	256	–	–
4–5	FC	256	–	–

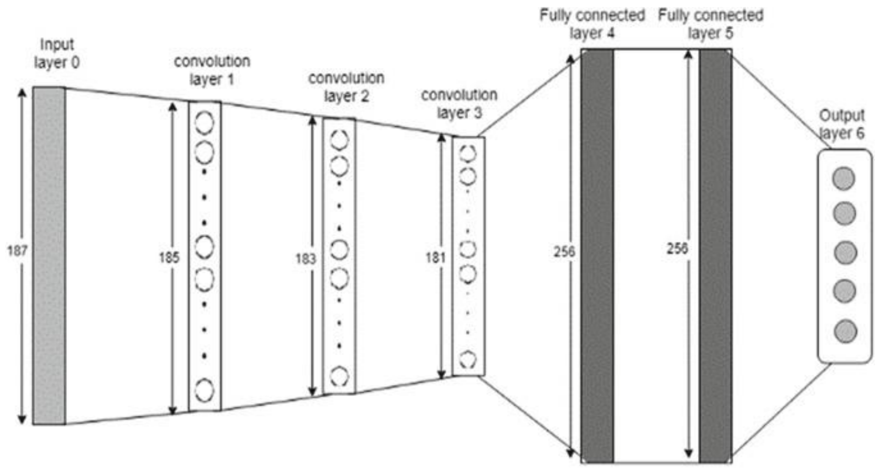


Fig. 2 Architecture of the proposed CNN model

linear unit (ReLU) is used as activation function. The fully connected layer of 256 neurons each and 5 output neurons at the output with softmax function.

3.3 Forward Propagation

Forward propagation is the propagation of the input from input layer to output layer, passing different layers in between. After the forward propagation, error is calculated at the output. The equations for the output of the forward propagation are shown in Eq. (3). Here Z represents the output from the linear layer before applying the activation function.

$$Z[i] = W[i] \times A[i-1] + b[i] \tag{3}$$

where W denotes the weights and b is the bias.

The output of the ith layer after applying the activation function is shown in Eq. (4) where $i > 1$, $i = 0$ gives the input. 'A' represents the output after applying the activation function.

$$A^{[i]} = g^{[i]}\left(Z^{[i]}\right) \tag{4}$$

3.4 Backpropagation

Backpropagation works by propagating from output to input by updating the weights of the layers to reduce the error calculated after the forward propagation (3.3). The equations representing the backpropagation through the layers are shown from Eqs. (5) to (12). dX denotes the gradient value of X, l denotes the final layer and A[0] represents the input signal. Equation (5) calculates the gradient of the output, and Eqs. (6) and (7) show how gradient with respect to weights and bias is calculated at the final output layer. Similarly, for the other layers formulas are mentioned in Eqs. (8), (9) and (10) for gradient of outputs, weights and bias, respectively. The weight and bias updating are done by subtracting the previous value by their respective gradient multiplied by learning rate. Their calculation is shown in Eqs. (11) and (12). In Eq. (5) $dZ^{[l]}$, $A^{[l]}$ and Y represent gradient of output and expected output, respectively. $dZ^{[i]}$, $W^{[i+1]T}$, and $g^{[i]}$ denote gradient of output, weights at layer $i + 1$, and output before applying activation, respectively, in Eq. (8). In (9), $db^{[i]}$ and $dZ^{[i]}$ denote gradient of bias and gradient of output, respectively. In Eq. (11), $dW^{[i]}$ and lr represent gradient of weight and learning rate. In the following equations, 'm' denotes number of output neurons.

$$dZ^{[l]} = A^{[l]} - Y \tag{5}$$

$$dW^{[l]} = 1/m\left(dZ^{[l]} \times A^{[l]T}\right) \tag{6}$$

$$db^{[l]} = 1/m\left(\sum dZ^{[l]}\right) \tag{7}$$

$$dZ^{[i]} = \left(W^{[i+1]T} \times dZ^{[i+1]}\right) \times \left(g^{[i]} \times \left(Z^{[i]}\right)\right) \tag{8}$$

$$dW^{[i]} = 1/m\left(dZ^{[l]} \times A^{[i-1]T}\right) \tag{9}$$

$$db^{[i]} = 1/m\left(\sum dZ^{[i]}\right) \tag{10}$$

$$W^{[i]}(t+1) = W^{[i]}(t) - lr \times dW^{[i]} \tag{11}$$

$$b^{[i]}(t+1) = b^{[i]}(t) - lr \times db^{[i]} \tag{12}$$

4 Transmission

Transmission of ECG data is done in two means; the test report generated by the CNN model is sent via a system generated email and also sent to the CVD forum.

4.1 System Generated Email

Simple Mail Transfer Protocol (SMTP) is a communication protocol for electronic mail transmission which is used for transmitting the system generated email to the cardiac specialist for the review. Figure 3 illustrates the flow diagram of SMTP communication protocol where the system generated mail is sent to the receiver over the Internet via mail server. Post Office Protocol 3 (POP3) is the standard protocol for receiving email.

5 CVD Forum Android Application

An android application for hosting a forum, where the ECG report of a cardiovascular disease patient can be posted, and the verified members of cardio specialists can give their review on it which will be valuable in generating a real-time opinion and validation for the system generated report; also, the CVD case history can be maintained for future reference. CVD forum android application is developed in android studio which supports all the android devices with android version 4.4 (KitKat) and above. Google Firebase is used as a database for uploading the ECG report to the

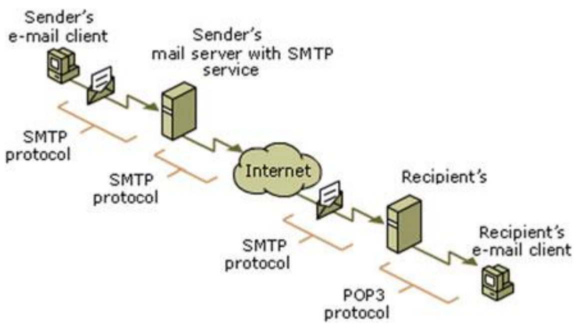


Fig. 3 System generated email using SMTP communication protocol

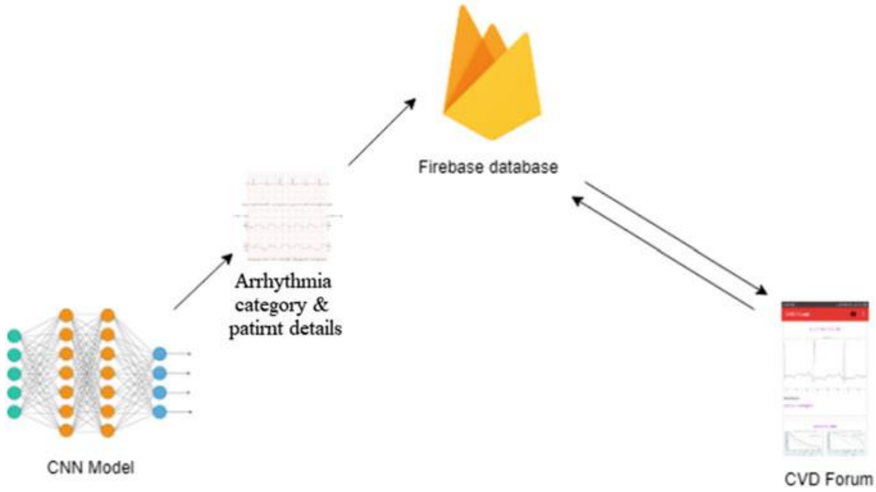


Fig. 4 Flow diagram of ECG data transmission

cloud, and the data is retrieved back to the android application CVD forum. Flow diagram of the ECG data transmission is illustrated in Fig. 4. The diagnosed ECG test sample is directly uploaded to the firebase database after the test run using firebase and pyrebase library. The application can have unlimited registered user accounts with email address verification of 1000 per day and password reset option of 150 per day.

The screenshots for the CVD forum are as shown in Figs. 5 and 6. Figure 5 is the main page of the application which is designed using xml language where all the post related to the CVD case are present in recycle review format and Fig. 6 is the upload page where a verified cardiac specialist himself can upload a CVD case to the forum from the gallery.

A cardiac specialist can give review on a CVD case by long pressing on that particular post and selecting the option review as shown in Fig. 7. After exercising the write review option, the user is diverted to a review page and is asked to enter authenticated user name, specialty and email address as shown in Fig. 8.

6 Results

The proposed CNN algorithm was trained on Google Colab. A total of 109,446 ECG samples were divided into two sets, i.e. training set and testing set, with the ratio of 4:1; approximately, 87,553 ECG samples were used for training and 21,891 ECG samples were used for testing. It took approximately 45 min to complete 30 training epochs. The model was fine-tuned (layers, batch size, epochs, learning rate, etc.) if

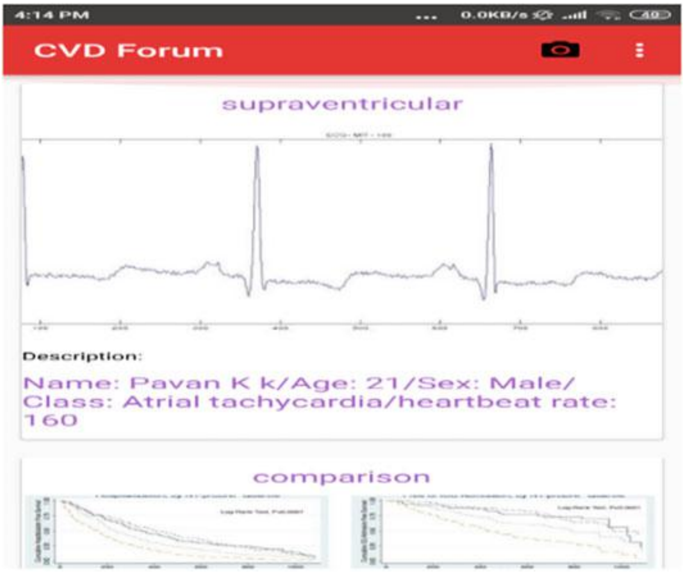


Fig. 5 CVD forum main page

Fig. 6 CVD forum upload page

expected accuracy is not reached. The model with the best accuracy was saved for the real-time prediction of the patient's ECG test sample input.

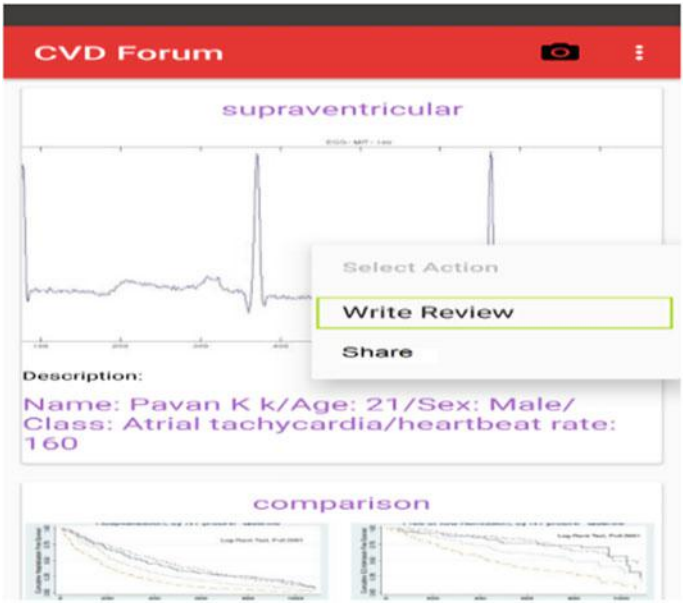


Fig. 7 Screenshot of selecting a review process

Fig. 8 CVD forum review page

6.1 *Confusion Matrix*

Table 2 shows the confusion matrix of ECG heartbeats where *x*-axis represents the predicted heartbeats and *y*-axis represents actual heartbeats. The highlighted diagonal elements represent the correctly classified ECG heartbeats.

Table 2 Confusion matrix of the ECG samples classified to its respective category

		Predicted				
		Normal	Supraventricular ectopic	Ventricular ectopic	Fusion	Unknown
Actual	Normal	18,013	31	54	9	10
	Supraventricular ectopic	171	379	5	0	1
	Ventricular ectopic	83	6	1342	15	2
	Fusion	38	1	15	108	0
	Unknown	34	1	6	0	1567

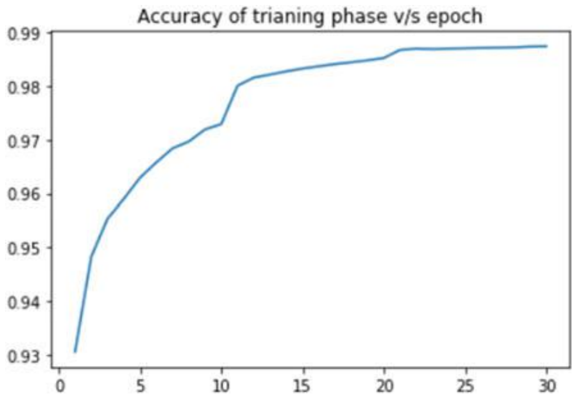


Fig. 9 Training accuracy versus epoch

6.2 Training Accuracy and Validation Accuracy

The CNN model was trained with 30 epochs, each iteration of epoch yielding its efficiency closer to the max efficiency of the model by adjusting the weights and reducing the losses.

The training and validation accuracy after each epoch in percentage is traced in Figs. 9 and 10, where the x-axis represents the epoch interval, y-axis represents the training/validation accuracy.

6.3 Accuracy, Sensitivity and Specificity of Each Category

See Table 3.

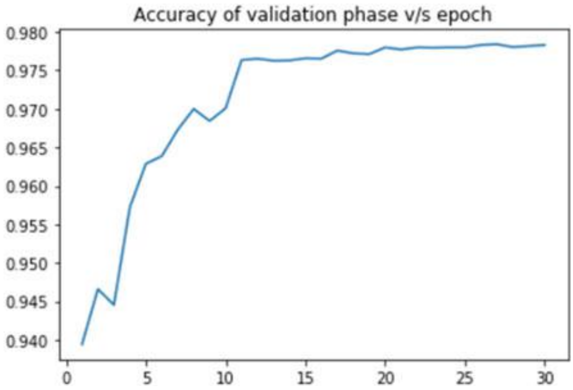


Fig. 10 Validation accuracy versus epoch

Table 3 Sensitivity and specificity for different classes

Category	Accuracy (%)	Sensitivity (%)	Specificity (%)
Normal	98.03	91.20	99.41
supraventricular ectopic	99.01	99.82	67.80
Ventricular ectopic	99.15	99.58	92.40
Fusion	99.64	99.88	66.66
Unknown	99.75	99.92	97.32

6.4 CVD Forum Review Reports from the Specialist

Figures 11 and 12 illustrate the review feature in the CVD forum android application on the supraventricular ectopic post in the news feed and the specialist writing his review on the same post in the review page, respectively.

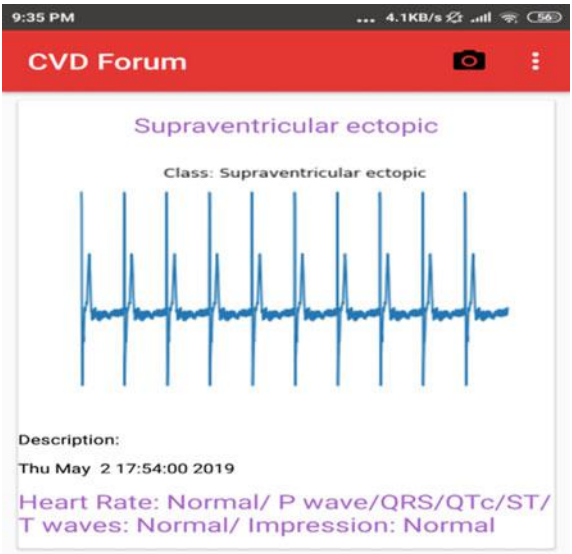
The review from the cardiac specialist is retrieved back from the firebase database in the form of '.txt' file with file name same as the post id as shown in Fig. 13 for the supraventricular ectopic post in the CVD forum.

7 Comparison

Recent works in the field of arrhythmia diagnosis are compared in Table 4 along with our proposed work. We have compared the methodologies used for the classification and the accuracy obtained for the respective methods.

RR: RR intervals, WPD: Wavelet packet decomposition, WPE: Wavelet packet entropy, PCA: Principal component analysis, CNN: Convolutional neural network, sen: Sensitivity, spc: Specificity, acc: Accuracy.

Fig. 11 CVD forum supraventricular ectopic



8 Novelty

- A CNN architecture with least number of convolutional layer (i.e. three layers) and with least epoch obtaining a significant efficiency of 97.772%.
- Social media android application specifically for cardiovascular disease with feature like authenticated users, news feed, upload from gallery, review a post.
- Database for maintaining the CVD case history for future reference and case study.

9 Conclusion

In this research, the authors developed an analysis, diagnosis and transmission system for ECG signal for identifying five heartbeat categories under arrhythmia using 1D convolutional neural network. The proposed CNN algorithm was trained and validated on ECG signal obtained from MIT-BIH (which constitute real ECG signal collected from patients for medical research). The result obtained in detecting all the categories correctly was significantly high with accuracies of 98.03%, 99.01%, 99.15%, 99.64% and 99.75% for classes normal, supraventricular ectopic, ventricular ectopic, fusion and unknown, respectively. Also, the tested results were transmitted to the CVD forum for reviews from cardio specialist. A confusion matrix was generated for the validation data set. Validation accuracy of 97.78%, average sensitivity of 98.08% and average specificity of 84.72% were achieved. In future, the authors

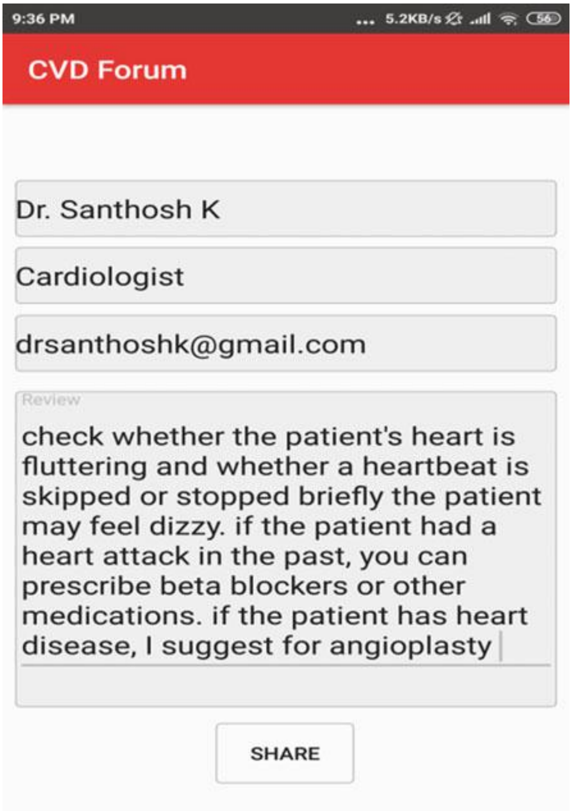


Fig. 12 CVD forum review for supraventricular ectopic post

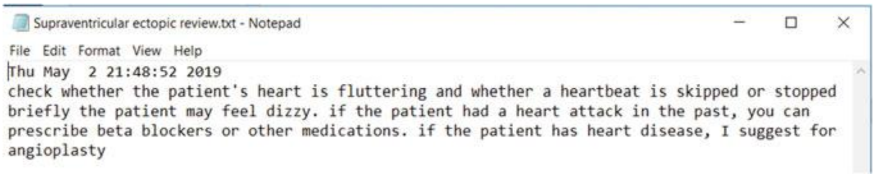


Fig. 13 Review from the cardiac specialist on supraventricular ectopic post

would like to implement a computer-aided design (CAD) ECG system for real-time analysis and diagnosis for 12-lead ECG signal, and further, increase the number of ECG heartbeat categories for more specific nature of arrhythmia and for other CVDs.

From this research, we can learn that a simple deep neural network model is able to classify the arrhythmia beats very accurately. With the addition of more classes of data, more advanced model can be built and can be integrated into the current ECG

Table 4 Comparison of proposed work with previous works

Authors	Year	Method	Result
Li and Min [17]	2016	Conventional machine learning approach: WPD + RR, WPE, WPE + RR, RF classifier	WPE + RR: acc: 94.61%
Dewangan and Shukla [18]	2016	Neural network-based global classifier	acc: 87% sen: 65% spe: 92%
Rama Raju et al. [19]	2017	Wavelet transform classification using neural network classifier with PCA	acc: 95.7 acc: 97.1(PCA)
Acharya et al. [20]	2017	R-peak detection. 11-layer deep neural network. No feature selection or feature reduction.	acc: 93.5 sen: 93.71 spc: 92.81
Acharya et al. [21]	2017	Nine-layer deep convolutional neural network Two sets of experiment	acc: 94.03 sen: 96.71 spc: 91.54
Kachuee et al. [22]	2018	Deep residual CNN	acc: 93.4
Savalia and Emamian [23]	2018	Deep neural network algorithms such as multi-layer perceptron (MLP) and convolutional neural network (CNN)	acc: 88.7(MLP) acc: 83.5(CNN)
Proposed work	2019	Three deep convolutional neural network layers	acc: 97.72 sen: 98.08 spc: 84.71

systems so that instead of getting only the ECG waveforms on a paper, more detailed reports can also be generated.

References

1. Cardiovascular Diseases (CVDs): World Health Organization (2019). https://www.who.int/cardiovascular_diseases/en/
2. National Heart, Lung, and Blood Institute (2011) What Is Arrhythmia? Archived from the original on 2 March 2015. Retrieved 7 March 2015
3. Lilly, L.S., (ed.): Pathophysiology of Heart Disease: A Collaborative Project of Medical Students and Faculty, 6th edn., p. 74. Lippincott Williams & Wilkins, Philadelphia (2016). ISBN 978-1451192759
4. Linh, T.H., Osowski, S., Stodolski, M.: On-line heartbeat recognition using Hermite polynomials and neuro-fuzzy network. IEEE Trans. Instrum. Meas. **52**(4), 1224–1231 (2003). https://doi.org/10.1109/TIM.2003.816841

5. Güler, İ., Übeyli, E.D.: ECG beat classier designed by combined neural network model. Pattern Recognit. **38**(2), 199–208 (2005). https://doi.org/10.1016/j.patcog.2004.06.009
6. Ceylan, R., Özbay, Y.: Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network. Expert Syst. Appl. **33**(2), 286–295 (2007). https://doi.org/10.1016/j.eswa.2006.05.014
7. Linh, T.H., Osowski, S., Stodolski, M.: On-line heartbeat recognition using Hermite polynomials and neuro-fuzzy network. IEEE Trans. Instrum. Meas. **52**(4), 1224–1231 (2003). https://doi.org/10.1109/TIM.2003.816841
8. Song, M.H., Lee, J., Cho, S.P., et al.: Support vector machine based arrhythmia classification using reduced features. Int. J. Control Autom. Syst. **3**(4), 571–579 (2005). https://doi.org/10.1016/j.artmed.2008.04.007
9. Polat, K., Gunes, S.: Detection of ECG Arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine. Appl. Math. Comput. **186**(1), 898–906 (2007). https://doi.org/10.1016/j.amc.2006.08.020
10. Melgani, F., Bazi, Y.: Classification of electrocardiogram signals with support vector machines and particle swarm optimization. IEEE Trans. Inf. Technol. Biomed. **12**(5), 667–677 (2008). https://doi.org/10.1109/TITB.2008.923147
11. Beyli, E.D.: Combining recurrent neural networks with eigenvector methods for classification of ECG beats. Digit. Signal Proc. **19**(2), 320–329 (2009). https://doi.org/10.1016/j.dsp.2008.09.002
12. Kumar, R.G., Kumaraswamy, Y.S.: Investigating cardiac arrhythmia in ECG using random forest classification. Int. J. Comput. Appl. **37**(4), 31–34 (2012). https://doi.org/10.5120/4599-6557
13. Jun, T.J., Park, H.J., Yoo, H., et al.: GPU based cloud system for high-performance arrhythmia detection with parallel k-NN algorithm. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5327–5330 (2016). https://doi.org/10.1109/embc.2016.7591930
14. Jun, T.J., Park, H.J., Minh, N.H., et al.: Premature ventricular contraction beat detection with deep neural networks. In: IEEE International Conference on Machine Learning and Applications, pp. 859–864 (2016). https://doi.org/10.1109/icmla.2016.0154
15. Rajpurkar, P., Hannun, A.Y., Haghpanahi, M., et al.: Cardiologist-level arrhythmia detection with convolutional neural networks (2017). arXiv:1707.01836, https://doi.org/10.1038/s41591-018-0268-3
16. Kiranyaz, S., Ince, T.: Real-Time Patient-Specific ECG Classification by 1-D Convolution Neural Networks, pp. 1558–2531. IEEE, Piscataway (2016). https://doi.org/10.1109/tbme.2015.2468589
17. Li, T.Y., Min, Z.: ECG classification using wavelet packet entropy and random forests. Entropy **18**(8), 1–16 (2016)
18. Dewangan, N.K., Shukla, S.P.: ECG arrhythmia classification using discrete wavelet transform and artificial neural network. In: 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, pp. 1892–1896 (2016)
19. Rama Raju, N.N.S.V., Malleswara Rao, V., Jagadesh, B.N.: Identification and classification of cardiac arrhythmia using neural network. HELIX **7**(5), 2041–2046
20. Acharya, U.R., Fujita, H., Oh, S.L., Hagiwara, Y., Tan, J.H., Adam, M.: Application of deep convolutional neural network for automated detection of myocardial infarction using ecg signals. Inf. Sci. **415**, 190–198 (2017)
21. Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adam, M., Gertych, A., San, T.R.: A deep convolutional neural network model to classify heartbeats. Comput. Biol. Med. **89**, 389–396 (2017). https://doi.org/10.1016/j.compbiomed.2017.08.022
22. Kachuee, M. Fazeli, S., Sarrafzadeh, M.: ECG heartbeat classification: a deep transferable representation (2018). arXiv:1805.00794v2
23. Savalia, S., Emamian, V.: Cardiac arrhythmia classification by multi-layer perceptron and convolution neural networks. Bioengineering **5**, 35 (2018)

The Effect of Real-Time Feedback on Consumer's Behavior in the Energy Management Sector: Empirical Study

Muhammad Nabeel, Basel Ali, and Allam Hamdan

Abstract Despite efforts exerted by the Electricity and Water Authority of the Kingdom of Bahrain (EWA), the Kingdom has one of the highest energy consumptions per capita at a rate of 11,500 kWh as of 2017. The rate is estimated to increase by 6–8% a year. The drop in oil prices has led to an increase in electricity prices in the Kingdom, which directly impacted the consumers. This in turn has opened doors for research in the domain of energy conservation. One way to achieve energy conservation is by avoiding energy-wasting behaviors and habits. In the modern era of information-based decision making, feedback technologies have shown a great potential in reducing energy consumption. Up to 20% reduction has been reported with the latest advancements in the feedback technology. In this study, an experiment was conducted by installing feedback devices in five different houses. In the first 20 days, the participants were only educated on electricity saving practices recommended by EWA and were not given access to feedback displays. This was done to monitor their regular electricity consumption behavior using the same devices to ensure accuracy. In the second phase of the experiment, participants were provided with access to the feedback monitors with displays real-time consumption information. Consumption data of the 5 houses were recorded and analyzed. Additionally, interviews were conducted with the families at the termination of the experiment to understand the unquantifiable aspects of the study. The research study reported an overall reduction of 12% in the electricity consumption with a variation from 3.9 to 25.7%.

Keywords Real-Time feedback · Energy management sector · Consumer's behavior · Bahrain

M. Nabeel · B. Ali
The George Washington University, Washington, USA
e-mail: zahabnabeel@hotmail.com

B. Ali
e-mail: bali@email.gwu.edu

A. Hamdan (✉)
Ahlia University, Manama, Bahrain
e-mail: allamh3@hotmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_58

1 Introduction

Wastage of electricity is a significant issue in the Kingdom of Bahrain and has unfortunately become a norm. Despite efforts by the authorities to create awareness and instill a culture of energy saving, the response from the people has been very casual toward the cause. By the end of 2017, Bahrain continues to be one of the highest consumers of electricity with a per capita rate of 11,500 kWh and a 6–8% growth in annual demand [963, 1]. Additionally, Bahrain also emits CO_2 at a very high rate of 26.8 metric tons per capita, which is more than three times of the emissions observed in Europe [902, 2]. The use of real-time feedback technology as a tool to change consumer behavior and reduce electricity consumption has been investigated in many countries around the globe including the USA, EU countries, UK, Australia, and Canada and has shown positive results [916, 3; 927, 4; 928, 5].

This paper focuses on testing the feasibility of such methodology on the Bahraini population who are still alien to the concept. The topic is more relevant at this time as authorities in Bahrain are currently investing in smart meters. Investigating the potential effect of feedback technology on electricity consumers' behavior in the Kingdom of Bahrain may help justify this investment. The findings of this study are expected to benefit the general public in finding means of reducing their electricity bills by better understanding their usage. It will also benefit the authorities, who are facing a challenge in reducing electricity waste in the Kingdom.

2 Methodology

The experiment included installing feedback devices in five different households. The choice of houses would later be discussed in the sampling section, whereas the number of houses (i.e., 5) was selected due to budget and time limitations. The feedback device chosen for the experiment was Blue Line Innovations (BLI) Power Cost Monitor (PCM), BLI-28000ER which provides real-time feedback on electricity usage [946, 6]. The device consists of two parts: an In-Home Display (IHD) and an optical sensor. The IHD displays the corresponding wirelessly transmitted real-time readings. The sensor mounts on the meter, as shown in Fig. 1, and reads from the LED (in digital meters) or the black dot on the rotating disk of the disk meters.

The BLI-28000ER comes with a couple of helpful features including a fast refresh rates (32 s or less), the capability of the device to show the total and real-time consumption in kWh as well as dollar amount. The device also makes use of a predictive analysis and computes the estimated usage at the end of 30 days again in both kWh and dollar units. This provides the consumer a good insight into how his/her electricity bill would look like with the current usage behavior and provides an opportunity to make changes. The device is also programmable for different tier as

Fig. 1 Sensor installation on a participant's meter

well as peak and off-peak rates. It is helpful in case electricity is charged at different rates at different times of the day.

2.1 *Experiment Phases and Timeframes*

Houses chosen for this experiment were of different sizes, levels of income, and backgrounds. Table 1 summarizes relevant information about each house. The experiment lasted for 40 days divided in two equal phases of 20 days each. In the first phase, the

Table 1 Research sample

Sample	Family members	Age ranking	Gender	Household type	Socioeconomic status	Education level
House 1	7	00–10: 1 20–30: 3 30–40: 1 50–60: 2	3 m/4 f	3 Bedrooms flat	Middle class	Medium
House 2	3	20–30: 1 50–60: 2	1 m/2 f	4 Bedrooms small villa (old)	Middle class	High
House 3	3	20–30: 1 50–60: 2	2 m/1 f	4 Bedrooms spacious flat	Higher middle class	High
House 4	9	00–10: 2 20–30: 5 50–60: 2	5 m/4 f	4 Bedrooms flat	Lower middle class	Low
House 5	8	10–20: 1 20–30: 5 50–60: 2	5 m/3 f	5 Bedrooms small villa	Lower middle class	Low

feedback devices were installed in the houses to collect data, but the family members had no access to the feedback. In the second phase, all family members were aware of the feedback devices installed at home and had access to the displays. The IHD was placed in central areas where it was visible at most times to the family members. A list of recommendations to save energy was also placed near the IHD to give family members ideas on how to avoid wasting energy. The daily measurements of electricity consumption at each household were recorded along with any comments or issues which might affect normal electricity usage (the family had people over for a gathering, for example). Even though the experiment was performed in the summer season when the weather remains consistent daily, the outside temperature, humidity, and heat index were also recorded as they may have an effect on our results.

2.2 Data Collection

Data collected over the 40 days were analyzed using SPSS. The data for each phase of each house were plotted using boxplots to examine both central tendency and variability and display any outliers or extreme scores that might be included in the distribution.

Bar charts were also used to display the daily consumption and the mean consumption for each phase as well as values of temperature, humidity, and heat index. The goal was to find out if there is any evidence of correlation between power consumption and changes in metrological variables. To examine if the decrease in consumption after the implementation of feedback technology was statistically significant, test of normality was first performed on the data using SPSS by inspecting the graphical outputs of histograms and Normal Q-Q plots as well as the numerical outputs of skewness, kurtosis, and Shapiro–Wilk test. Table 2 showcases the summarized results for each house. For normally distributed data, independent samples t-test was performed to determine if the difference in electricity consumption was statistically significant. On the other hand, in the case of non-normal data, Mann–Whitney test was used.

3 Data Analysis

3.1 House 1 Analysis

The boxplot in Fig. 2 shows the median in the after-feedback phase to be lower by a very small margin, whereas the bottom whisker is lower quite considerably. Both phases have two outliers in the lower range. The lower outliers in the first phase were explained by the household to be associated with an “all-day” visit to a relative’s

Table 2 Summarized normality test results

Sample		Visual outputs	Skewness	Kurtosis	Shapiro–Wilk *p*-value	Deduction
House 1	Before	Not normal	−0.635 ± 0.512	3.421 ± 0 .992	0.042	Not normal
	After	Not normal	−1.405 ± 0.512	2.065 ± 0.992	0.010	Not normal
House 2	Before	Not normal	0.828 ± 0.512	1.957 ± 0.992	0.025	Not normal
	After	Not normal	1.646 ± 0.512	2.775 ± 0.992	0.002	Not normal
House 3	Before	Approx. normal	0.809 ± 0.512	1.642 ± 0.992	0.245	Approx. normal
	After	Approx. normal	−0.187 ± 0.512	−0.775 ± 0.992	0.719	Approx. normal
House 4	Before	Approx. normal	−0.381 ± 0.512	−0.240 ± 0.992	0.881	Approx. normal
	After	Approx. normal	−0.570 ± 0.512	−0.754 ± 0.992	0.137	Approx. normal
House 5	Before	Approx. normal	−0.222 ± 0.512	0.371 ± 0.992	0.880	Approx. normal
	After	Not normal	−0.585 ± 0.512	−1.596 ± 0.992	0.001	Not normal

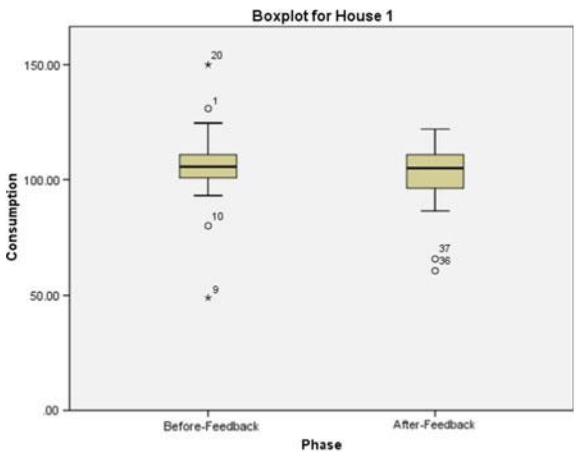


Fig. 2 House 1 electricity consumption before- and after-feedback

house, whereas the lower outliers in the second phase had no specific reason and were regarded to the reduction efforts exerted by the family members in response to the feedback provided. The before-feedback phase also had two questionable upper range values. The higher outlier on day 20 was explained by higher humidity level on the day along with a visit from guests, whereas the outlier observed on day 1 was associated with hosting guests at home.

Figure 3 summarizes the daily electricity consumption in house 1 before and after the implementation of the feedback technology. The figure also shows the daily values of temperature, humidity, and heat index.

Figure 3 shows no clear correlation between the metrological factors and electricity consumption. While a small increase and decrease in temperature have been observed, both the maximum and minimum values do not correspond to the maxima and minima of the temperature, respectively. The maximum electricity consumption value happens on the same day with the highest value of humidity. This can be explained by noting that the efficiency of air conditioners drops in high humidity levels causing an increase in electricity consumption. Higher values of humidity and heat index corresponded to the maximums in both phases, whereas the minimum values showed no pattern with these variables. They were merely associated with the absence of family members in the first phase and efforts to reduce consumption in the second. An important point to note here is that despite the higher values of all three variables in the feedback phase, the average consumption showed a decent decrease.

An overall decrease of 4.1% is observed in average daily consumption of electricity after the introduction of feedback technologies. Based on the normality test results, the data for house 1 were shown to have non-normal distribution. Therefore, Mann–Whitney test was used to inspect if the consumption after-feedback decreased significantly from the consumption before-feedback. Using a 1-tailed test with alpha = 0.05, the calculated p-value 0.313 concludes that the decrease in consumption in the second phase is not statistically significant. However, this decrease has been observed despite the increase in temperature, humidity, and heat index in the second

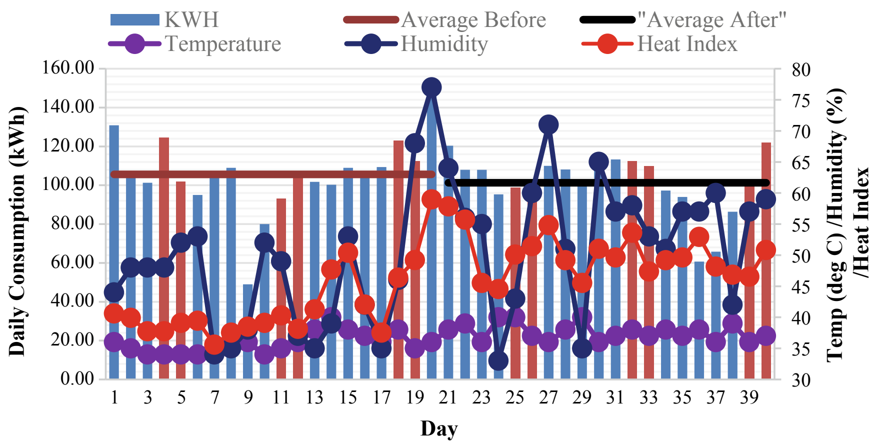
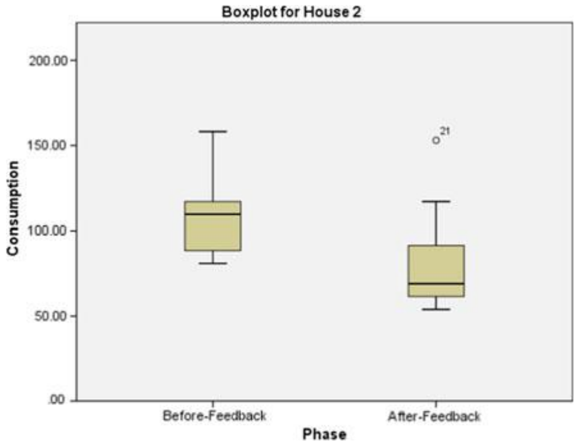


Fig. 3 Energy consumption for house 1 along with metrological factors

Fig. 4 House 2 electricity consumption before- and after-feedback



phase by 5%, 16.2%, and 19.5%, respectively. These considerable increases in temperature, humidity, and heat index call for an increase in usage as well, especially in case of the air conditioners.

3.2 House 2 Analysis

The boxplot in Fig. 4 shows a very clear difference between the medians of the two phases. The after-feedback phase has a clear reduction shown in terms of whiskers and percentiles. Only one outlier exists in the second phase at the upper level for this household. This was due to some guests visiting (Fig. 5).

Using the Mann–Whitney test, the calculated p-value was 0.000 which indicated that the 25.7% decrease in the daily electricity consumption was statistically significant. This high value of reduction was seen despite the increase temperature, humidity, and heat index by 3.9%, 12.2%, and 13.7%, respectively, during the feedback phase of the experiment.

3.3 House 3 Analysis

A clear reduction in the percentiles is observed in Fig. 6 in the after-feedback phase of the experiment. Only one outlier is observed in phase 1, which can be explained by a dinner party hosting 20 guests that night.

Again, Fig. 7 shows no set pattern is observed with any of the three metrological factors. None of the factors seemed to be dominant in this case; high values corresponded to temperature sometimes and humidity the other times. The maximum (outlier in this case) and minimum do not fall at the maxima or minima of any of the

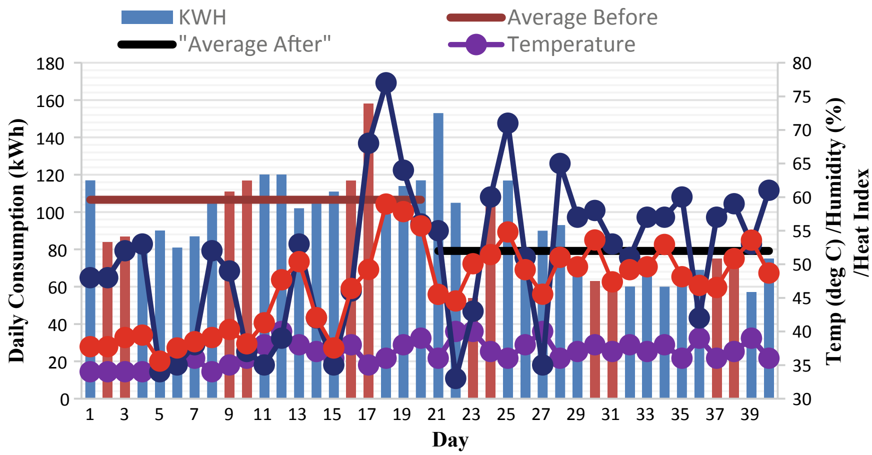
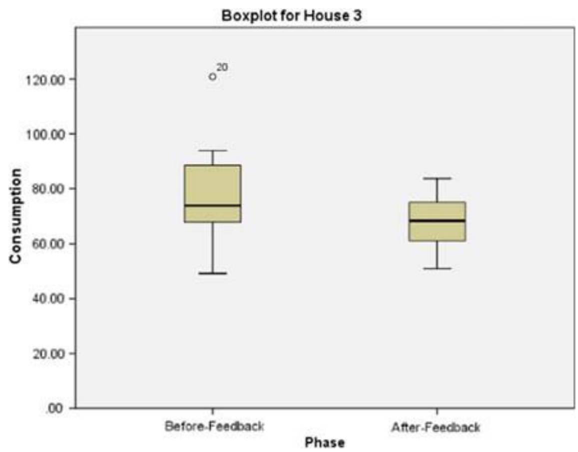


Fig. 5 Energy consumption for house 2 along with metrological factors

Fig. 6 House 3 electricity consumption before- and after-feedback



variables. They corresponded to guest parties or family outings instead. The average temperature conditions remained constant in both phases but the humidity increased which caused the heat index to increase as well. Despite this increase, the house showed a decrease in electricity consumption.

Data collected for house 3 were approximately normal; therefore, independent samples t-test with alpha = 0.05 was conducted in this case. The p-value was found to be 0.018; thus, the 13.1% reduction in the consumption was statistically significant. This reduction was observed despite the increase in factors like humidity and heat index by 4.2% and 2.7%, respectively.

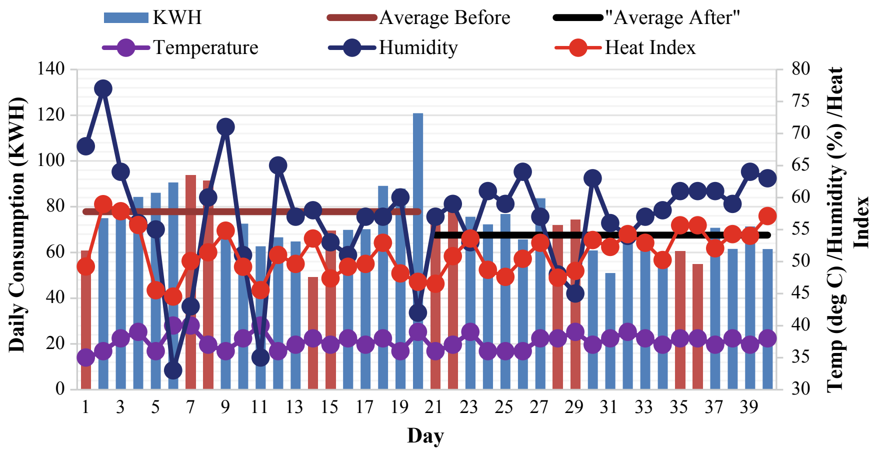


Fig. 7 Energy consumption for house 1 along with metrological factors

3.4 House 4 Analysis

A clear reduction in the median can be seen from the boxplot in Fig. 8. The upper whisker has come down but the lower one is still more than the first phase of the experiment. No outliers are observed in the boxplot (Fig. 8).

This house showed no set patterns with any of the metrological factors under consideration; see Fig. 9. The maximum and minimum values of consumption in both phases did not correspond to the maxima or minima of any of the factors. The family members explained the maximum value to be associated with the preparation of a food festival event they participated in, whereas the minimum value corresponded to a wedding for which most of the members were out of the house. All metrological

Fig. 8 House 4 electricity consumption before- and after-feedback

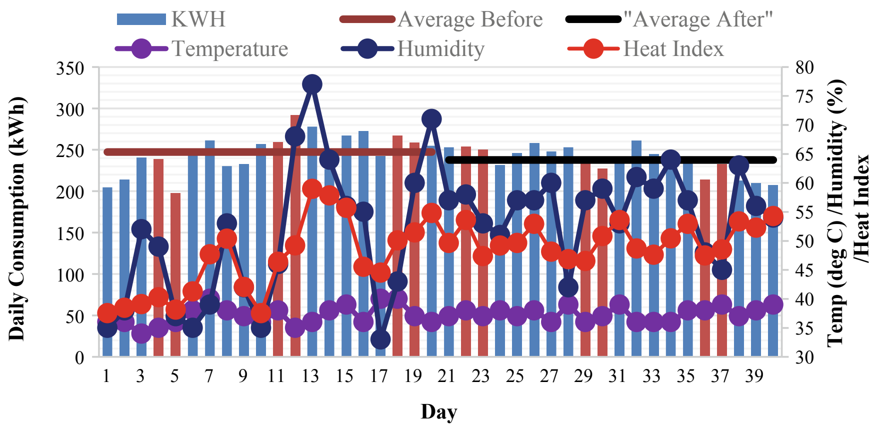


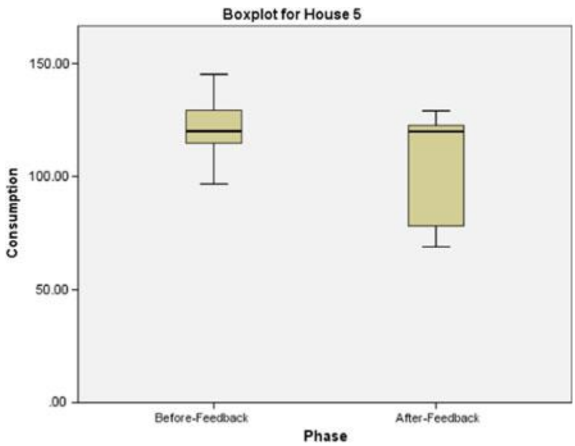
Fig. 9 Energy consumption for house 4 along with metrological factors

variables showed a considerable increase in the second phase of the experiment. Despite this increase, an overall decrease of 3.9% was observed. Though this decrease was not found to be statistically significant by means of t-test, it was observed despite the increase in variables like temperature, humidity, and heat index which increased by 1.1%, 12.4%, and 8.3%, respectively.

3.5 *House 5 Analysis*

The boxplot in Fig. 10 does not show much of a difference in the median but the

Fig. 10 House 5 electricity consumption before- and after-feedback



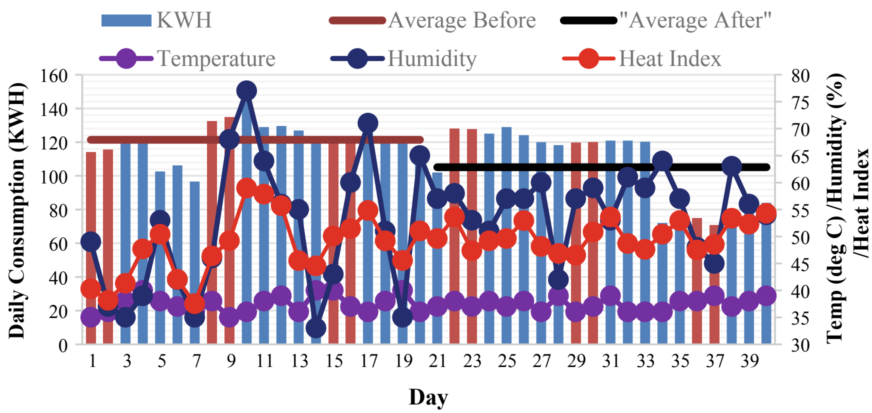


Fig. 11 Energy consumption for house 5 along with metrological factors

lower whisker has shown a noticeable decrease as well as the upper whisker has come down. No outliers were observed in the boxplot.

Figure 11 shows the data for house 5 which also showed no set pattern with the metrological factors. Humidity was again found to be the most dominant factor and the main culprit for increased consumption. The maximum value of the entire experiment corresponded to the maximum humidity level. On the other hand, the minimum values showed no pattern or correlation with the variables. The family members associated with these values with family outings instead. The second phase of the experiment despite an increase in the humidity showed a great amount of reduction in the usage. Mann–Whitney test was used to get a p-value of 0.032, thus concluding that the 13.3% reduction in electricity consumption was statistically significant. The weather conditions were much harsher in the second phase of the experiment when the household was provided with the feedback device, yet they still managed to reduce their consumption. The temperature average remained the same in both phases but the humidity and heat index increased by 9.6% and 4.8%, respectively.

4 Conclusion and Recommendations

An average reduction of 12% with variability from 3.9 to 25.7% savings was observed for the 5 households under the experiment. Though no statistical dependence of consumption on metrological variables could be established (as no definite patterns were observed), higher values of humidity were observed to play a role in increasing consumption. The results of all 5 houses supported the hypothesis that "Real-time cost feedback will help conserve electricity in the Kingdom of Bahrain." Thus, feedback technologies showed a great potential to encourage electricity consumers in the Kingdom of Bahrain to cut back their consumption and reduce electricity wastage.

References

1. Bin Ali Mirza, A.: Is Bahrain ready for sustainable energy? Gulf Insider (2017). http://www.gulf-insider.com/bahrain-ready-sustainable-energy/. Accessed 11 Nov. 2017
2. Worlddata.info: Energy consumption in Bahrain (2017). https://www.worlddata.info/asia/bahrain/energy-consumption.php. Accessed 23 Jun. 2017
3. Alahmad, M., Wheeler, P., Schwer, A., Eiden, J., Brumbaugh, A.: A comparative study of three feedback devices for residential real-time energy monitoring. IEEE Trans. Industr. Electron. **59**(4), 2002–2013 (2012)
4. Oltra, C., Boso, A., Espluga, J., Prades, A.: A qualitative study of users' engagement with real-time feedback from in-house energy consumption displays. Energy Policy **61**, 788–792 (2013)
5. Houde, S., Todd, A., Sudarshan, A., Carrie Armel, K.: Real-time feedback and electricity consumption: a field experiment assessing the potential for savings and persistence. Energy J. **34**(1) (2013)
6. Bluelineinnovations.com: Power Cost Monitor | Features (2016). https://www.bluelineinnovations.com/features. Accessed 10 Aug. 2017

Synchronization Control in Fractional Discrete-Time Systems with Chaotic Hidden Attractors

Adel Ouannas, Giuseppe Grassi, Ahmad Taher Azar, and Amina Aicha Khennaoui

Abstract This paper is addressed to investigate the synchronization between two-dimensional fractional discrete-time systems with chaotic hidden attractors. A non-linear control strategy is proposed for the complete synchronization based on the stability theory of linear fractional discrete-time systems. Some numerical examples have been carried out to test the validity of the synchronization methods developed in this paper.

Keywords Chaos · Discrete chaos · Synchronization · Hidden fractional discrete calculus · Lyapunov stability

1 Introduction

Fractional calculus is a subject which is over 300 years old. Such mathematical phenomena allow a real object to be represented more precisely than the conventional

A. Ouannas
Department of Mathematics, LAMIS Laboratory,
University of Larbi Tebessi, 12002 Tebessa, Algeria
e-mail: ouannas.a@yahoo.com

G. Grassi
Dipartimento Ingegneria Innovazione, Università del Salento, 73100 Lecce, Italy
e-mail: giuseppe.grassi@unisalento.it

A. T. Azar (✉)
Robotics and Internet-of-Things Lab (RIOTU), Prince Sultan University, Riyadh, Saudi Arabia
e-mail: ahmad_T_azar@ieee.org; aazar@psu.edu.sa

Faculty of computers and Artificial Intelligence, Benha University, Banha, Egypt
e-mail: ahmad.azar@fci.bu.edu.eg

A. A. Khennaoui
Departement of Mathematics and Computer Sciences,
University of Larbi Ben M'hidi, Oum El Bouaghi, Algeria
e-mail: kaminaaicha@yahoo.fr

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_59

method of integer-order. The FC was therefore used in engineering and scientific approaches in various fields [2, 5, 6, 9, 11, 13, 19, 20, 30, 35, 36].

Discrete-time chaotic systems play an important role in practice. Several mathematical models in all kinds of fields have been proposed in the literature using non-linear discrete models [10]. Hence, the study of synchronization in chaotic discrete systems has recently increased [38], and since then it has attracted more attention of researchers. [18, 22, 23, 26, 31, 33, 34, 37, 39]. Recently, interest has grown in the subject of chaotic systems with hidden attractors. From a computational perspective, they refer to attractors where there basin of attraction does not connect to any fixed point. Hidden attractors can be chaotic or periodic. They appear in systems to which any existing fixed point is stable or with no fixed point [15]. This property can be applied in many areas of science and engineering [24]. Up to now, there exists only a few investigations of hidden chaotic attractors in discrete-time models [14, 16, 17].

Recently, a series of works about fractional discrete calculus and its application has been published [12, 25]. This gave rise to many fractional discrete-time chaotic systems such as [18, 22, 23, 27, 29, 32, 34, 40]. The synchronization of such discrete-time systems have been investigated [7, 21, 28]. Studies have shown that fractional discrete-time chaotic systems have rich dynamics with both system parameters and fractional order. On the other hand, chaos synchronization of discrete dynamical system with fractional order and hidden chaotic attractor has not been studied so far. Thus, it remains to this day a new topic and unexplored field. The purpose of this work is to propose a control scheme between two fractional maps with hidden chaotic attractors to achieve synchronization. A numerical example is simulated to verify the effectiveness of the nonlinear controller.

The rest of the paper is organized as follows. In Sect. 2, preliminaries of fractional calculus is proposed. Section 3 describes the master and slave systems. In Sect. 4, a synchronization scheme is introduced. Finally, concluding comments are summarize in Sect. 5.

2 Preliminaries

In this section, preliminaries of fractional calculus are presented.

Definition 1 ([4]) The fractional sum of order $\nu > 0$ of a function $X(t) : \mathbb{N}_a \to \mathbb{R}$, with $\mathbb{N}_a = \{a + n : n \in \mathbb{N}\}$

$$\Delta_a^{-\nu} X(t) = \frac{1}{\Gamma(\nu)} \sum_{s=a}^{t-\nu} \frac{\Gamma(t-s)}{\Gamma(t-s-\nu+1)} X(s), \tag{1}$$

where Γ is $\Gamma-$ function.

Definition 2 ([1]) For $0 < \nu < 1$ and $X(t) \in \mathbb{N}_a$, the νth fractional-order Caputo-like difference operator is defined as:

$$ {}^{C}\Delta_a^{\nu} X(t) = \Delta_a^{t-(1-\nu)} \Delta X(t) = \frac{1}{\Gamma(1-\nu)} \sum_{s=a}^{t-(1-\nu)} \frac{\Gamma(t-s)}{\Gamma(t-s-\nu+1)} \Delta_s X(s), \quad (2) $$

where $t \in \mathbb{N}_{a+n-\upsilon}$.

Theorem 1 *Consider for $\nu \in (0, 1]$, the following fractional difference equation:*

$$ \begin{cases} {}^{C}\Delta_a^{\nu} u(t) = f(t+\nu-1, u(t+\nu-1)), \\ u(a) = u_0, \end{cases} \quad (3) $$

which is equivalent to the integral equation given by [3]:

$$ u(t) = u_0 + \frac{1}{\Gamma(\upsilon)} \sum_{s=1-\nu}^{t-\nu} \frac{\Gamma(t-s)}{\Gamma(t-s-\nu+1)} f(s+\nu-1, u(s+\nu-1)), t \in \mathbb{N}_1, \quad (4) $$

where u_0 is the initial condition.

Theorem 2 ([8]) *The zero equilibrium of the following linear fractional discrete system is: discrete-time system*

$$ {}^{C}\Delta_a^{\nu} e(t) = \mathbf{M} e(t+\nu-1), \quad (5) $$

with $e \in \mathbb{R}^n$, $0 < \nu \leq 1$, $\mathbf{M} \in R^{n \times n}$ and $\forall t \in \mathbb{N}_{a+1-\nu}$, is asymptotically stable if $|\lambda| < \left(2\cos\frac{|\arg|-\pi}{2-\upsilon}\right)^{\nu}$ and $|\arg \lambda| > \frac{\nu\pi}{2}$, are satisfied for all the eigenvalues λ of **M**.

3 Master–Slave Systems

Here, the master system describes a fractional-order model of two nonlinear difference equation [29] is given as:

$$ \begin{cases} {}^{C}\Delta_a^{\nu} x_m(t) = y_m(t-1+\nu) - x_m(t-1+\nu), \\ {}^{C}\Delta_a^{\nu} y_m(t) = x_m(t-1+\nu) + a_1^2 x_m(t-1+\nu) + a_2 y_m^2(t-1+\nu) \\ \qquad -a_3 x_m(t-1+\nu) y_m(t-1+\nu) - y_m(t-1+\nu) - a_4, \end{cases} \quad (6) $$

where a_1, a_2, a_3, a_4 are bifurcation parameters, $\nu \in [0, 1[$ is the fractional order and $t \in \mathbb{N}_{a+1-\nu}$. In the following, we refer to Eq. (6) as master system where the subscript m in the states x and y stand for the master system. To the best of our knowledge, the discrete model (9) is the first example of fractional map that presents hidden chaotic attractor. For example, when $(a_1, a_2, a_3, a_4) = (0.2, 0.71, 0.91, 1.14)$,

$(x(0), y(0)) = (0.93, -0.44)$, $\nu = 0.985$ and $a = 0$, the map possesses hidden chaotic attractor. Numerical simulation in Fig. 1 illustrates the dynamical behavior of the model (6) versus a_2.

The slave system is given by:

$$\begin{cases} {}^{C}\Delta_a^{\nu} x_s(t) = y_s(t-1+\nu) + u_1, \\ {}^{C}\Delta_a^{\nu} y_s(n) = -\alpha|y_s(t-1+\nu)| - x_s(t-1+\nu)y_s(t-1+\nu) \\ \qquad\qquad + \beta x_s^2(t-1+\nu) - \gamma y_s^2(t-1+\nu) + \delta + u_2. \end{cases} \tag{7}$$

Similarly, the slave system is denoted by s and the function u_i, $i = 1, 2$, is the synchrinization controller to be determined later. The uncontrolled slave system (7) possesses a hidden chaotic attractor when the parameters $(\alpha, \beta, \gamma, \delta)$ are taken as $(0.01, 0.1, 2, 0.1)$, $x(0) = 1.5$, $y(0) = 0.5$, and $\nu = 0.985$. Figure 2 shows the dynamical behavior of the model (7) versus γ.

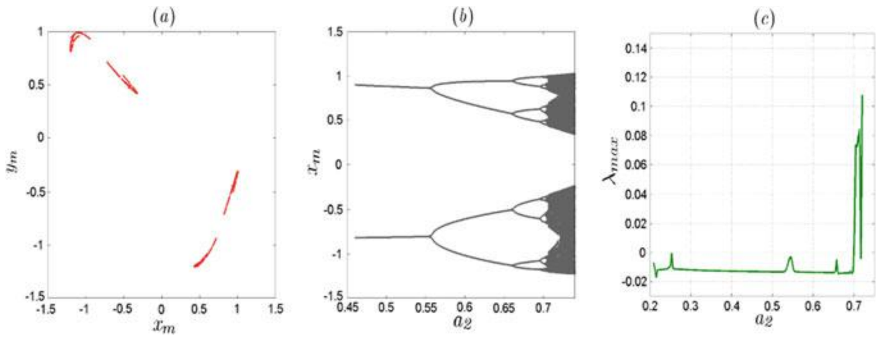


Fig. 1 **a** Hidden chaotic attractor of the master system (6) for $\nu = 0.985$; **b** bifurcation diagram for the master system (6) versus a_2 for $\nu = 0.985$; **c** largest Lyapunov exponent diagram versus a_2 for $\nu = 0.985$

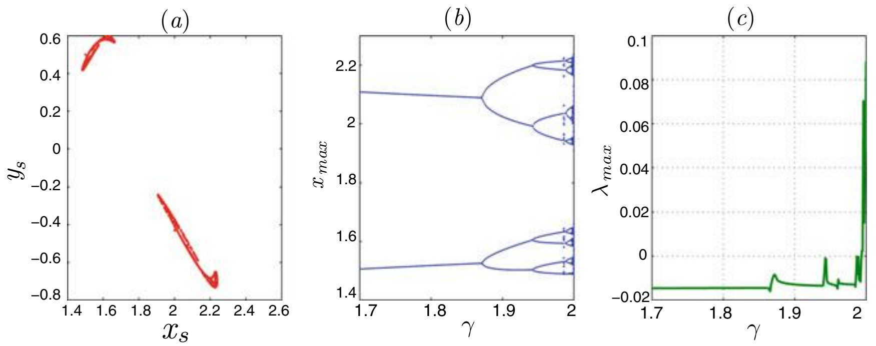


Fig. 2 **a** Hidden chaotic attractor of the slave system (7) without control for $\nu = 0.985$; **b** bifurcation diagram for the slave system (7) versus γ for $\nu = 0.985$; **c** largest Lyapunov exponent diagram versus γ for $\nu = 0.985$

In the following, a complete synchronization scheme is presented by designing a proper controller u such that the error between the two dynamical systems (6) and (7) is asymptotically stable.

4 Synchronization Scheme

The goal of synchronization is to realize harmonica between the slave system (7) and master system (6), by forcing a synchronization error to zero. In this case, the synchronization error (8) has the following fractional-order dynamical system:

$$\begin{cases} {}^C\Delta_a^\nu e_1\,(t) = y_s(t-1+\nu) + u_1\,(t) - y_m(t-1+\nu) + x_m(t-1+\nu), \\ {}^C\Delta_a^\nu e_2\,(t) = -\alpha|y_s(t-1+\nu)| - x_s(t-1+\nu)y_s(t-1+\nu) \\ \qquad\qquad + \beta x_s^2(t-1+\nu) - \gamma y_s^2(t-1+\nu) + \delta + u_2\,(t) \\ \qquad\qquad -x_m(t-1+\nu) - a_1^2 x_m(t-1+\nu) - a_2 y_m^2(t-1+\nu) \\ \qquad\qquad + a_3 x_m(t-1+\nu)y_m(t-1+\nu) + y_m(t-1+\nu) + a_4, \end{cases} \tag{8}$$

where $e_1 = x_s - x_m$ and $e_2 = y_s - y_m$. It is noted that the slave system (7) is completely synchronized with master system (6), if there exists a controller u_1, u_2 such that the states of the fractional-error discrete-time system (8) are stabilized. Theorem 3 present the main results:

Theorem 3 *The synchronization of systems (6) and (7) can be achieved, if the controller parameter u_1 and u_2 are defined as:*

$$\begin{cases} u_1 = x_s - 2x_m, \\ u_2 = -\frac{9}{4}x_s + \frac{13}{4}x_m - 2y_s + y_m + \alpha|y_s| + x_s y_s - \beta x_s^2 \\ \qquad +\gamma y_s^2 - \delta + a_1^2 x_m + a_2 y_m^2 - a_3 x_m y_m - a_4. \end{cases} \tag{9}$$

Proof Substituting Eq. (9) into Eq. (8) yields:

$$\begin{cases} {}^C\Delta_a^\upsilon e_1\,(t) = e_1(t-1+\upsilon) + e_2(t-1+\upsilon), \\ {}^C\Delta_a^\upsilon e_2\,(t) = -\frac{9}{4}e_1(t-1+\upsilon) - 2e_2(t-1+\upsilon). \end{cases} \tag{10}$$

Next, it is observed that the zero equilibrium of (10) converge to zero. The fractional-error dynamical system(7) can be rewritten as the matrix form:

$$ {}^C\Delta_a^\upsilon\,(e_1\,(t)\,, e_2\,(t))^T = \mathbf{M} \times (e_1\,(t-1+\upsilon)\,, e_2\,(t-1+\upsilon))^T\,, \tag{11}$$

where

$$\mathbf{M} = \begin{pmatrix} 1 & 1 \\ -\frac{9}{4} & -2 \end{pmatrix}. \tag{12}$$

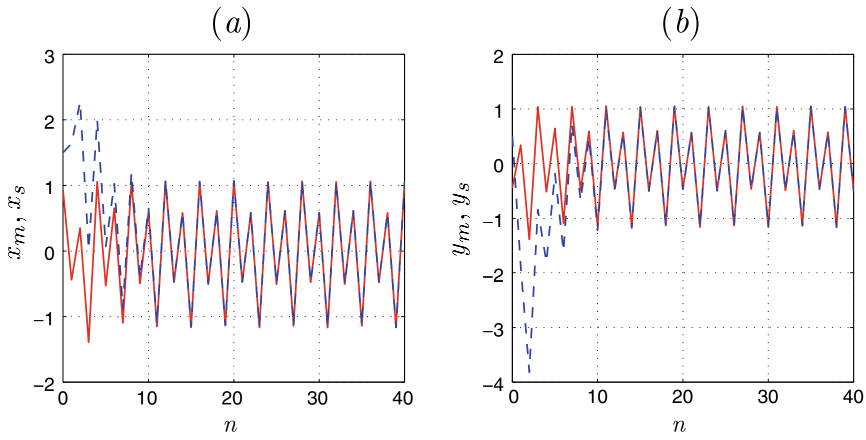


Fig. 3 Synchronization states of master and slave systems (9–10)

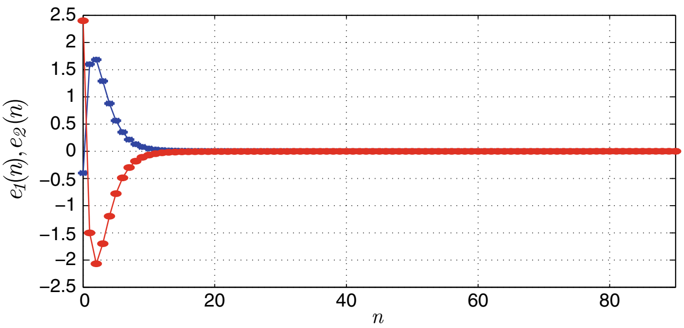


Fig. 4 Evolution of errors (10); (e_1 blue dots and error e_2 red dots)

Based on stability theorem 2, one can see that the eigenvalues of the matrix **M** satisfy the condition (2). This indicates that the zero solution of a system (11) is globally asymptotically stable. Hence, the systems (6)–(7) are synchronized.

Figure 3 shows the variation of states variable of systems (6)–(7) and the states e_1, e_2 of the errors fractional system (10) are depicted in Fig. 4, respectively. It is obvious that the proposed synchronization is achieved well for these master and slave systems.

5 Conclusion

In this article, a control scheme is proposed to synchronize discrete models with hidden chaotic attractors characterized by fractional forms. The proposed nonlinear controller was designed by linearization method. The corresponding numerical simulation confirms that the response fractional discrete model is well synchronized with the master system with respect to the nonlinear controller.

References

1. Abdeljawad, T.: On riemann and caputo fractional differences. Comput. Math. Appl. **62**(3), 1602–1611 (2011)
2. Ammar, H.H., Azar, A.T.: Robust path tracking of mobile robot using fractional order pid controller. In: Hassanien, A.E., Azar, A.T., Gaber, T., Bhatnagar, R., Tolba, M.F. (eds.) The International, Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019), pp. 370–381. Springer International Publishing, Cham (2020)
3. Anastassiou, G.A.: Principles of delta fractional calculus on time scales and inequalities. Math. Comput. Model. **52**(3), 556–566 (2010)
4. Atici, F., Eloe, P.: Discrete fractional calculus with the nabla operator. Electronic Journal of Qualitative Theory of Differential Equations [electronic only] 2009:Paper No. 3, 12 p., electronic only–Paper No. 3, 12 p., electronic only (2009). http://eudml.org/doc/227223
5. Azar, A.T., Serrano, F.E.: Fractional order two degree of freedom pid controller for a robotic manipulator with a fuzzy type-2 compensator. In: Hassanien, A.E., Tolba, M.F., Shaalan, K., Azar, A.T. (eds.) Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018, pp. 77–88. Springer International Publishing, Cham (2019)
6. Azar, A.T., Vaidyanathan, S., Ouannas, A. (eds.): Robust Adaptive Supervisory Fractional Order Controller for Optimal Energy Management in Wind Turbine with Battery Storage, pp. 165–202. Springer International Publishing, Cham (2017)
7. Bendoukha, S., Ouannas, A., Wang, X., Khennaoui, A., Pham, V.T., Grassi, G., Huynh, V.: The co-existence of different synchronization types in fractional-order discrete-time chaotic systems with non-identical dimensions and orders. Entropy **20**(710), 1–13 (2018)
8. Cermak, J., Gyori, I., Nechvatal, L.: On explicit stability conditions for a linear fractional difference system. Fract. Calc. Appl. Anal. **18**(3), 651–672 (2015)
9. Djeddi, A., Dib, D., Azar, A., Abdelmalek, S.: Fractional order unknown inputs fuzzy observer for takagi-sugeno systems with unmeasurable premise variables. Mathematics **7**(10), 984 (2019)
10. Elaydi, S.N.: Discrete Chaos: With Applications in Science and Engineering. Chapman and Hall/CRC, Boca Raton (2007)
11. Ghoudelbourk, S., Dib, D., Omeiri, A., Azar, A.T.: Mppt control in wind energy conversion systems and the application of fractional control (piα) in pitch wind turbine. Int. J. Model. Identif. Control **26**(2), 140–151 (2016)
12. Goodrich, C., Peterson, A.C.: Discrete Fractional Calculus. Springer, Cham (2015)
13. Gorripotu, T.S., Samalla, H., Jagan Mohana Rao, C., Azar, A.T., Pelusi, D.: Tlbo algorithm optimized fractional-order pid controller for agc of interconnected power system. In: Nayak, J., Abraham, A., Krishna, B.M., Chandra Sekhar, G.T., Das, A.K. (eds.) Soft Computing in Data Analytics, pp. 847–855. Springer, Singapore (2019)
14. Huynh, V.V., Ouannas, A., Wang, X., Pham, V.T., Nguyen, X.Q., Alsaadi, F.E.: Chaotic map with no fixed points: entropy, implementation and control. Entropy **21**(3), 279 (2019)
15. Jafari, S., Sprott, J.C., Nazarimehr, F.: Recent new examples of hidden attractors. Eur. Phys. J. Spec. Top. **224**(8), 1469–1476 (2015)

16. Jiang, H., Liu, Y., Wei, Z., Zhang, L.: Hidden chaotic attractors in a class of two-dimensional maps. Nonlinear Dynam. **85**(4), 2719–2727 (2016a)
17. Jiang, H., Liu, Y., Wei, Z., Zhang, L.: A new class of three-dimensional maps with hidden chaotic dynamics. Int. J. Bifurc. Chaos **26**(12), 1650,206 (2016b). https://doi.org/10.1142/S0218127416502060
18. Jouini, L., Ouannas, A., Khennaoui, A.A., Wang, X., Grassi, G., Pham, V.T.: The fractional form of a new three-dimensional generalized hénon map. Adv. Differ. Equ. **122**, 1–12 (2019)
19. Khan, A., Singh, S., Azar, A.T.: Combination-combination anti-synchronization of four fractional order identical hyperchaotic systems. In: Hassanien, A.E., Azar, A.T., Gaber, T., Bhatnagar, R., Tolba, M.F. (eds.) The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019), pp. 406–414. Springer International Publishing, Cham (2020a)
20. Khan, A., Singh, S., Azar, A.T.: Synchronization between a novel integer-order hyperchaotic system and a fractional-order hyperchaotic system using tracking control. In: Hassanien, A.E., Azar, A.T., Gaber, T., Bhatnagar, R., Tolba, F.M. (eds.) The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019), pp. 382–391. Springer International Publishing, Cham (2020b)
21. Khennaoui, A.A., Ouannas, A., Bendoukha, S., Grassi, G., Wang, X.: Pham VT (2018a) Generalized and inverse generalized synchronization of fractional-order discrete-time chaotic systems with non-identical dimensions. Adv. Differ. Equ. **1**, 303 (2018)
22. Khennaoui, A.A., Ouannas, A., Bendoukha, S., Wang, X., Pham, V.T.: On chaos in the fractional-order discrete-time unified system and its control synchronization. Entropy **20**(7), 530 (2018b)
23. Khennaoui, A.A., Ouannas, A., Bendoukha, S., Grassi, G., Lozi, R.P., Pham, V.T.: On fractional-order discrete-time systems: chaos, stabilization and synchronization. Chaos Solitons Fractals **119**, 150–162 (2019)
24. Kuznetsov, N.V., Leonov, G.A., Yuldashev, M.V., Yuldashev, R.V.: Hidden attractors in dynamical models of phase-locked loop circuits: limitations of simulation in matlab and spice. Commun. Nonlinear Sci. Numeri. Simul. **51**, 39–49 (2017)
25. Ostalczyk, P.: Discrete Fractional Calculus. World Scientific, Singapore (2016). https://doi.org/10.1142/9833
26. Ouannas, A., Khennaoui, A.A., Bendoukha, S., Vo, T., Pham, V.T., Huynh, V.: The fractional form of the tinkerbell map is chaotic. Appl. Sci. **8**(12), 2640 (2018a)
27. Ouannas, A., Khennaoui, A.A., Grassi, G., Bendoukha, S.: On the q-s chaos synchronization of fractional-order discrete-time systems: general method and examples. Discrete Dynam. Nat. Soc. **2018**(Article ID 2950357), 1–8 (2018b)
28. Ouannas, A., Khennaoui, A.A., Grassi, G., Bendoukha, S.: On the q-s chaos synchronization of fractional-order discrete-time systems: general method and examples. Discrete Dyn. Nat. Soc. **2950357**, 1–8 (2018c)
29. Ouannas, A., Wang, X., Khennaoui, A., Bendoukha, S., Pham, V.T., Alsaadi, F.: Fractional form of a chaotic map without fixed points: chaos, entropy and control. Entropy **20**(720), 1–14 (2018d)
30. Ouannas, A., Grassi, G., Azar, A.T., Singh, S.: New control schemes for fractional chaos synchronization. In: Hassanien, A.E., Tolba, M.F., Shaalan, K., Azar, A.T. (eds.) Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018, pp 52–63. Springer International Publishing, Cham (2019a)
31. Ouannas, A., Khennaoui, A.A., Bendoukha, S., Grassi, G.: On the dynamics and control of a fractional form of the discrete double scroll. Int. J. Bifurc. Chaos **29**(06), 1950,078 (2019b)
32. Ouannas, A., Khennaoui, A.A., Bendoukha, S., Grassi, G.: On the dynamics and control of a fractional form of the discrete double scroll. Int. J. Bifurc. Chaos **29**(06), 1950,078 (2019c). https://doi.org/10.1142/S0218127419500780
33. Ouannas, A., Khennaoui, A.A., Grassi, G., Bendoukha, S.: On chaos in the fractional-order grassi-miller map and its control. J. Comput. Appl. Math. **358**, 293–305 (2019d)

34. Ouannas, A., Khennaoui, A.A., Odibat, Z., Pham, V.T., Grassi, G.: On the dynamics, control and synchronization of fractional-order ikeda map. Chaos Solitons Fractals **123**, 108–115 (2019e)
35. Ouannas, A., Grassi, G., Azar, A.T.: Fractional-order control scheme for q-s chaos synchronization. In: Hassanien, A.E., Azar, A.T., Gaber, T., Bhatnagar, R., Tolba, F.M. (eds.) The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019), pp. 434–441. Springer International Publishing, Cham (2020a)
36. Ouannas, A., Grassi, G., Azar, A.T.: A new generalized synchronization scheme to control fractional chaotic systems with non-identical dimensions and different orders. In: Hassanien, A.E., Azar, A.T., Gaber, T., Bhatnagar, R., Tolba, F.M. (eds.) The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019), pp. 415–424. Springer International Publishing, Cham (2020b)
37. Ouannas, A., Grassi, G., Azar, A.T., Khennaouia, A.A., Pham, V.T.: Chaotic control in fractional-order discrete-time systems. In: Hassanien, A.E., Shaalan, K., Tolba, M.F. (eds.) Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2019, pp 207–217. Springer International Publishing, Cham (2020c)
38. Ouannas A., Al-sawalha M.M.: Synchronization of Chaotic Dynamical Systems in Discrete-Time, Springer International Publishing, Cham, pp 101–132 (2016). https://doi.org/10.1007/978-3-319-30340-6_5
39. Wu, G.C., Baleanu, D.: Discrete chaos in fractional delayed logistic maps. Nonlinear Dynam. **80**(4), 1697–1703 (2015)
40. Wu, G.C., Baleanu, D., Xie, H.P., Chen, F.L.: Chaos synchronization of fractional chaotic maps based on the stability condition. Phys. A: Stat. Mech. Appl. **460**, 374–383 (2016)

Employment of Cryptographic Modus Operandi Based on Trigonometric Algorithm and Resistor Color Code

Rajdeep Chowdhury and Sukhwant Kumar

Abstract In the progressive world, safekeeping is a substantial apprehension, and formulation of inventive cryptographic modus operandi is one of the approaches to be equipped with safekeeping. Copious cryptographic modus operandi has been projected over the years, and most of them have its stake of shortcoming.The work-in-hand exemplifies a modus operandi by which an assortment of statements could be verified to ensure cossetted safekeeping. In the projected algorithm, the plain text is encoded with trigonometric function, ASCII table and resistor color code to formulate the ciphertext, and the content would be engendered. The unscrambled content would be hopped on by applying the switch technique, especially swapping. The calculation could be implemented on any thoughtful record as it is realized at bit level.The excellence of the devised stratagem is fragmented in the formulated paper with distinct sections. The freshly shaped cryptographic technique calculation through encryption is familiarized in the endeavor.

Keywords Trigonometry · Resistor color code · ASCII · Swapping · Encryption · Decryption

1 Introduction

In the flair orb of communiqué, it is imperative to envision information larceny and obvious inclination toward data shielding. Safekeeping has developed into a decisive facet for blooming networks and Webs. Cryptography is not only universally acknowledged but extensively implemented [5–8].

R. Chowdhury (✉)
Chinsurah, Hooghly 712101, West Bengal, India
e-mail: dujon18@yahoo.co.in

S. Kumar
Department of Mechanical Engineering, JIS College of Engineering, Block–A, Phase–III, Kalyani, Nadia 741235, West Bengal, India

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_60

The techniques have numerous applications in engineering and unalike fields, viz. shielding military communications, confidentiality of emails, fiscal information, organizational knowledge, private records, etc.

Cryptography is an imperative means of data safekeeping through mathematical operations, along with being associated in unfathomable format for unsanctioned individuals. The mechanism employs scientific outlines and procedures to scramble information into inscrutable text [10].

In the proposed work, employment of trigonometric function and resistor color code in formulation of the encryption modus operandi eventually aids in upholding safekeeping of the algorithm [1–4, 9].

Initially, in the projected work, plain text is engaged, and it regenerates into ASCII code.

Then, the ASCII code is operated with the aid of mathematical functions as the text could be redeveloped into encryption range. The operated value is arranged in ascending sequence, and the initial and the ultimate digit of apiece numeral is engaged as keys. Next, the keys are formulated in the manner, as mentioned beneath, namely

$$\mathrm{K} + \mathrm{Y} = \mathrm{KY} = \mathrm{R} \quad (1)$$

Subsequently, the ASCII values and R are swapped exclusively, amid the initial and the ultimate digit of their numerals, respectively. Finally, the swapped value would be compared with the resistor color code and assigned with color in a triplet band form [3, 4].

The formulated paper could be applied as reference paper for modern academicians and scholars undertaking research exclusively on papers interrelated to progress of cryptographic algorithms or in conjunction with data warehouse protection employing inventive and novel cryptographic modus operandi.

Apparently, the devised paper could furnish loads of prized contributions for opportune integration by particular scholars in diverse fields of work.

2 Proposed Work

The proposed work section essentially comprises the inventive notion ensured in transition amid encryption and decryption modus operandi with key engendering, along with accumulation of ciphertext, demonstrated by specimen and flowchart.

The proposed work has been pictorially demonstrated by the flowchart stated beneath (Fig. 1).

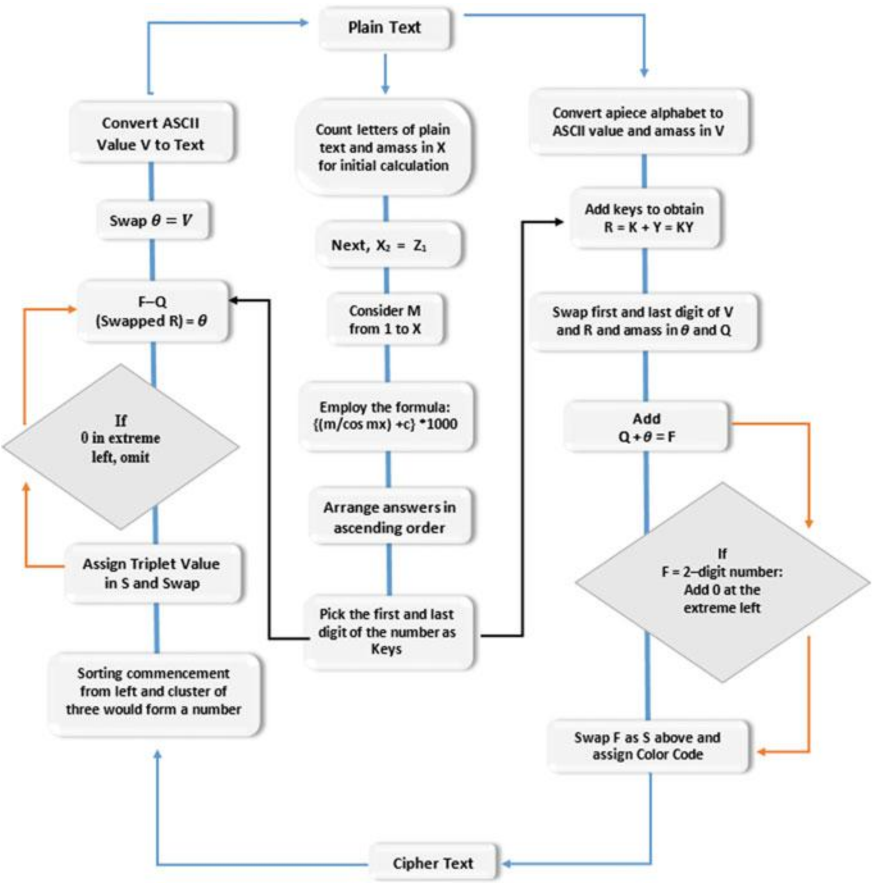


Fig. 1. Flowchart of the proposed work

2.1 Key Engendering

The key is being engendered by the formula, namely

$$Z = \left(\int m^2 \sec mx \tan mx \, dx \right) * 10000 \qquad (2)$$

Here,

$$\begin{aligned} &= d/dx[(\sec(mx)/m) + c] \\ &= (1/m)d/dx(\sec mx) + d/dx(c) \\ &= (1/m)(m \sec mx \tan mx) + 0 \\ &= \sec mx \tan mx \end{aligned}$$

So,

$$\left(\int m^2 \sec mx \tan mx \ dx\right) * 10000 = \left(\left(m^2 \sec mx/m\right) + c\right) * 10000$$
$$= [(m/(\cos mx)) + c] * 10000 \tag{3}$$

where 10000 is the multiplier, which confirms the outcome to be an integer value rather than being a floating point value taking up to 4 decimal places.

Initially, Z_1 "X" is the value that describes the number of letters taken by the user.

After that,

$$\begin{aligned} X &= Z_{1.} \\ X &= Z_2 \\ &\vdots \\ X &= Z_{(x-1)} \end{aligned} \tag{4}$$

"M" is another variable which has range from 1 to number of letters taken, and it increments by 1.

C is constant, and its value is 1.

The sign "+" or "–" in the outcome is overlooked.

Finally, the outcome is arranged in ascending order, and the initial and the ultimate digits of the numeral are selected, which are the engendered key "K" and "Y".

2.2 Encryption Modus Operandi

The encryption modus operandi is explicated for ease in understanding, namely

Using the formula, the key is engendered.

The key is employed to encrypt the furnished text.

Next, apiece character in the furnished text is converted to their respective ASCII values and named "V".

Next, ASCII values of first and last digit would be swapped and amassed in "θ".

$$123 = 321$$

Now, both the keys K and Y are solved employing the formula.

$$K + Y = KY = R \tag{5}$$

The outcome "R" is then swapped in the aforementioned identical manner and amassed in "Q".

Note: In swapping, 0 would be considered as a digit and would not be omitted.

Then, the swapped values are added to obtain “F”.

$$\theta + Q = F \tag{6}$$

Note: If “F” is obtained as two-digit number, 0 is placed in the extreme left of the number.

Subsequently, it would be swapped in the aforementioned manner to obtain “S”, and then, the usual color code for resistors with apiece digit and respective color is assigned for obtaining the ciphertext.

2.3 Decryption Modus Operandi

The decryption modus operandi is explicated for ease in understanding, namely

The ciphertext is sorted, either from left or right, in cluster of triplets.

Next, apiece triplet is compared with the resistor color code, and the respective digit or numeric value from left-hand side is amassed in “S”.

The first and the last digit of “S” are swapped with one another and amassed in “F”.

Note: If 0 is retrieved at the extreme left of the number “F”, it would be omitted.

Subtract Q from it to get “θ,”

$$\theta = F - Q \tag{7}$$

After obtaining “θ,” swap again to obtain “V”.

Finally, after acquiring the ASCII values, the corresponding characters or the plain text could be effortlessly retrieved.

2.4 Example Elucidation

Explication with apposite example is specified beneath, namely

“@12Saket” is the plain text to be encrypted.

Hence, eight numbers would be engendered.

Let, X be 8 and M be another variable incrementing by 1.

C is constant with value 1.

The sign “+” or “–” in the outcome is overlooked.

Subsequently,

Table 1. Key engendering

Numerals	Key 1 (K)	Key 2 (Y)
20098	2	8
25765	2	5
105911	1	1
130000	1	0
136190	1	0
175342	1	2
320685	3	5
413113	4	3

$$[(1/(\cos 1 * 8)) + \mathrm{c}] * 10000 = 20098$$
$$[(2/(\cos 2 * 20098)) + \mathrm{c}] * 10000 = 25765$$
$$[(3/(\cos 3 * 25765)) + \mathrm{c}] * 10000 = 105911$$
$$[(4/(\cos 4 * 105911)) + \mathrm{c}] * 10000 = 175342$$
$$[(5/(\cos 5 * 175342)) + \mathrm{c}] * 10000 = 136190$$
$$[(6/(\cos 6 * 136190) + \mathrm{c}] * 10000 = 130000$$
$$[(7/(\cos 7 * 130000) + \mathrm{c}] * 10000 = 413113$$
$$[(8/(\cos 8 * 413113) + \mathrm{c}] * 10000 = 320685$$

Finally, the outcome is arranged in ascending order, and the ensuing list transpires, namely (20098, 25765, 105911, 130000, 136190, 175342, 320685, 413113).

Then, the first and the last digits are selected, and it forms the keys "K" and "Y". (Tables 1, 2, 3).

3 Result Analysis

In this section, comparative learning amid the stated modus operandi, namely TRCC, AES (128 bits) and Triple–DES (168 bits) have been performed on fifteen files (.doc) with varying source file sizes. Outcome has been engendered for encryption and decryption time contrast (Table 4).

Apiece simulation program for the devised algorithm calculates the entire encryption time and decryption time at culmination. Time engaged is the difference amid processor clock ticks at commencement and culmination of implementation. All time ticks are in milliseconds (ms). Lesser the time engaged, the healthier it is for an end user. Since CPU clock ticks are engaged as time, trivial discrepancy in real time might be encountered (Figs. 2 and 3).

Table 2. Encryption modus operandi

Character	ASCII Value (V)	Keys (K), (Y)	R (K + Y = KY)	Swap (V) = θ	Swap (R) = Q	Sum F (Q +θ)	Swap (F) = S	Converted color code
@	64	2, 8	28	46	82	128	821	
1	49	2, 5	25	94	52	146	641	
2	50	1, 1	11	05	11	016	610	
S	83	1, 0	10	38	01	039	930	
a	97	1, 0	10	79	01	080	080	
k	107	1, 2	12	701	21	722	227	
e	101	3, 5	35	101	53	154	451	
t	116	4, 3	43	611	34	645	546	

Table 3. Decryption modus operandi

Converted color code	Value (S)	Swap (S) = F	Subtract (F–Q) = θ	Swap θ= (V)	Value (Q)	Swap (Q) = R	Keys {R = KY = K + Y}	ASCII value (V)	Character
	821	128	128–82 = 46	64	82	28	2, 8	64	@
	641	146	146–52 = 94	49	52	25	2, 5	49	1
	610	016	016–11 = 05	50	11	11	1, 1	50	2
	930	039	039–01 = 38	83	01	10	1, 0	83	S
	080	080	080–01 = 79	97	01	10	1, 0	97	a
	227	722	722–21 = 701	107	21	12	1, 2	107	k
	451	154	154–53 = 101	101	53	35	3, 5	101	e
	546	645	645–34 = 611	116	34	43	4, 3	116	t

Exception: If K1, K2 and ASCII value is adhered as 0 (applicable in terms of null character), ciphertext would be considered as 000 and vice versa for decryption modus operandi

Table 4. Encryption and decryption time breakup (For.doc files)

Sl.	Source file name	Source file size	TRCC (In ms)		AES (In ms)		TDES (In ms)	
		(In bytes)	Enc.	Dec.	Enc.	Dec.	Enc.	Dec.
1	F01.doc	24,164	16	01	01	32	17	17
2	F02.doc	39,524	17	16	17	16	17	01
3	F03.doc	50,788	17	01	17	16	17	17
4	F04.doc	72,292	16	01	17	17	16	17
5	F05.doc	117,860	16	17	17	17	33	32
6	F06.doc	351,332	16	33	17	32	110	79
7	F07.doc	544,356	33	16	17	47	126	111
8	F08.doc	749,668	32	32	32	48	142	157
9	F09.doc	1,130,596	78	32	33	63	235	329
10	F10.doc	1,596,104	100	113	64	79	563	329
11	F11.doc	2,019,428	133	136	64	95	408	422
12	F12.doc	2,287,204	141	147	126	189	673	485
13	F13.doc	2,724,452	188	221	79	126	673	563
14	F14.doc	3,043,940	199	202	79	126	813	657
15	F15.doc	3,347,556	215	213	173	141	766	861

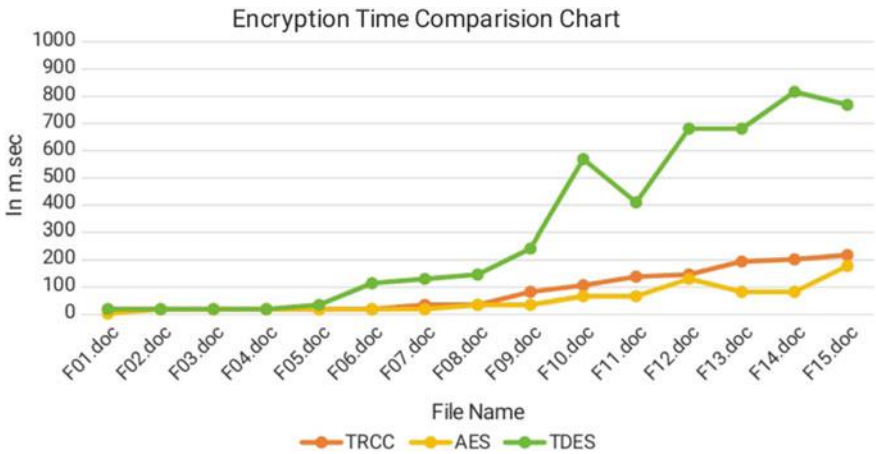


Fig. 2. Encryption time comparison chart for TRCC, AES and TDES

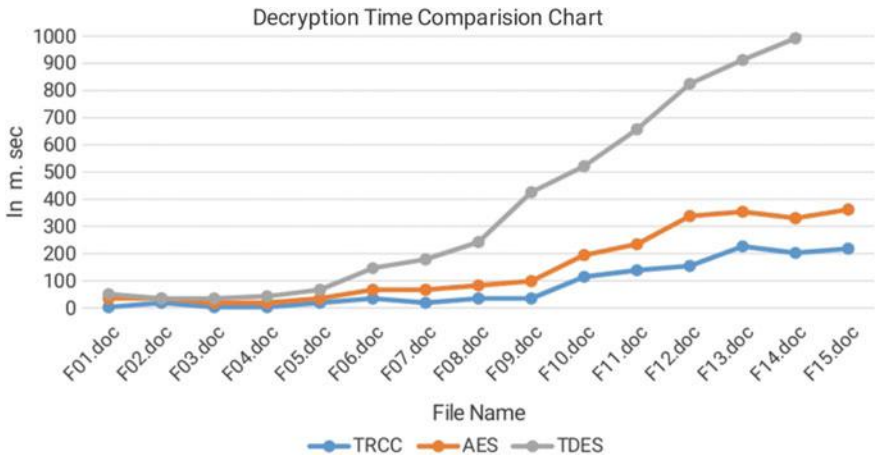


Fig. 3. Decryption time comparison chart for TRCC, AES and TDES

4 Conclusion

In the formulated paper, a newfangled cryptographic modus operandi has been implemented, which is an amalgamation of diverse disciplines, namely mathematics, physics and prominently computer science and engineering. The amalgamation of these distinct disciplines aid in the proceedings of the projected work, and the novel aspects of calculations have been ensured, inclusive of key engendering, encryption, decryption, etc. Incorporation of multiple engendered keys accentuates the safekeeping stratum and aids the users from malevolent intrusions.

References

1. Pateriya, P.K., Rai, M., Mukherjee, K.P., Raj, G.: A lightweight cryptographic algorithm using trigonometric technique. In: Proceedings of 2018 International Conference on Advances in Computing and Communication Engineering, ICACCE 2018, pp. 111–114. Published and Archived in IEEE Digital Xplore (2018). ISBN–978-1-5386-4486-7
2. Gowda, S.N., Vrishabh, D.N.: A secure trigonometry based cryptography algorithm. In: Proceedings of 2017 International Conference on Communication and Signal Processing, ICCSP 2017, pp. 0106–0109. Published and Archived in IEEE Digital Xplore (2017). ISBN–978-1-5090-3801-5
3. Kanchan, M., Shah, A., Gupta, J., Naik, S.: advance cryptography using colour code based substitution with multi–language support. In: International Journal of Computer Applications, Proceedings of The National Conference on Role of Engineers in National Building, NCRENB 2016, vol. 1, pp. 05–09 (2016). ISSN–0975 8887
4. Gaitonde, A.: Colour coded cryptography. Int. J. Sci. Eng. Res. **3**(7), 01–03 (2012). ISSN–2229 5518

5. Chowdhury, R., Roy, O., Datta, S., Dasgupta, S.: Virtual data warehouse model employing crypto–math modus operandi and intelligent sensor algorithm for cosseted transference and output augmentation. In: Knowledge Computing and Its Applications, pp. 111–129. Springer, Singapore (2018). ISBN (O)–978-981-10-6680-1; ISBN (P)–978-981-10-6679-5
6. Chowdhury, R., Datta, S., Dasgupta, S., De, M.: Implementation of central dogma based cryptographic algorithm in data warehouse for performance enhancement. Int. J. Adv. Comput. Sci. Appl. **6**(11), 29–34 (2015). ISSN (O)–2156 5570; ISSN (P)–2158 107X
7. Chowdhury, R., Dey, K.S., Datta, S., Shaw, S.: Design and implementation of proposed drawer model based data warehouse architecture incorporating dna translation cryptographic algorithm for security enhancement. In: Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2014, Mysore, pp. 55–60. Published and Archived in IEEE Digital Xplore (2014). ISBN–978-1-4799-6629-5
8. Chowdhury, R., Chatterjee, P., Mitra, P., Roy, O.: Design and implementation of security mechanism for data warehouse performance enhancement using two tier user authentication techniques. Int. J. Innov. Res. Sci. Eng. Technol. **3**(6), 165–172 (2014). ISSN (O)–2319 8753; ISSN (P)–2347 6710
9. Munoz–Rodriguez, J.: Computational cryptography based on trigonometric algorithms and intensity superposition. Imaging Sci. J. **58**(02), 61–80 (2010). ISSN (O)–1743 131X; ISSN (P)–1368 2199
10. Kahate, A.: Cryptography and Network Security, 4th Edition. Published by the Tata McGraw–Hill Publishing Company Limited, New Delhi (2008). ISBN (10)–0-07-064823-9; ISBN (13)–978-007-06-4823-4

Experimental and Dimensional Analysis Approach for Human Energy Required in Wood Chipping Process

V. M. Sonde, P. P. Shirpurkar, M. S. Giripunje, and P. P. Ashtankar

Abstract Motion is one of the most causes of energy for gathering. This article shows a serious role in the construction of human energy that is a motivation research related to wood chipping process. A successful precise model has been established for three parameters, i.e., power required for speeding up the flywheel, blood pressure rise, and time required to speed up the flywheel. The five persons of different physical characteristic were used for pedaling the energy unit and accordingly the power required per person was calculated. The process from pedaling of bicycle to the clutch engagement can be the separate field of research which is done in this research work. The separate experimental plan was established for measuring the human energy and various variables were identified for detail study. The major phase attempt has been made for calculation of human energy required for any machine operation.

Keywords Human energy · Blood pressure · Pulse rate · Flywheel · And time

1 Introduction

The wood chipper cutter using human-powered flywheel motor basically includes two phenomena [1]. First is the method from pedaling of bicycle to the clutch engagement and second is the method after engagement of clutch to the process unit. All previous researchers about human-powered concept have done the research work only after engagement of clutch to the process unit. Nobody takes attempt on the process from rider effort to the clutch the engagement in human-powered systems [2]. Here, the human energy required for pedaling the bicycling unit up to the required rpm is measured. The separate experimental plan existed advanced for measuring the human energy. The energy is measured for all gear sets according to the developed experimental plan.

When rider begins pedaling of bicycle, the load of foot touching the pedal was measured with the assistance of load cell [3]. The load acting on both the pedal

V. M. Sonde (✉) · P. P. Shirpurkar · M. S. Giripunje · P. P. Ashtankar
Priyadarshini College of Engineering, Nagpur, India
e-mail: mailtovivekms@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_61

was recorded by using two separate load cell for both the pedal. The torque is then calculated by measuring the length of sprocket arm. (Load acting on pedal into length of sprocket arm). The calculated torque is then converted into the power required by dividing total time taken for flywheel rise [4]. The separate console system was designed so as to measure the load of the rider acting on the bicycle pedal.

As this research belongs to the human energy calculation, other parameters like blood pressure and pulse rate measurement must have been measured for detailed study [5]. The initial rise in blood pressure and pulse rate was measured for each person. The sixty observations were recorded for each second and for each rider. This research includes study of various human body factors that are having a direct impact on experimentation. The number of observations taken itself shows that blood pressure rise and power produced are inversely proportional with each other likewise pulse rate rise and time [6]. Ultimately, all the variations are been studied by means of this research. The mathematical model is formed by measuring the human body variables like power required for speeding up the flywheel, blood pressure rise, and pulsing rate rise [7].

2 Identification of Variables

The various dependent and independent variables identified for measuring human energy required for wood chipper cutter are as follows [8] (Table 1):

Table 1 Variables related to human energy required for wood chipper cutter

S.N	Variables	Unit	MLT	Nature (Dependent/ Independent)
1	P_h = Power required to speed up the flywheel	N-mm/sec	ML^2T^{-1}	Dependent
2	B_p = Blood pressure rise	N/mm	MT^{-2}	Dependent
3	t_f = Time required to speed up the flywheel	Sec	T	Dependent
4	P_r = Pulse rate rise	1/sec	T^{-1}	Dependent
5	G = Gear ratio	–	$M^0L^0T^0$	Independent
6	ω_f = Angular speed of flywheel	Rad/sec	T^{-1}	Independent
7	W_p = Weight of person	N	M	Independent
8	H_p = Height of person	mm	L	Independent
9	g = Acceleration due to gravity	mm/s^2	LT^{-2}	Independent

3 Formation of Pi (π) Terms for All Dependent and Independent Variables

The Buckingham's π Theorem is used for the dimensional analysis of proposed machine [9]. The power required to speed up the flywheel P_h, blood pressure rise B_p, time required to speed up the flywheel t_f, and pulse rate rise P_r is function of gear ratio (G), acceleration due to gravity (g), angular speed of flywheel (ω_f), weight of person (W_p), and height of person (H_p)

Total number of variable = 9
Number of dependent variables = 4
Number of independent variables, n = 5, therefore

$$P_h/B_p/t_f/P_r = f\big(G, \omega_f, W_p, H_p, g\big) \quad \text{or} \quad f\big(G, \omega_f, W_p, H_p, g\big) = 0$$

g, W_p, and H_p are considered as the repeating variables, therefore, m = 3
Total number of independent variables = n = 5
Number of Π terms = n – m = 5 – 3 = 2, therefore, $\Pi_{D1} = f_1\ (\Pi_1, \Pi_2) = 0$

4 Formation of Pi Terms for Dependent Variables

4.1 *First Π Term*

$$\Pi_1 = (g)^{al}\big(W_p\big)^{b1}\big(H_p\big)^{cl}G$$
$$(M)^0(L)^0(T)^0 = \big(LT^{-2}\big)^{al}(M)^{b1}L\Big)^{c1}\big(M^0L^0T^0\big)$$

The values of a_1, b_1 and c_1 are computed by equating the power of M, L, T on both sides as given below:

For 'M'	For 'L'	For 'T'
$M \rightarrow 0 = b_1$ $b_1 = 0$	$L \rightarrow 0 = a_1 + c_1$ (From equation of T, subst. $a_1 = 0$) $0 = 0 + c_1$, Hence, $c_1 = 0$	$T \rightarrow 0 = -2a_1$ Hence, $a_1 = 0$ Subst. in equation of L to get value of b_1

Substituting the values of a_1, b_1, and c_1 in the equation of Π_1 term, we have:

$$\Pi_1 = (g)^0 (D_s)^0 (E_c)^0 G$$

$$\boldsymbol{\pi}_1 = \boldsymbol{G}$$

4.2 Second Π Term

$$\Pi_2 = (g)^{a2} \left(W_p\right)^{b2} \left(H_p\right)^{c2} \omega_f$$
$$(M)^0 (L)^0 (T)^0 = \left(LT^{-2}\right)^{a2} (M)^{b2} (L)^{c2} T^{-1}$$

The values of a_2, $b_{2,}$ and c_2 are computed by equating the power of M, L, and T on both sides as given below:

For 'M'	For 'L'	For 'T'
$M \rightarrow 0 = b_2$	$L \rightarrow 0 = a_2 + c_2$ $0 = a_2 + c_2 - 0$ (From equation of T, subst. $a_2 = -1/2$) $0 = -1/2 + c_2 - 0$, Hence, $c_2 = 1/2$	$T \rightarrow 0 = -2a_2 - 1$ $0 = -2a_2 - 0 - 1$, Hence, $a_2 = -1/2$ Subst. in equation. of L to get value of c_2

Substituting the values of a_2, $b_{2,}$ and c_2 in the equation of Π_2 term, we have: (Table 2)

$$\Pi_2 = (g)^{-1/2} \left(W_p\right)^0 \left(H_p\right)^{1/2} \omega_f,$$

$$\boldsymbol{\pi}_2 = \boldsymbol{\omega}_f \sqrt{\frac{\boldsymbol{H}_p}{\boldsymbol{g}}}$$

Table 2 Pi terms for independent variables

pi terms	pi terms equations
π_1	$\pi_1 = G$
π_2	$\pi_2 = \omega_f \sqrt{\frac{H_p}{g}}$

5 Formation of Pi Terms for Dependent Variables

5.1 First Dependent Π Term

$$\Pi_{D1} = (g)^{aD1}\left(W_p\right)^{bD1}\left(H_p\right)^{cD1}P_h$$
$$(M)^0(L)^0(T)^0 = \left(LT^{-2}\right)^{aD1}\left((M)^{bD1}L\right)^{cD1}\left(ML^2T^{-1}\right)$$

The values of a_{D1}, $b_{D1,}$ and c_{D1} are computed by equating the power of M, L, and T on both sides as given below:

For ‘M’	For ‘L’	For ‘T’
$M \rightarrow 0$ $b_{D1} + 1 = 0$ $b_{D1} = -1$	$L \rightarrow 0 = a_{D1} + c_{D1} + 2$ (From equation of T, subst. $a_{D1} = -1/2$) $0 = -1/2 + c_{D1} + 2$, Hence, $c_{D1} = -3/2$	$T \rightarrow 0 = -2a_{D1} - 1$ Hence, $a_{D1} = -1/2$ Subst. in equation of L to get value of c_{D1}

Substituting the values of a_{D1}, $b_{D1,}$ and c_{D1} in the eq. of Π_{D1} term, we have:

$$\Pi_{D1} = (g)^{-1/2}(D_s)^{-1}(E_c)^{-3/2}P_h$$

$$\boldsymbol{\pi_{D1}} = \frac{\boldsymbol{P_h}}{\boldsymbol{g^{1/2}W_pH_p^{3/2}}}$$

5.2 Second Dependent Π Term

$$\Pi_{D2} = (g)^{aD2}\left(W_p\right)^{bD2}\left(H_p\right)^{cD2}B_p$$
$$(M)^0(L)^0(T)^0 = \left(LT^{-2}\right)^{aD2}\left((M)^{bD2}L\right)^{cD2}\left(MT^{-2}\right)$$

The values of a_{D2}, $b_{D2,}$ and c_{D2} are computed by equating the power of M, L, and T on both sides as given below:

For 'M'	For 'L'	For 'T'
$M \rightarrow 0$ $b_{D2} + 1 = 0$ $b_{D2} = -1$	$L \rightarrow 0 = a_{D2} + c_{D2}$ (From equation of T, subst. $a_{D2} = -1$) $0 = -1 + c_{D2}$, Hence, $c_{D2} = 1$	$T \rightarrow 0 = -2a_{D2} - 2$ Hence, $a_{D2} = -1$ Subst. in equation of L to get value of c_{D2}

Substituting the values of a_{D2}, $b_{D2,}$ and c_{D2} in the equation of Π_{D2} term, we have:

$$\Pi_{D2} = (g)^{-1}\left(W_p\right)^{-1}\left(H_p\right)^{1}B_p$$

$$\boldsymbol{\pi}_{\boldsymbol{D}2} = \frac{B_p H_p}{g W_p}$$

5.3 Third Dependent Π Term

$$\Pi_{D3} = (g)^{aD3}\left(W_p\right)^{bD3}\left(H_p\right)^{cD3}t_f$$
$$(M)^0(L)^0(T)^0 = \left(LT^{-2}\right)^{aD3}\left((M)^{bD3}L\right)^{cD3}(T)$$

The values of a_{D3}, $b_{D3,}$ and c_{D3} are computed by equating the power of M, L, and T on both sides as given below:

For 'M'	For 'L'	For 'T'
$M \rightarrow 0$ $b_{D3} = 0$	$L \rightarrow 0 = a_{D3} + c_{D3}$ (From equation of T, subst. $a_{D3} = 1/2$) $0 = 1/2 + c_{D3}$, Hence, $c_{D3} = -1/2$	$T \rightarrow 0 = -2a_{D3} + 1$ Hence, $a_{D3} = 1/2$ Subst. in equation of L to get value of c_{D3}

Substituting the values of a_{D3}, $b_{D3,}$ and c_{D3} in the equation of Π_{D3} term, we have:

$$\Pi_{D3} = (g)^{1/2}\left(W_p\right)^{0}\left(H_p\right)^{-1/2}t_f$$

$$\boldsymbol{\pi}_{\boldsymbol{D}3} = \boldsymbol{t}_{\boldsymbol{f}}\sqrt{\frac{\boldsymbol{g}}{\boldsymbol{H}_{\boldsymbol{p}}}}$$

Table 3 Pi terms for dependent variables

pi terms	pi terms equations
π_{D1}	$\frac{P_h}{g^{1/2} W_p H_p^{3/2}}$
π_{D2}	$\frac{B_p H_p}{g W_p}$
π_{D3}	$t_f \sqrt{\frac{g}{H_p}}$
π_{D4}	$P_r \sqrt{\frac{H_p}{g}}$

5.4 *Forth Dependent Π Term*

$$\Pi_{D4} = (g)^{aD4}\left(W_p\right)^{bD4}\left(H_p\right)^{cD4} P_r$$
$$(M)^0 (L)^0 (T)^0 = \left(\left(LT^{-2}\right)^{aD4} (M)^{bD4} L\right)^{cD4} \left(T^{-1}\right)$$

The values of a_{D4}, $b_{D4,}$ and c_{D4} are computed by equating the power of M, L, and T on both sides as given below:

For 'M'	For 'L'	For 'T'
$M \to 0$ $b_{D4} = 0$	$L \to 0 = a_{D4} + c_{D4}$ (From equation of T, subst. $a_{D4} = -1/2$) $0 = -1/2 + c_{D4}$, Hence, $c_{D4} = 1/2$	$T \to 0 = -2a_{D4} - 1$ Hence, $a_{D4} = -1/2$ Subst. in equation of L to get value of c_{D4}

Substituting the values of a_{D4}, $b_{D4,}$ and c_{D4} in the eq. of Π_{D4} term, we have: (Table 3)

$$\Pi_{D4} = (g)^{-1/2}\left(W_p\right)^0\left(H_p\right)^{1/2} P_r$$

$$\boldsymbol{\pi_{D4} = P_r \sqrt{\frac{H_p}{g}}}$$

6 Formulation of Experimental Data Based Model

$$\frac{P_h}{g^{1/2} W_p H_p^{3/2}} = f_1 \left\{ (G) \left(\omega_f \sqrt{\frac{H_p}{g}} \right) \right\} \tag{1}$$

$$\frac{B_p H_p}{g W_p} = f_1 \left\{ (G) \left(\omega_f \sqrt{\frac{H_p}{g}} \right) \right\} \tag{2}$$

$$t_f \sqrt{\frac{g}{H_p}} = f_1 \left\{ (G) \left(\omega_f \sqrt{\frac{H_p}{g}} \right) \right\} \tag{3}$$

$$P_r \sqrt{\frac{H_p}{g}} = f_1 \left\{ (G) \left(\omega_f \sqrt{\frac{H_p}{g}} \right) \right\} \tag{4}$$

7 Applicability

The above research shows that how the human energy can be produced and how it will helpful in application to processing of wood chipper units [10]. The human inputs such as muscular power, blood pressure, and pulse rate are the most important parameters for a general study [11]. The research is applicable to the any field where the human energy is utilized for processing. The variation in weight and height of person can give the different impacts to the process [12]. It is quite sure that the strong person with a good physique can generate the excellent result in less time. The attempt can be made to implement this study to the various field of research like mechanical units, production units, and manufacturing units as well as it can be used for medical sector, construction sector, and agricultural sector.

8 Conclusion

Thus, the dimensional equations developed in compact mode so as to form the entire experimentation method results in minimum time taking and impact in generation of best knowledge through this research. The experimental knowledge is going to be generated for invention of the mathematical model. The experimental setup will be designed which will include the power required for speeding up the flywheel, blood pressure rise, and time required flywheel rise. The indices of mathematical model are going to be developed using multivariate analysis. The analysis technique used for this work can comprise of sensitivity analysis, determination of limiting values, improvement, dependability, and AI technique are establish ANN model and to cut back error amongst experimental and mathematical knowledge. Supported the results, conclusions and hypothesis are going to be created.

References

1. McCartney, N., Heigenhauser, G.F., Jones, N.L.: Power output and fatigue of human muscle in maximal cycling exercise. J. Appl. Physiol. **55**, 218–224 (1983)
2. Sargeant, A.J., Dolan, P.: Effect of prior exercise on maximal short-term power output in humans. J. Appl. Physiol. **63**, 1475–1480 (1987)
3. Modak, J.P.: Human powered flywheel motor concept, design, dynamics and applications, pp 1–50
4. Rao, S.S.: Optimization Theory & Applications, 2nd edn. Wiley Eastern Ltd., Hoboken (1984)
5. Costill, D.L., Thomason, H., Roberts, E.: Fractional utilization of aerobic capacity during distance running. Med. Sci. Sports **5**, 248–252 (1973)
6. Hooper, D., Miller, R., Layzer, R., Giannini, D., Milner-Brown, H.S., Koretsky, A., Weiner, M.: Correlation between high energy phosphates and fatigue in human muscle. Soc. Mag. Res. Med. **1**, 481 (1985)
7. Margaria, R., Aghemo, P., Rovelli, E.: Measurement of muscular power (anaerobic) in man. J. Appl. Physiol. **21**, 1662–1664 (1966)
8. Sakhale, C.N., Bapat, P.M., Singh, M.P.: Design of experimentation and application of methodology of engineering experimentation to investigation of processing torque, energy and time required in bamboo processing operations. Int. J. Bamboo Rattan **9**(1–2), Jbr 284, 13–27 (2011)
9. Bunge, M.: A mathematical theory of dimensions and units of physical quantities. In: Bunge, M., et al. (eds.) Problems in the Foundations of Physics. Montreal. Springer, Berlin (1971)
10. Kapur, T.: Energy requirement, sources and conservation opportunities for agro-processing operations. In: Course Manual for Winter School on Entrepreneurship Development of Rural Woman through Appropriates Agro-processing Technologies, Nov. 12–Dec. 2, pp 3.1–3.6. Central Institute of Agricultural Engineering, Bhopal (2002)
11. Bergstrom, J., Hultman, E.: Muscle glycogen synthesis after exercise: an enhancing factor localized to the muscle cells in man. Nature **210**, 309–310 (1966)
12. Courtice, F.C., Douglas, C.G.: The effect of prolonged muscular exercise on the metabolism. Proc Roy Soc B, **119**, 381–439 (1936)

Impact of High-K Gate Dielectric and Work functions Variation on Electrical Characteristics of VeSFET

Gurpurneet Kaur, Sandeep Singh Gill, and Munish Rattan

Abstract In this paper, the impact of high-k dielectric gate oxide and work function variation on the electrical characteristics of the Vertical Slit Field Effect Transistor through 3D-TCAD Cogenda simulation tools has been studied. The gate work function is varied from 4.25 to 4.5 eV and high-k gate dielectric permittivity lies in the range of 10 to 35 for tied gates of device. The transistor electrical parameters such as on-current, off-current, the ratio of on-current to off-current, subthreshold swing, transconductance and transconductance generation factor are extracted from the device characteristics curve. Higher the gate work function is, the smaller is the leakage current produces at the cost of on-current. On the other side, the use of higher value of gate dielectric constant (k = 35) results in lower subthreshold swing (69 mV/dec), reduced off-current (2.26E-13), higher on-current (6.13E-5), enhanced on-current to off-current ratio (in order of 10^8), improved tranconductance and transconductance generation factor.

Keywords Subthreshold swing · VeSFET · Dielectric constant · Work function · On-current

1 Introduction

Recently, devices such as FinFET, ultra-thin body and buried-oxide fully depleted SOI (FDSOI), gate-all around nanowire MOSFET are considered as solution to lessen the short channel effects (SCEs) for sub-20 nm technology node [1, 2]. But,

G. Kaur (✉)
I. K. Gujral Punjab Technical University, Jalandhar, Punjab, India
e-mail: gurpurneetkaur@gmail.com

S. S. Gill
National Institute of Technical Training and Research, Chandigarh, Punjab, India
e-mail: ssg270870@yahoo.co.in

M. Rattan
Guru Nanak Dev Engineering College, Ludhiana, Punjab, India
e-mail: rattanmunish@gndec.ac.in

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_62

the fabrication-related issues like a gate patterning around the fin in FinFET and uniform etching of the gate beyond the channel in GAA make them less reliable for scalability perspective [3]. Therefore, IC industry is focusing on substitute device along with advantages of these devices. A novel Vertical Slit Field Effect Transistor (VeSFET) with excellent I_{ON}/I_{OFF}, small gate capacitances and greater immunity to SCEs are introduced. The circuit layout of VeSFET device is highly regular, in which all the transistors are placed on a canvas formed by a pillar array with same pillar height 'h'. Previous studies of VeSFET applications indicate that it is a prominent device for low-power circuit applications. Also, the circuit implementation of VeSFET as compared to bulk MOSFETs devices at 65 nm technology node demonstrates that it has the least intrinsic capacitance and lowest minimum energy among others [4]. Lin et al. [5] revealed that VeSFET-based circuits consume about 70% less power than the 65 nm low-power CMOS can deliver. The n-channel junctionless VeSFET device has been designed and fabricated on SOI wafer using CMOS process. The electrical characteristics of designed device show excellent results for I_{ON}, I_{ON}/I_{OFF} ratio, subthreshold slope and drain-induced Barrier lowering [2]. Twin gate n-channel and p-channel VeSFET devices designed with polysilicon gate oxide materials demonstrate outstanding electrical characteristics for a radius of 50 nm and tox = 4 nm [6]. The authors investigated n-type VeSFET device on SOI wafer with high-k dielectric at sub-7 nm node using TCAD environment. The obtained results are compared with reference FinFET in the term of I_{ON}/I_{OFF} ratio, capacitance and threshold voltage [7]. In [8], a n-type VeSFET with slit doping concentration 10^{17} atoms/cm^3 in the dimension of 50/80 nm-wide (top/bottom) and 110 nm-tall is fabricated on a boron-doped (10^{15} atoms/cm^3) SOI wafer. The results show a subthreshold swing (SS) of 65 mV/decade and on current (I_{ON}) to off current (I_{OFF}) ratio over the order of 10^8. The high-density circuits have been designed using VeSFET arrays, which result in low fabrication cost [9]. The VeSFET-based AND and OR logic circuits and SRAM memory cells have been implemented due to its additional feature of independent control of two gates [10–12]. Tunable ring oscillator and Schmitt trigger circuits have also been designed using the adjustable threshold voltage concept of VeSFET device [13].

Further, VeSFET device performances can be improved by integrating VeSFET with high-k gate dielectric materials. The semiconductor industry is focusing toward high-k gate dielectrics in transistor manufacturing processes to meet the need for higher speed transistors while keeping power consumption under control. EOT scaling below 2 nm along with the use of high-k dielectric gate oxide in place of SiO_2 results in the reduction of gate leakage current, random threshold voltage variation and SCEs [14, 15]. The downward scaling of dielectric material along with a high dielectric constant was suggested as a solution to achieve a better transistor performance [16]. The gate dielectric thickness reduction below 2 nm results in increased capacitive coupling between the gate and the substrate. A figure of merit, equivalent oxide thickness (EOT $=$ (k{SiO_2}/k{high-k})dk) is used to judge the thickness of high-k gate dielectric layer [17]. Different gate oxide dielectric materials Si_3N_4 ($K = 7.5$), Al_2O_3 ($K = 10$), $LaAlO_3$ ($K = 15$), HfO_2/ZrO_2 ($K = 25$), La_2O_3 ($K = 27$), TiO_2 ($K = 40$) have been used for reducing the direct tunneling leakage current and were found as the best alternative for SiO_2. The new materials for high-k gate dielectrics include ZrO_2, HfO_2, and their silicate alloys have good thermal stability, high crystallization temperature and sound interface qualities [18]. The other parameter which can improve the device performance is a gate work function. The performance of FDSOI MOSFET device for different metal gate work functions has been demonstrated. It has been observed that with the increase in work function, the leakage current reduces, and drive current increases. Moreover, short channel effects viz. drain-induced barrier lowering (DIBL) and subthreshold leakage current have also been decreased. The current ratio, I_{ON}/I_{OFF}, is increased to 10^4 orders, and threshold voltage (V_T) is increased linearly when metal gate work functions vary from 4.2 to 4.5 eV [19].

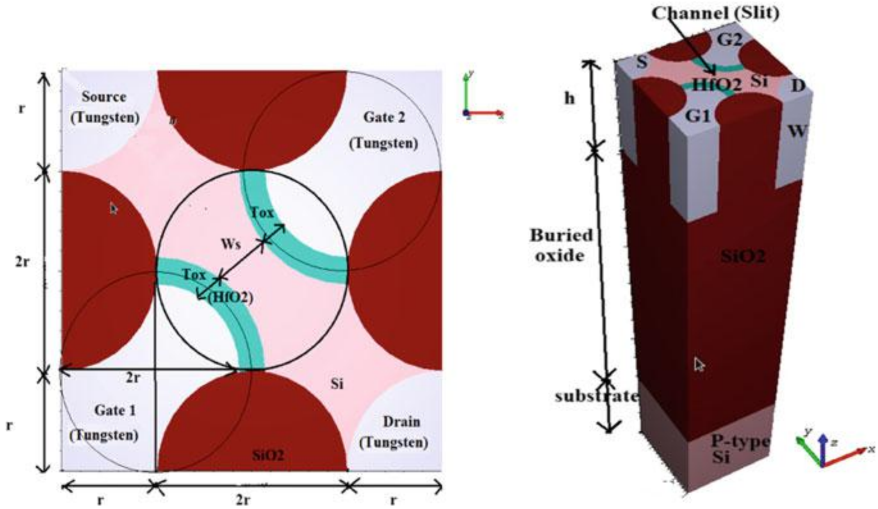
It is observed from previous research that VeSFET is a strong competitor of FinFET devices for future technologies. In this work, n-type VeSFET device has been designed with silicon and high-k gate dielectric materials on SOI wafer. The electrical characteristics of device for two channel doping levels, such as 1E16 cm^{-3} and 2E18 cm^{-3} with different high-k gate dielectric oxides and work functions, are analyzed and compared.

2 Device Design and Simulation Methodology

VeSFET is three dimensional, two gate junctionless device with one horizontal channel and four vertical terminals viz. drain, source and two independent symmetrical high-k dielectric gate oxides implemented by four cylindrical tungsten metal pillars. The structure of n-channel VeSFET is shown in Fig. 1 along with the critical dimensions of pillar radius 'r' and height 'h'. An identical circle of radius 'r' is used for making all regions of the device. All four terminals are separated by low-k isolation oxide material, i.e., SiO_2 where curved shaped high-k (HfO_2) material is placed between the gate and channel region with a thickness (T_{ox}) of 2.7 nm. The slit is narrowest channel region as mentioned in Fig. 1, has with of thickness 'Ws' can be calculated by using Eq. 1 [7].

$$W_s = 4\sqrt{2r} - 2 \times (\sqrt{2} + 1)r - T_{ox} \approx 0.828r - T_{ox} \tag{1}$$

The device is built on 150 nm thick oxide and 50 nm thick p-type silicon substrate. The source/drain doping of device is highly doped, i.e., 1E20 cm^{-3} for reducing source/drain resistance, while channel is lightly doped with two doping concentration viz. 1E16 and 2E18 cm^{-3}. The temperature and drain supply voltage used are 298 K and 0.65 V, respectively. The two independent gates G1 and G2 are tied together, and –0.2 V supply is applied to them. The equivalent gate length, contacted gate pitch (CGP), radius, height and other specifications of device are mentioned in Table 1, [7]. In n-channel VeSFET and p-channel VeSFET, the channel can be virtually undoped or moderately doped with different concentration to form junctionless device. The



(a) Top View (b) Side View

Fig. 1 VeSFET structure with critical dimensions. (**a**) Top View (**b**) Side View

Table 1 Dimensions of VeSFET device

Parameters	Lateral FinFET [7, 20]		VeSFET
	7 nm	r	r = 10 nm
Contacted gate pitch(CGP)	44–55	4r (CGP)	40
Gate length (Lg)	12–16	Eq.Lg = 1.036r, W_s = 0.828r-T_{ox}	10.36, 5.58
T_{ox}	1.5–2	2.7	2.7
Spacer	5–7	–	–
Contact	12–16	1.4r (contact diameter)	14
Fin width	6-8	–	–
Fin height	35	4r	40
Total width	75	8r	80

structure is made, and device simulations are done in the Cogenda TCAD environment. The device models for simulation of the device include drift diffusion, a quantum correction model which is the solution to Schrodinger equation, band-to-band tunneling model, lucent mobility model and carrier transport model.

3 Results and Discussion

A n-type VeSFET device with a radius of 10 nm was designed and simulated using Cogenda TCAD simulator for dimensions as shown in Table 1. The electrical characteristics are calculated and analyzed by plotting V-I characteristic curve of the device as shown in Fig. 2. The performance parameters such as I_{ON}, I_{OFF} and SS are measured for different high-k gate oxides and work functions at 0.65 drain voltage (V_d) and gate voltage (V_g). The drive current (I_{ON}) is defined as the value of current calculated at $V_d = 0.65$ V and $V_{g1} = V_{g2} = -0.2$ V. Leakage current (I_{OFF}) is referred as current value when applied input gate voltage is zero. Subthreshold swing (SS) is the inverse of subthreshold slope (S); it is defined as the ratio of change in applied gate voltage to the decade change in drive current ($SS = \frac{\partial V_{gs}}{\partial \log_{10} I_d}$) [21].

Figure 2 shows the variation of drain current (I_d) versus gate voltage (V_g) characteristics of FinFET for different values of drain voltages. It was observed that $V_d = 0.65$ V has higher on-current as compare to $V_d = 50$ mV.

3.1 Impact of High-K Gate Oxide Variation

The gate dielectric materials play very crucial role in the design of high-performance nanoscale devices. The high-k gate dielectrics are used in place of SiO_2 as they decreased threshold voltage and improved leakage characteristics of the device. The various electrical parameters for channel doping of 1E16 and 2E18 cm^{-3} are examined for different values of dielectric constants and are shown in Tables 2 and 3. Figure 3 indicates the extracted value of I_{ON} and I_{OFF} for channel doping concentration of 2E18cm^{-3} for different high-k gate dielectric materials of the device. It is observed that the device with higher dielectric constant has larger on-current and reduced off-current. The enhanced value of on-current to off-current ratio and decreased subthreshold swing is demonstrated in Table 3. The lower the SS, the higher the switching speed of digital circuits. Therefore, the device can be used for implementation in digital logics. In Tables 2 and 3, it is observed that for channel

Fig. 2 V-I characteristic curve of VeSFET

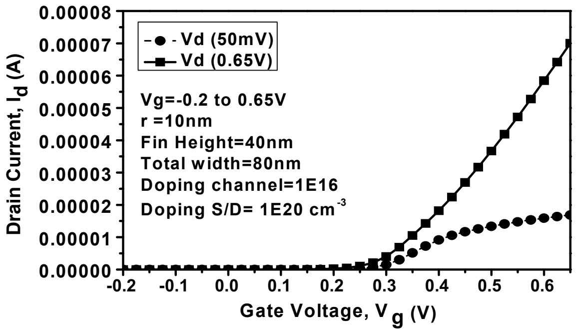


Table 2 Comparison of different dielectric constants for channel doping of 1E16 cm^{-3}

Parameters	Dielectric permittivity (k)						
	35	25	17	15	13	10	3.9
I_{ON} (A)	7.0E-5	6.5E-5	5.9E-5	5.78E-5	5.57E-5	5.19E-5	4.08E-5
I_{OFF}(A)	6.8E-13	1.83E-12	7.9E-12	1.38-11	2.77E-11	1.13E-10	5.22E-8
I_{ON}/I_{OFF}	1.02E8	3.55E7	0.746E7	0.418E7	0.201E7	0.519E6	0.781E3
SS(mV/dec)	70.8	74.8	79	80.9	81.5	90.3	147
g_m (μS) (V_d = 50 mV)	86	80.6	72.7	69.6	66.39	59.92	36.4
TGF (V^{-1}) (V_d = 50 mV)	22.2	21.8	21	20.8	20.42	19.5	15.17

Table 3 Comparison of different dielectric constants for channel doping of 2E18 cm^{-3}

Parameters	Dielectric permittivity(k)						
	35	25	17	15	13	10	3.9
I_{ON} (A)	6.13E-5	5.62E-5	5.0E-5	4.8E-5	4.57E-5	4.17E-5	1.58E-5
I_{OFF}(A)	2.26E-13	5.68E-13	2.13E-12	3.5E-12	6.54E-12	2.31E-11	5.9E-9
I_{ON}/I_{OFF}	2.71E8	0.98E8	2.34E7	1.37E7	0.69E7	0.18E7	0.267E4
SS(mV/dec)	69.8	72.2	78	81	83.5	89.2	132
g_m (μS) (V_d = 50 mV)	62.9	57.7	50.6	48.07	45.5	40.6	23.65
TGF (V^{-1}) (V_d = 50 mV)	22.26	21.78	21.08	20.8	20.44	19.64	15.3

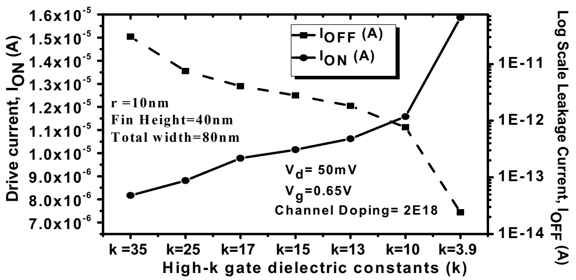


Fig. 3 Dependence of drive current and leakage current for different high-k gate dielectric constants

doping 1E16 and 2E18 cm^{-3}, drive current decreases and leakage current increases with an increase in k value. The maximum values of analog parameters viz. transconductance ($g_m = \partial I_d/\partial V_g$) and transconductance generation factor ($TGF = g_m/I_d$) are measured at k = 35 as shown in Figs. 4 and 5 [18, 21–25].

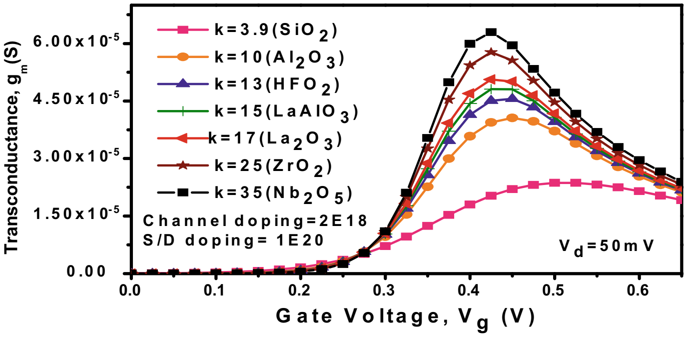


Fig. 4 Transconductance (g_m) versus gate voltage for different high-k gate dielectric constants

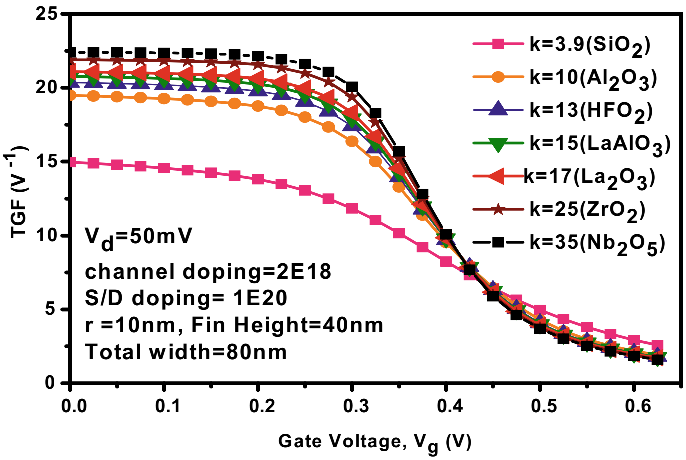


Fig. 5 Transconductance generation factor (TGF) versus gate voltage for different high-k gate dielectric constants

Tables 2 and 3 demonstrate that the drive current of high-k gate dielectric oxide (k = 35) increases by 1.75 times and 4 times as compared to SiO_2 for the channel doping of 1e16 cm^{-3} and 2E18 cm^{-3}, respectively. Also, the leakage current reduces by the order of 10^4, SS diminishes by 2 times, g_m enhances by 2.5 times, and TGF improves by 47% as compared to conventional gate oxide material for both cases. For the higher channel doping, the current ratio enhances twice, comparable SS and TGF are observed, and 67% reduction in off-current is examined at the cost of on-current and g_m for high-k gate dielectric materials as outlined in Tables 2 and 3.

3.2 Impact of Work function Variation

Work function is defined as least energy required for removing electron from a solid to a point in vacuum. The metal gate work function has been varied from 4.25 to 4.5 eV by taking their different combination for tied gates in proposed devices. It is demonstrated that the device made with work function combination of (4.35 eV, 4.35 eV) for both gates has maximum I_{ON} as compare to others, whereas for gate work function 4.5 eV for both gates has a minimum leakage current and SS because of device high threshold voltage as shown in Figs. 6 and 7. Therefore, it is concluded that device on-current is sacrificed for increased gate work function of proposed device. I_{ON}/I_{OFF} current ratio obtained from the device simulations has been found to improve significantly with the increase in metal gate work function of VeSFET. Although the device on-current reduces to some extent with an increase in gate work function, but an increase in on–off current ratio is a clear indication of overall improvement in drive current with a required low off-state leakage current for LSTP logic technology. The devices with high I_{ON}/I_{OFF} ratio have improved switching speed of the device. [19, 26].

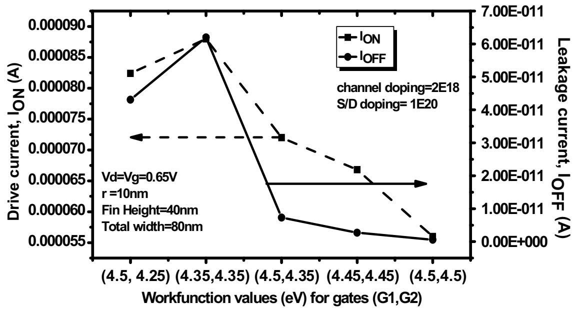


Fig. 6 Drive current and leakage current for gate work function variation from 4.25 to 4.5 eV

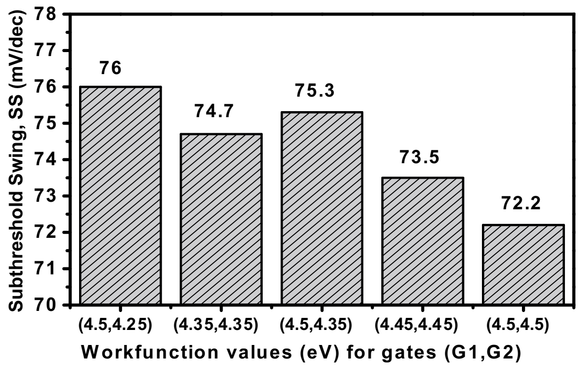


Fig. 7 Drive current and leakage current for gate work function variation from 4.25 to 4.5 eV

4 Conclusion

It is concluded that the devices made with higher value of dielectric permittivity (k) have improved I_{ON}/I_{OFF} ratio and reduced I_{OFF} at the cost of I_{ON} for channel doping 2E18 cm^{-3} as compared to 1E16 cm^{-3} in saturation region. The current ratio of VeSFET device increases in the order of 10^2 as compared to 7 nm technology FinFET device [27]. Also, SS is declined by 11% in comparison with existing work for VeSFET device integrated with the high-k gate oxide material [7]. The study also presents the effectiveness of gate work function engineering for the proposed structure. This work also explains the capability of higher gate work function materials with high-k gate dielectric materials. During simulation, it is observed that improved SS and reduced leakage current are obtained for 4.5 eV as compared to others, and such devices can be used for LSTP applications.

Acknowledgements The authors would like to thanks to Guru Nanak Dev Engineering College, Ludhiana, for providing the Research Lab facility. The authors would like to explicit their appreciation to Dean RIC, I. K. Gujral Punjab Technical University, Kapurthala. The first author would also provide thanks to Mr. Amit Saini for software support.

References

1. Colinge, J.P.: FinFETs and Other Multi-Gate Transistors. Integrated Circuits and Systems Series. Springer, New York, USA (2008)
2. Othman, N., Arshad, M.K.M., Sabki, S.N., Hashim, U.: Ultra-thin body and buried oxide (UTBB) SOI MOSFETs on suppression of short-channel effects (SCEs): a review. In: Nanotechnology in the Service of Health, Environment and Society (NANO-SCITECH 2014), pp. 257–261. Trans Tech Publications, Zürich (2015)
3. Chen, Z., Kamath, A., Singh, N., Shen, N., Li, X., Lo, G.Q., Kwong, D.L., Kasprowicz, D., Pfitzner, A., Maly, W.: N-channel junction-less vertical slit field-effect transistor (VeSFET): fabrication-based feasibility assessment. In: International Conference on Solid-State and Integrated Circuit (ICSIC 2012), pp. 1–5. IACSIT Press, Singapore (2012)
4. Qiu, X., Sadowska, M.M., Maly, W.: Vertical slit field effect transistor in ultra-low power applications. In: 13th International Symposium on Quality Electronic Design, pp. 384–391. Santa Clara, CA, USA (2012)
5. Lin, Y.W., Sadowska, M.M., Maly, W., Pfitzner, A., Kasprowicz, D.: Is there always performance overhead for regular fabric? In: IEEE International Conference on Computer Design, pp. 557–562, Lake Tahoe, CA, USA (2008)
6. Chaudhary, T., Khanna, G.: Performance analysis of vertical slit field effect transistor. Int. J. Eng. Res. Technol. (IJERT) **3**(1), 2093–2098 (2014)
7. Yang, P.L., Hook, T.B., Oldiges, P.J., Dori, B.B.: Vertical slit FET at 7-nm node and beyond. IEEE Trans. Electron Devices **63**(8), 3327–3334 (2016)
8. Maly, W.P., Singh, N., Chen, Z., Shen, N., Li, X., Pfitzner, A., Kasprowicz, D., Kuzmicz, W., Lin, Y.W., Sadowska, M.M.: Twin gate, vertical slit FET (VeSFET) for highly periodic layout and 3D integration. In: 18th International Conference on Mixed Design of Integrated Circuits and Systems, pp. 145–150, Gliwice, Poland (2011)
9. Lin, Y.W., Sadowska, M.M., Maly, W.P.: Cell layout-performance relationships in VeSFET-based, high-density regular circuits. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **30**(2), 229–241 (2011)

10. Weis, M., Pfitzner, A., Kasprowicz, D., Lin, Y.-W., Fischer, T., Emling, R., Marek-Sadowska, M., Schmitt-Landsiedel, D., Maly, W.: Low power SRAM cell using vertical slit field effect transistor (VeSFET), ESSCIRC Fringe (2008)
11. Weis, M., Emling, R., Schmitt-Landsiedel, D.: Circuit design with independent double gate transistors. Adv. Radio Sci. **7**, 231–236 (2009). https://doi.org/10.5194/ars-7-231-2009
12. Weis, M., Pfitzner, A., Kasprowicz, D., Emling, R., Maly, W., and Schmitt-Landsiedel, D.: Adder circuits with transistors using independently controlled gates In: Proc. IEEE International Symposium on Circuits and Systems ISCAS, pp. 449–452 (2009b)
13. Weis, M., Schmitt-Landsiedel, D.: Circuit design with adjustable threshold using the independently controlled double gate feature of the Vertical Slit Field Effect Transistor (VESFET). Adv. Radio Sci. **8**, 275–278 (2010). https://doi.org/10.5194/ars-8-275-2010
14. Wilk, G.D., Wallace, R.M., Anthony, J.M.: High-k gate dielectrics: current status and materials properties considerations. J. Appl. Phys. **89**, 5243–5275 (2001)
15. Nirmal, D., Nalini, B., Vijaya Kumar, P.: Nano sized high-k dielectric material for FINFET'. Integr. Ferro Electron. **121**, 31–35 (2010)
16. Bouazra, A., Abdi-Ben Nasrallah, S., Said, M., Poncet, A.: Current tunnelling in MOS devices with Al2O/SiO2 gate dielectric. Res. Lett. Phys. **2008**, 1–5 (2008)
17. Huang, A.P., Yang, Z.C., Chu, P.K.: Hafnium-based high-k gate dielectrics. In: Advances in Solid State Circuits Technologies, pp. 333–350. IntechOpen, London
18. Nirmal, D., Vijayakumar, P., Samuel, P.P.C., Jebalin, B.K., Mohankumar, N.: Subthreshold analysis of nanoscale FinFETs for ultra low power application using high-k materials. Int. J. Electron. **100**(6), 1–15, (2013)
19. Ana, F., Din, N.U.: Gate work function engineering for deep sub-micron MOSFET's: motivation, features and challenges. IJECT **2**, 30–35 (2011)
20. Doris, B. B., Hook, T.: Scalability of planar FDSOI and FinFETs and what's in store for the future beyond that? In Proceeding of ESSDERC, pp. 160–163, Graz, Austria (2015)
21. Mohapatra, S.K., Pradhan, K.P., Singh, D., Sahu, P.K.: The role of geometry parameters and Fin aspect ratio of sub-20 nm SOI-FinFET: an analysis towards analog and RF circuit design. IEEE Trans. Nanotech. **14**(3), 546–554 (2015)
22. Nayak, K., Bajaj, M., Konar, A., Oldiges, P.J., Natori, K., Iwai, H., Murali, K.V., Rao, V.R.: CMOS logic device and circuit performance of Si gate all around nanowire MOSFET. IEEE Trans. Electron Devices **61**(9), 3066–3074 (2014)
23. Mohsenifar, S., Shahrokhabadi, M.H.: Gate stack high-k materials for Si-based MOSFETs past, present, and futures, Microelectron. Solid State Electron. **4**(1), 12–24 (2015)
24. Song, J., Han, C., Lai, P.T.: Comparative study of Nb_2O_5, NbLaO, and La_2O_3 as gate dielectric of InGaZnO thin-film transistor. IEEE Trans. Electron Devices **63**(5), 1928–1933 (2016)
25. Bahari, A., Gholipur, R., Mianaee1, Z.K., Mousavi, S.H.: The effect of Zirconium incorporation in La_2O_3 nanocrystallites as gate dielectric in MOSFET. J. Basic Appl. Sci. Res. **2**(4), 3470–3475 (2012)
26. Mustafa, M., Bhat, T.A., Beigh, M.R.: Threshold voltage sensitivity to metal gate work-function based performance evaluation of double-gate n-FinFET structures for LSTP technology. World J. Nano Sci. Eng. **3**, 17–22 (2013)
27. Sicard, E.: Introducing 7-nm FinFET technology in Microwind. https://hal.archives-ouvertes.fr/hal-01558775. Accessed 23 Sep. 2019

Correlating Personality Traits to Different Aspects of Facebook Usage

Maheshkumar B. Landge, Deepali Mahajan, and C. Namrata Mahender

Abstract Nowadays Internet is widely used and social networking sites has become the important part of human's life. The impact of such influence is needed to be studied as the physical world is becoming reality and real-world hell or less important. Social networking site is not only for communication but games, videos, status and many more. The use of social networking sites is diverse, and it can be from daily shopping to chitchatting, to education, learning and even business. This diversity attracts all stock holders making it more powerful, and at the same time, nonproper scheduling is causing the new term Internet addiction. There are many numbers of social networking sites with various different specialties they offer, and each has its own remarkable story, for example, YouTube for video sharing, Twitter for text messaging or tagging generally termed as tweets, Skype for video chatting and many more. The most popular of them is Facebook; according to 2018 July report, 2196 million users across the world. The main aim of this paper is to create awareness about the addiction and then provide the aspect of personality traits. In this paper, Facebook is considered to showcase the necessary features which should be properly surveyed like Poke, Tag, Like, fear of missing out, etc., so that its relationship can be derived to analyze the 'big five' personality trait is openness, conscientiousness, extraversion, agreeableness and neuroticism. We designed Facebook questionnaire of 36 questions, from which we considered four questions for analysis. On the basis of option frequencies of these questions, we can analyze daily Facebook usage time, Facebook use on holiday and in exam time.

Keywords Personality traits · Social networking sites · Facebook features

---

M. B. Landge (✉) · D. Mahajan · C. Namrata Mahender
Department of CS and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MS, India
e-mail: maheshkumar.landge@gmail.com

D. Mahajan
e-mail: deepali.borole@gmail.com

C. Namrata Mahender
e-mail: nam.mah@gmail.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_63

1 Introduction

The world seems too small, the cultural barriers are turning to melodious songs, everything is online, nothing is left unseen, associations amongst strangers are tight, and we just have anything to fight. The days and nights have no importance as communication is anytime anywhere. The more savvy the more the thrill, the world is felt on fingertip. The computer is not limited to learning or telecommunication. For just communication, they have become media for education, entertainment learning, socializing, business and many more. Recently, social networking sites (SNSs) like Facebook, MySpace, Twitter, LinkedIn, etc., are channels to explore as per individual requirement and interest. According to Kaplan and Haenlein, SNSs are a group of Internet-based apps that build on Web 2.0's ideological and technological foundations and enable content produced by users to be created and exchanged [1].

The statistics portal report of all social networking sites for year 2019 is shown in below table [2]

Table 1 shows that Facebook is the topmost according to number of users of Facebook. The history of social networking sites (SNSs) dates back to 1997 when the first SNS six degrees emerged as a result of the idea that individuals are linked via six degrees of separation [3, 4] and are conceived as "the small world problem," in which society is viewed as becoming increasingly interconnected [5]. In 2004, Facebook was launched as an online community for students of Harvard University and has since become the world's most popular SNS [6].

Similarly if we want to see both sides of the coin i.e. positive and negative effect of SNSs. There are some positive effects of SNSs like it can be helpful in increasing creativity, interaction, learning, easiest way of sharing information and also some negative effects SNSs like cyber bullying, depression, loneliness, anxiety, negative feeling [7]. Many researchers are trying to measure "Big five" personality traits by surveying on the usage of social media. This paper is only focusing on Facebook and also the features of it, which should be incorporated in any survey related to SNS.

2 Personality Traits

Personality traits represent the distinctive patterns of ideas, emotions and behaviors of individuals [8]. Personality traits reflect basic dimensions on which people differ [9]. A significant characteristic of personality traits is that they reflect continuous distributions rather than separate kinds of personality.

Table 1 The statistical portal report of social networking sites use for year 2019

Sr. no.	Name of social networking site and established year	Monthly active users	Daily active users	Daily usage	Daily time spent
1	Facebook (Est. 2004)	2.4 Billion	1.6 Billion	300 Million photos uploaded everyday	58 min
2	YouTube (Est. 2005)	1.9 Billion	149 Million	5 Billion video views daily	40 min
3	WhatsApp (EST. 2009)	1.5 Billion	1 Billion	60 Billion messages daily	—
4	Instagram (Est. 2010)	1 Billion	600 Million	95 Million posts daily	—
5	Twitter (Est. 2006)	330 Million	134 Million	140 Million daily tweets	—
6	Reddit (Est. 2005)	330 Million	14 Billion	25 Million votes daily	13 min
7	LinkedIn (Est. 2002)	303 Million	5.3 Million	30 Million company pages	10 min
8	Snapchat (Est. 2011)	301 Million	109 Million	10 Billion daily video views	20 min (25 year and above), 30 min (below 25 year)
9	Pinterest (EST. 2010)	291 Million	1 Billion	175 Billion pins create	14 min

2.1 Brief on Big Five Personality Traits

The big five personality traits are also termed as OCEAN, O—Openness, C—Conscientiousness, E—Extraversion, A—Agreeableness and N—Neuroticism [10].

(i) Openness:

This trait features characteristics such as imagination and insight. People who are high in this trait also tend to have a broad range of interests. They are curious about the world and other people and eager to learn new things and enjoy new experiences.

(ii) Conscientiousness:

Standard features of this dimension include high levels of thoughtfulness, good impulse control and goal-directed behaviors. Highly conscientious people tend to be organized and mindful of details. They plan ahead, think about how their behavior affects others and are mindful of deadlines.

(iii) Extraversion:

Extraversion is characterized by excitability, sociability, talkativeness, assertiveness and high amounts of emotional expressiveness. People who are high in extraversion are outgoing and tend to gain energy in social situations. Being around other people helps them to feel energized and excited.

(iv) Agreeableness:

This personality dimension includes attributes such as trust, altruism, kindness, affection and other prosocial behaviors. People who are high in agreeableness tend to be more cooperative, while those low in this trait tend to be more competitive and sometimes even manipulative.

(v) Neuroticism:

Neuroticism is a trait characterized by sadness, moodiness and emotional instability. Individuals who are high in this trait tend to experience mood swings, anxiety, irritability and sadness. Those low in this trait tend to be more stable and emotionally resilient.

3 Facebook Features

We selected some Facebook features as given below:

(i) Tagged

"Tagging" is a social feature that started on Facebook that involves attaching a friend's name to a social media post or image online.

In the beginning, Facebook tagging could only be done with photos. Today, we can incorporate tagging into any type of Facebook post at all. Tagging basically involves attaching a friend's name to one of your posts. This made a lot of sense back when it was exclusively meant for photos because anyone who uploaded photos could tag their friends who appeared in them to put a name to each face [11].

(ii) Poked

A poke is a form of electronic nudge or touch that basically says "Hello!" However, a poke can also be used to flirt, annoy or check in with someone. Its meaning is in the mind of the "poker" and how the "pokee" views it [12].

(iii) Fear of missing out

Fear of missing out (FOMO) defines when a individual feels enhanced anxiety or stress because he or she misses social activities or is not invited to participate first [13].

(iv) Timeline

Every user shares pictures, messages and experiences on Facebook using Facebook Timeline. You can add a cover picture to your Timeline, edit your private data, view your Facebook activity log, highlight posts or pictures, update your Facebook status, share your Facebook app activity and add fresh activities to your profile [14].

4 Database

We have developed questionnaire for aspect checking on Facebook. In that questionnaire, 36 questions are used. We can understand impact of Facebook on individual's life through this questionnaire. We have collected database from 661 users, and there are 350 male and 311 female. The age group of users is 14–16. This is sample questionnaire for Facebook as shown in Fig. 1.

5 Identical Features of Facebook and Mapping to Personality Trait

Generally, when we take a walkthrough of Facebook, we find friend list, find friend, friend request, messages, groups, notification, timeline, post wall, poke, tags, etc. The most popular SNSs addiction test is Internet addiction test, but as every SNS have their own specific features, it is not possible to see the effect of it on the survey to analyze the personality trait. Figure 2 shows few features and approximate map with personality trait.

Rather than having questions only on:

- How many friends?
- How many physically you know?
- Number of groups joined?
- Time spent

The main focus showed be does something affects them positively or negatively

- Tags:

− How many tags?
− Do you like tagged?

Computational and Psycholinguistic Research Lab, Department of CS & IT, Dr. B. A. M. University Aurangabad

Name of School : ______________________ **Class:** __________,

Student Name:______________________ **Age:**____ **Gender: Male/Female**

Questionnaire for aspect checking on Facebook

1. Do you use Facebook?
   a. Yes b. No
2. Do you have your own login?
   a. Yes b. No
3. Are your parents in your friends list?
   a. Mother: Yes/No b. Father: Yes/No
4. Are your siblings in your friend list?
   a. Brother: Yes/No b. Sister: Yes/No
5. How much time you spend on Facebook?
   a. 0-2 hour's b. 2-5 hours c. More than 5 hours
6. What is daily Facebook using time in general?
   a. Early Morning b. Afternoon c. Evening d. Late Night
7. Do you enjoy while you are on Facebook?
   a. Yes b. No
8. How much time on holiday you spend on Facebook?
   a. 0-3 hour's b. 3-6 hours c. More than 6 hours
9. Are your teachers in your friend list?
   a. Yes b. No
10. Do you feel angry when you are not allowed to use Facebook?
    a. Never b. Rarely c. Sometimes d. Often e. Frequently
11. How many friends you have on Facebook?
    a. 0-100 b. 100-500 c. More than 500
12. What activities you do most on Facebook?
    a. Chatting b. Browsing through your homepage c. Posting articles/emotions/photos
13. Are you always excited to post on Facebook photos or any other material?
    a. Never b. Rarely c. Sometimes d. Often e. Frequently
14. Do you frequently check your post for friend's remark?
    a. Never b. Rarely c. Sometimes d. Often e. Frequently
15. If negative remarks are there do you feel,
    a. Sad b. Angry c. Both sad and angry
16. How much time do you spend on Facebook during examination?
    a. 0-1 hour's b. 1-2 hours c. More than 2 hours
17. If free time is there what will you prefer?
    a. Using Facebook c. Talking to friends/relatives on phone
    b. Watch movie d. Go to sleep
18. Do you share every detail of success on Facebook?
    a. Never b. Rarely c. Sometimes d. Often e. Frequently
19. Do you share every detail of failure on Facebook?
    a. Never b. Rarely c. Sometimes d. Often e. Frequently
20. Do share every detail of daily routine on Facebook?
    a. Never b. Rarely c. Sometimes d. Often e. Frequently

1

Acknowledge to CSRI DST Major Project NO. SR/CSRI/71/2015(G)

Fig. 1 Sample of questionnaire for aspect checking on Facebook

- Pokes:

− Do you get poked?
− Why are you poked?
− Do you get disturbed by pokes?

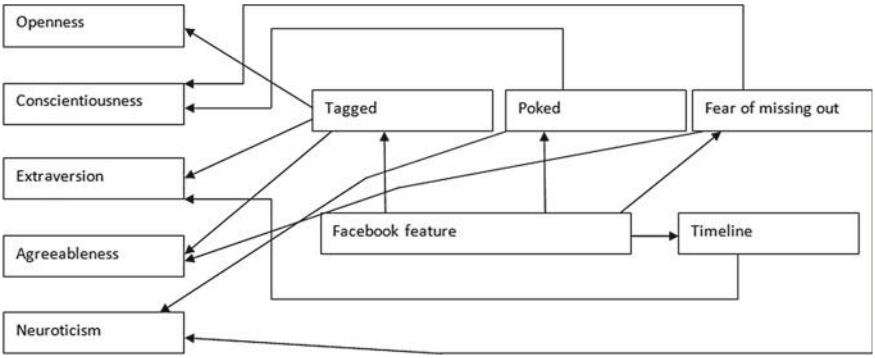


Fig. 2 Features and approximate map with personality trait

The answers will explain use the nature of impact like "I am happy when friends tag me" or "I get irritated due to poke." Such questions when added to survey will substantially increase the reach and able to get issues for analyzing and mapping to personality traits.

6 Result

We have selected Question 5, 6, 8 and 16 from database questionnaire for data analysis, which contains information related to Facebook usage time, daily Facebook using time, holiday Facebook usage time, Facebook usage in exam time.

Table 2 shows result of Question 5 for Facebook use for male and female.

Table 3 shows result of Question 6 for Facebook use for male and female.

Table 2 Result of Question 5 for Facebook use

Option	Male frequency	Female frequency
1	190	153
2	135	138
3	13	15
4	11	5

Table 3 Result of Question 6 for Facebook use

Option	Male frequency	Female frequency
1	204	150
2	28	19
3	32	54
4	47	64

Table 4 Result of Question 8 for Facebook use

Answer	Male frequency	Female frequency
1	200	150
2	125	138
3	13	18
4	11	5

Table 5 Result of Question 16 for Facebook use

Option	Male frequency	Female frequency
1	218	171
2	107	127
3	16	9
4	8	4

Table 4 shows result of Q 8 for male and female for Facebook use.
Table 5 shows result of Question 16 for Facebook use for male and female.

7 Analysis

Tables 2, 3, 4 and 5 show result for Question 5, 6, 8 and 16 for male and female for age group 14–16. Figures 3, 4, 5 and 6 show plots of Question 5, 6, 8 and 16 results. Fig. 3 shows frequency of male is more for spending time on Facebook as compared to female. Both male and female students spend time on Facebook up to

Fig. 3 Plot of Question 5 result

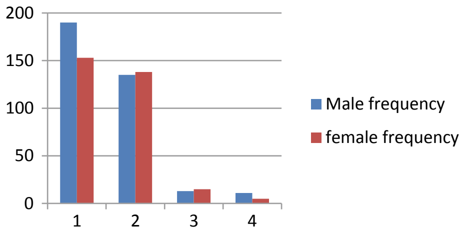
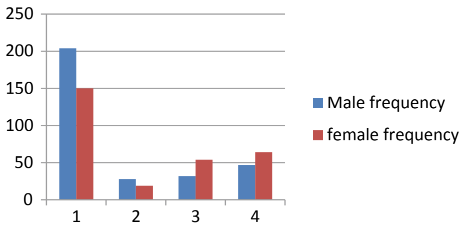


Fig. 4 Plot of Question 6 result



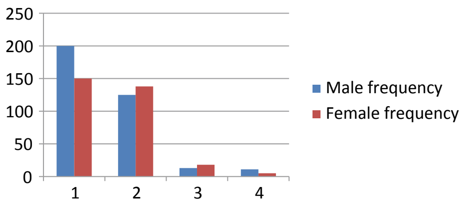


Fig. 5 Plot of Question 8 result

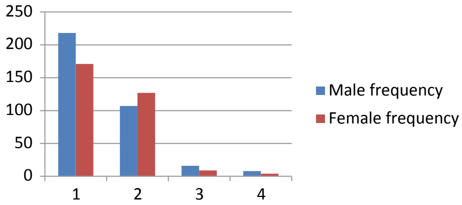


Fig. 6 Plot of Question 16 result

2 h. Fig. 4 shows both male and female students use Facebook more in early morning time. Frequency of male students is more than female students. Fig. 5 shows both male and female students use Facebook less than 3 h on holidays. Facebook usage on holiday ratio of male students is higher than female students. Fig. 6 shows both male and female students use Facebook in exam time less than 1 h. The ratio of Facebook usage in exam time is more for male students than female students.

8 Conclusion

Social networking sites are increasing day by day, but the main concern is that we all of any age groups, country, religion, profession are getting trapped in SNSs and moving away from the real life which makes things more complicated. Mainly, it disturbs the mental health and many times even physical health. So to study and create awareness is very important. But in India, not much efforts can be seen in this direction compared to countries like UK, Japan, China, USA. Even modern analysis is focusing on very general features, and it should actually have to impart on all the relative aspects of SNSs. This paper provides such perspective through Facebook and its relative mapping need for understanding personality traits. This paper provides the view of how and why it is important, which is theoretically mapped, and we have created database questionnaire for to study the Facebook aspect. We have selected few questions for data analysis. From analysis of selected questions, we come to know some conclusions like both male and female students spend time on Facebook up to 2 h daily, both male and female students use Facebook more in early morning time, both male and female students use Facebook less than 3 h on holidays, and both male and female students use Facebook in exam time less than 1 h. In all analysis of questions, frequency of male students is higher than female students.

Acknowledgements Authors would like to acknowledge and thanks to CSRI DST Major Project sanctioned No. SR/CSRI/71/2015 (G), Computational and Psycholinguistic Research Lab Facility supporting to this work and Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India.

References

1. Kaplan, A.M., et al.: Users of the world, unite! The challenges and opportunities of social media. Bus. Horiz. **53**(1), 59–68 (2010)
2. Stout, D.W: Social Media Statistics 2019: Top networks by the numbers. https://dustinstout.com/social-media-statistics/. Accessed 22 Sep. 2019
3. Dhokrat, A., Gite Hanumant, R., Namrata Mahender, C.: Add-on group system for filtered social networking by age, ISCA—100, Calcutta (2012)
4. Boyd, D.M., et.al.: Social network sites: definition, history, and scholarship. J. Comput. Mediat. Commun. **13**(1), 210–230 (2007)
5. Milgram, S.: The small world problem. Psychol. Today **2**, 60–67 (1967)
6. Kuss, D.J., et al.: Online social networking and addiction—a review of the psychological literature. Int. J. Environ. Res. Public Health **8**(9), 3528–3552 (2011)
7. UK Essays: Positive and negative effects of social media. https://www.ukessays.com/essays/education/positive-effects-of-social-networking.php?vref=1. Accessed 2 April 2020.
8. Diener, E., Lucas, R.E.: Personality traits. https://nobaproject.com/modules/personality-traits. Accessed 17/9/2019
9. Sanford, R.N., et al.: The measurement of implicit antidemocratic trends. The authoritarian personality, pp. 222–279 (1950)
10. Cherry, K.: The-big-five-personality-dimensions. https://www.verywellmind.com/the-big-five-personality-dimensions-2795422. Accessed 18 Sep. 2019
11. Moreau, E.: What is tagging on Facebook. https://www.lifewire.com/what-is-tagging-on-facebook-3486369. Accessed 17 Sep. 2019
12. Price, R.: What does it mean to poke someone on Facebook? https://www.techwalla.com/articles/what-does-it-mean-to-poke-someone-on-facebook. Accessed 19 Sep. 2019
13. Stroud, F.: Fear of missing out. https://www.webopedia.com/TERM/F/fomo.html. Accessed 21 Sep. 2019
14. Beal, V.: Facebook timeline. https://www.webopedia.com/TERM/F/facebook_timeline.html. Accessed 24 Sep. 2019

Fractional-Order Control of a Fuel Cell–Boost Converter System

Arezki Fekik, Ahmad Taher Azar, Nashwa Ahmad Kamal, Hakim Denoun, Khaled Mohamad Almustafa, Lamine Hamida, and Mustapha Zaouia

Abstract This paper processes the output voltage control system for energy source based on a fuel cell linked to a DC voltage generator. The suggested technique is based on the design of a fractional controller to regulate the output voltage of this DC–DC converter for distinct resistive load levels. Closed-loop system efficiency is assessed using simulation results. A regulation of the suitable output voltage and a solid conduct regarding the load variation and the DC output voltage are provided.

Keywords DC–DC boost converter · Fuel cells · Fractional-order control · Power load

---

A. T. Azar (✉)
Robotics and Internet-of-Things Lab (RIOTU), Prince Sultan University, Riyadh, Saudi Arabia
e-mail: ahmad_T_azar@ieee.org; aazar@psu.edu.sa; ahmad.azar@fci.bu.edu.eg

Faculty of computers and Artificial Intelligence, Benha University, Benha, Egypt

N. A. Kamal
Faculty of Engineering,
Cairo University, Giza, Egypt
e-mail: nashwa.ahmad.kamal@gmail.com

K. M. Almustafa
Prince Sultan University, Riyadh, Saudi Arabia
e-mail: kalmustafa@psu.edu.sa

A. Fekik · H. Denoun · L. Hamida · M. Zaouia
Electrical Engineering Advanced Technology Laboratory (LATAGE),
University Mouloud Mammeri of Tizi-Ouzou, Tizi Ouzou, Algeria
e-mail: a.fekik@univ-bouira.dz; arezkitdk@yahoo.fr

H. Denoun
e-mail: akim_danoun2002dz@yahoo.fr

L. Hamida
e-mail: l_hamida@yahoo.com

M. Zaouia
e-mail: zbmust@yahoo.fr

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_64

1 Introduction

The restricted conventional energy sources as well as the increasing energy demands highlight the role of renewable energy sources [1, 3, 17–19]. The fuel cell energy source could be used either as a stand-alone energy source or as a network energy source [9, 27, 30]. The objective in all instances is to generate a stable DC output voltage independent of the load current. The fuel cell does not generate acceptable output voltages [28]. For this reason, a DC–DC boost converter must be used to achieve the required voltage output value [15]. The fuel cell's output voltage relies on several parameters such as hydrogen and oxygen pressures, cell membrane temperature and moisture, and output current [2]. Furthermore, the fuel cell's control issue could be reduced by regulating the pressure and the flow of hydrogen to the required output voltage and compensating the impact of the charging current. A reasonable control of output voltage is not a simple task because the two systems are extremely nonlinear. In the field of control engineering, fractional-order controls were motivated by the ambition to improve the performance and robustness of control systems based on the advantageous properties of fractional-order systems [4, 5]. Most research analyzes the generalization of the design of classical controls. Fractional $PI^{\lambda}D^{\mu}$ control [22] and fractional adaptive control [13] are among the popular examples of fractional-order regulators. Fractional-order systems could model more real materials more appropriately than integer ones, providing an excellent modeling tool for describing many real dynamic processes [13, 23, 33, 34]. The subject of fractional controllers has attracted the attention of the control engineering research community for its high application potential [14, 24, 32]. Many fractional control structures based on the model predictive control (MPC) configuration have been developed in the literature [21, 26]. For these split-order systems, the control engineer must approximate the order division model with a rational transfer function using some approximation techniques.

This paper would address the voltage of the fuel cell-based power supply control system connected to the DC voltage generator. The proposed method is based on the design of a fractional regulator for the regulation of the output voltage of this DC–DC converter of the level of the resistive load. The efficiency of the closed-loop system is assessed using simulation results. Appropriate output voltage control with high tolerance for load variability and DC output voltage is given.

The paper is divided into six sections. Sections 2 and 3 deal with the modeling of fuel cells and DC–DC converters. Sections 4 and 5 describe the design of the control system and the simulation results. The conclusion is presented in Sect. 6.

2 Modeling of the Fuel Cell Stack

The proton exchange membrane fuel cell (PEMFC) basic system (shown in Fig. 1) can be demonstrated as two electrodes (anode and cathode) separated by a strong membrane acting as an electrolyte [31]. Hydrogen fuel flows via a network of anodes,

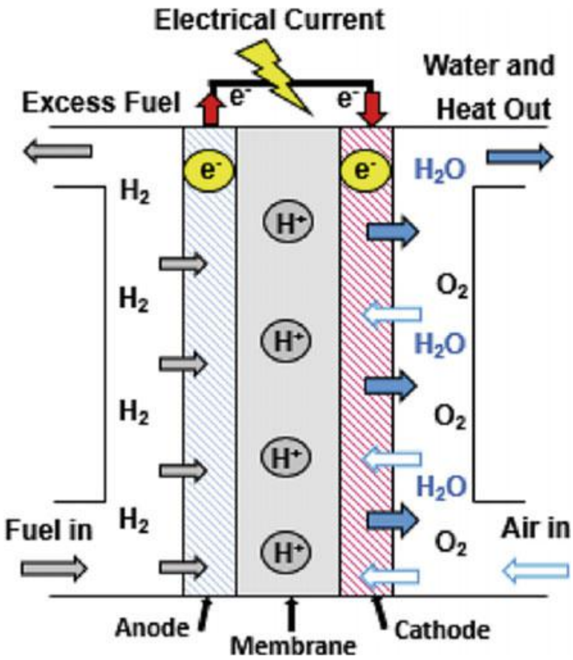


Fig. 1 Operating principle of PEMFC

dissociating itself into protons which in turn flow from the membrane to the cathode and electrons gleaned by an internal circuit which connects the two electrodes as an electrical current. The oxidant (air in this research) flows through a comparable network of channels to the cathode where oxygen mixes with the electrons in the internal circuit and the protons flowing through the membrane to produce water. [10–12]. The electrochemical reactions at the anode and cathode in a PEMFC are given below [16]:

Anode reaction

$$H_2 \Leftrightarrow 2H^+ + 2e^- \tag{1}$$

Cathode reaction

$$O_2 + 2H^+ + 2e^- \Leftrightarrow H_2O \tag{2}$$

General cell reaction

$$H_2 + \frac{1}{2}O_2 \Leftrightarrow H_2O \tag{3}$$

The mathematical model utilizes a set of parameters whose definition is crucial to obtain the best simulation outcomes. It is preferable to set the output voltage of a single cell at this stage which can be described as [6, 7]:

$$V_{cell} = E - \eta_{act} - \eta_{ohm} - \eta_{dif} \tag{4}$$

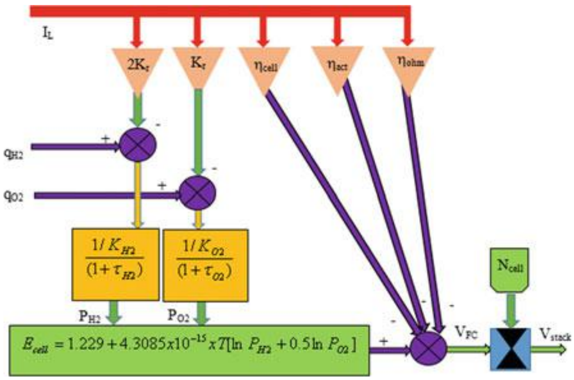


Fig. 2 Schematic model of PEMFC system

For N cells put in series, creating a stack, the voltage, V_{stack}, can be determined by:

$$V_{stack} = N_{cell} * V_{cell} \tag{5}$$

The electrical power produced by the cell to the load can be provided by:

$$P_{FC} = V_{FC} * I \tag{6}$$

where V_{Fc} is cell voltage for all operating requirements using the PEMFC equations given in the [8], and the PEMFC system can be modeled as shown in Fig. 2.

3 Modeling of the DC–DC Converter

Booster converters are typically boost power converters that have a low voltage input and an output at a higher voltage. An ideal DC–DC over-current converter circuit is shown in Fig. 3. The mathematical relationship between the input and output voltages is regulated by the duty cycle (d) as given by the equation below [8, 20]:

$$V_O = \left(\frac{1}{1-d}\right) V_{in} \tag{7}$$

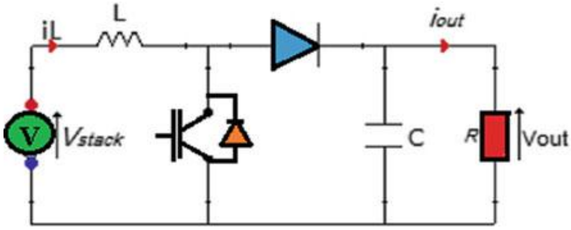
with $V_{in} = V_{stack}$

The voltage boost converter given in Fig. 3 includes an inducer, a transistor, a diode, a filter condenser and a resistor. It is driven by a generator of DC voltage.

The differential equations characterizing the current dynamic state of the inductor i_L and the dynamic state of the output voltage V_{out} are obtained when the diode is off and the transistor is activated, as follows:

$$\frac{di_L}{dt} = \frac{1}{L} V_{stack}, \quad \frac{dV_{out}}{dt} = -\frac{1}{C} i_{out} \tag{8}$$

Fig. 3 Stator flux with $\omega_{ref} = cst$



The differential equations that characterize the current dynamic state of the inductor i_L and the dynamic state of the output voltage V_{out} are obtained when the transistor is off and the diode is on, as follows:

$$\frac{di_L}{dt} = \frac{1}{L}(V_{stack} - V_{out}), \quad \frac{dV_{out}}{dt} = \frac{1}{C}(i_l - i_{out}) \tag{9}$$

4 Fractional-Order Proportional Integral (PI) Controller Design of Fuel Cell

4.1 Fractional-Order PID Controller

Recently, fractional-order controls are widely used in many papers with respect to these advantages, namely the robust performance of a proportional–integral–derivative (PID) controller. The fractional PID controller is used by [22] for a fractional control system. Afterward, several researchers in the field of engineering have designed fractional PID controller based on different designs and methods of tuning [25]. Fractional-order controls consider conventional PID extensions and are less sensitive to changes in the controlled system and controller parameters. Unlike the traditional PID controller, the fractional PID controller has five parameters to adjust, which provides two additional degrees of freedom over the conventional PID controller. This fractional PID controller provides the generalized transfer function as follows:

$$G_c(s) = K_p + \frac{K_I}{S^\lambda} + S^\mu \tag{10}$$

where λ and μ are order of fractional-order integration and order of fractional-order differentiation, respectively. In view of this, input to the system can be controlled by taking the following form of equation:

$$u(t) = K_p * e(t) + K_I * j^{-\lambda} * e(t) + K_D * D^\mu * e(t) \tag{11}$$

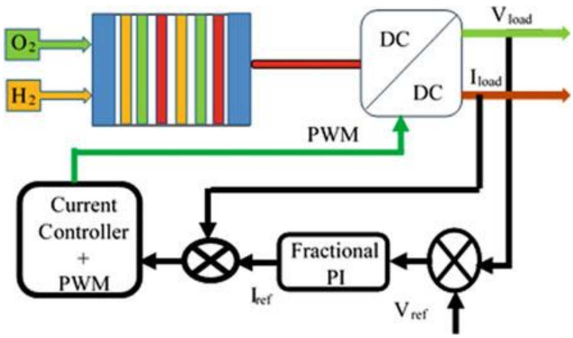


Fig. 4 Voltage load trajectory with $R_{load} = cst$

4.2 Fractional-Order PI Controller

In this portion, the fractional-order PI (FO-PI) controller is used to stabilize and enhance efficiency such as the PEMFC system output voltage reaction. The regulation on the closed loop is demonstrated in Fig. 4.

In FO-PI controller, I operator is usually of fractional orders. So by controlling the fractional integrator operator, one more freedom degree will be added to K_p and K_I variables. This variable, λ, is the fractional integrator. This additional degree of freedom makes FO-PI controller have a better performance than integer-order PI controller. The control signal of FO-PI in the time domain can be written as [13]:

$$u(t) = K_p * e(t) + K_I * D^{-\lambda} * e(t) \tag{12}$$

with: $e(t) = V_{ref} - V_{load}$
where D is differential operator. The transfer function is displayed as:

$$C(s) = K_p + \frac{K_I}{S^{\lambda}} \tag{13}$$

The current reference I_{ref} is produced based on a fractional action of PI on the output voltage V_{load} and its reference V_{ref} as shown in Fig. 4.

5 Simulation Results and Discussion

This part displays the results of output voltage regulation of PEMFC using classical and fractional-order PI controller. Simulations are done in the MATLAB/Simulink, and the fractional implementation was done by FOMCON toolbox [29]. Parameters of the proposed system are summarized in Table 1.

Table 1 Parameters of the proposed system

Parameter	Value
Voltage reference V_{ref}	48 V
Resistance load R	10 Ω
Inductance L	4.5e-6 H
Capacitance C	90e-6 F
PEMFC model	1.26 KW-24 Vdc

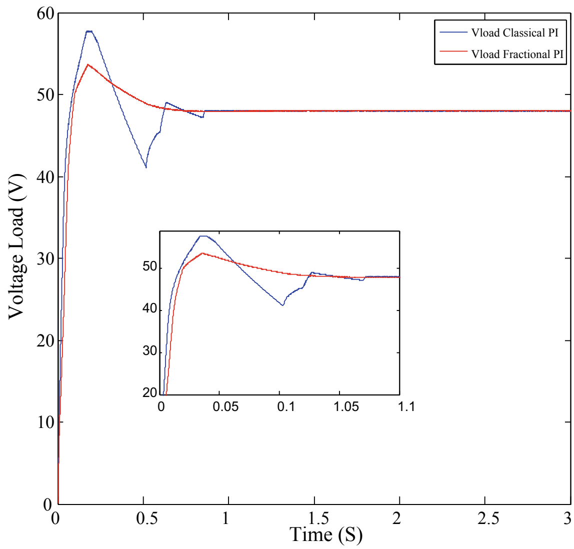


Fig. 5 Voltage load trajectory with $R_{load} = cst$

The performance testing scenarios are simulated as follows.
Simulation Test 1: Static load as R_{load} = 10Ω, and the second test: variation of load as R_{load} = 10 and 5 and 3.33Ω.

Figure 5 shows the voltage load trajectory with $R_{load} = 10\Omega$. From this figure, the time response using the proposed fractional-order PI controller is robust and better than the classical-order PI controller as shown at the zoom.

Figures 6 and 7 show the current load trajectory and power load trajectory with $R_{load} = 10\Omega$, respectively. It can be noted that the overshoot is considerably eliminated in the proposed fractional PI controller than the classical controller for the current and the power load.

The second test: variation of load as $R_{load} = 10$, 5 and 3.33Ω. Figure 8 shows the voltage load trajectory with variation of R_{load} at different levels. It can be noted that the response of the measured voltage follows its references for the different load levels with small overshoots for the two controllers. However, the robustness is improved when the fractional-order PI controller is used.

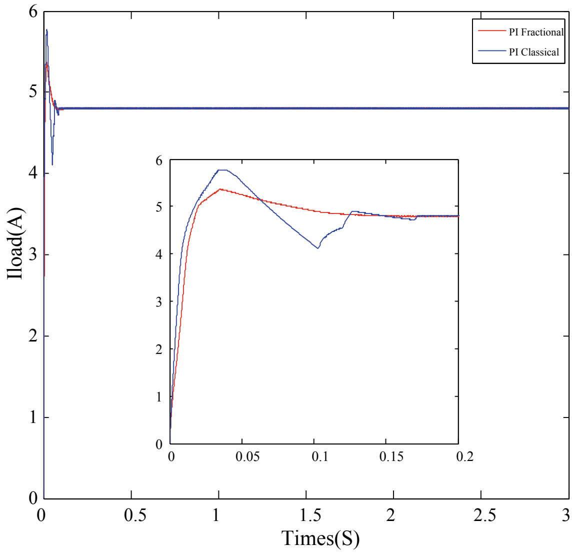


Fig. 6 Current load trajectory with $R_{load} = cst$

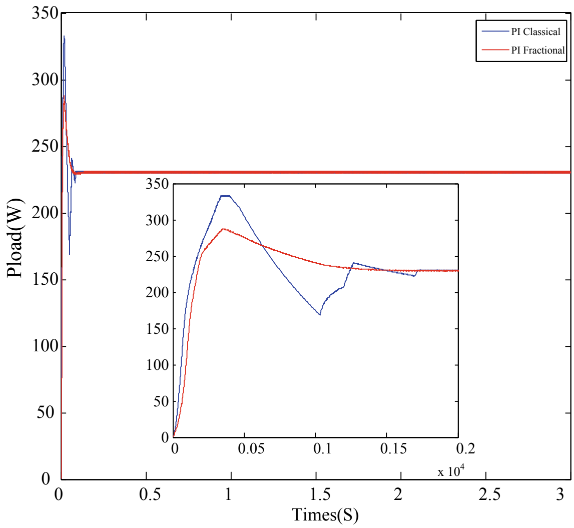


Fig. 7 Power load trajectory with $R_{load} = cst$

Figures 9 and 10 show the current load trajectory and the power load, respectively, with variation of R_{load} at different levels. It can be seen that the load power increases as well as the current during the load variation for both controllers. Robustness is improved when using the fractional-order PI controller compared to the conventional PI controller.

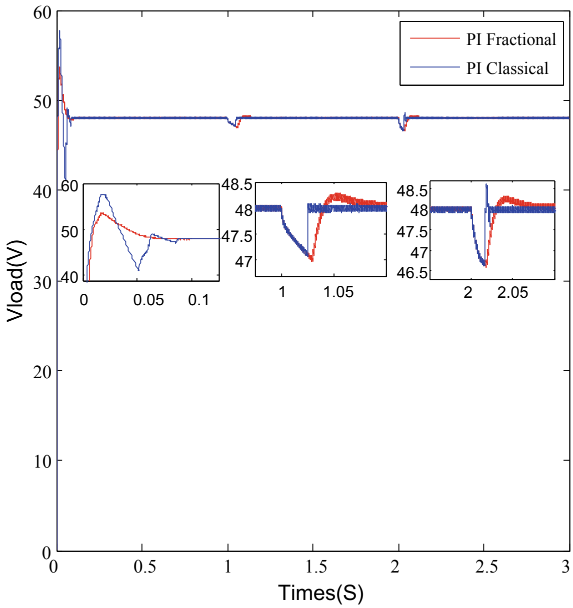


Fig. 8 Voltage load trajectory with variation of R_{load}

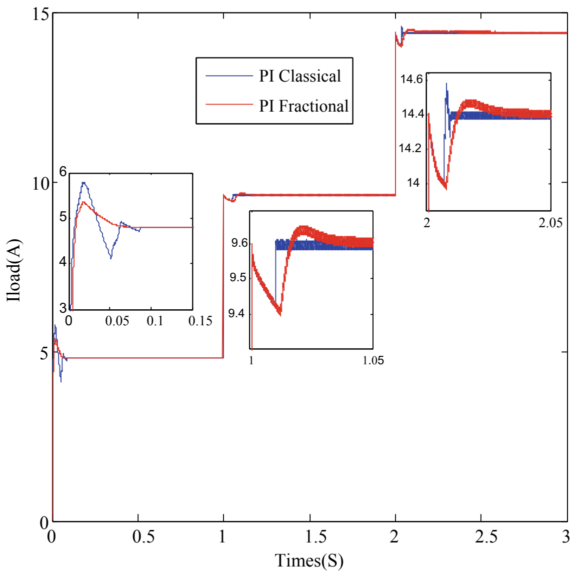


Fig. 9 Current load trajectory with variation of R_{load}

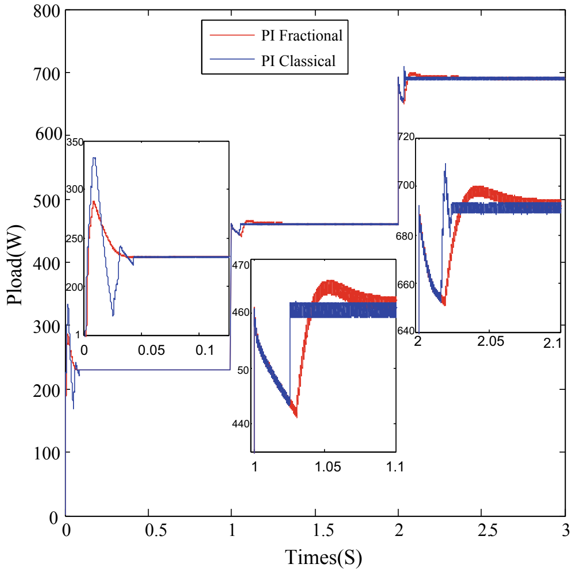


Fig. 10 Power load trajectory with variation of R_{load}

6 Conclusion

A fractional-order PI controller system for regulating the output voltage of a fuel cell–boost converter scheme is shown and presented in this work. A nonlinear fractional-order PI controller was obtained providing stability and dynamics for regulation. Accurate control, robust behavior against high unknown load variations and fuel cell and battery tension have been demonstrated. For future work, the control approach can be extended to have an adaptive controller based on a fuzzy logic.

Acknowledgements The authors would like to thank Prince Sultan University, Riyadh, Saudi Arabia, for supporting this work.

References

1. Abdelmalek, S., Azar, A.T., Dib, D.: A novel actuator fault-tolerant control strategy of dfig-based wind turbines using takagi-sugeno multiple models. Int. J. Control Autom. Syst. **16**(3), 1415–1424 (2018)
2. Al-Baghdadi, M.A.S.: Modelling of proton exchange membrane fuel cell performance based on semi-empirical equations. Renew. Energy **30**(10), 1587–1599 (2005)
3. Ammar, H.H., Azar, A.T., Shalaby, R., Mahmoud, M.I.: Metaheuristic optimization of fractional order incremental conductance (FO-INC) maximum power point tracking (MPPT). Complexity 2019(Article ID 7687891), 1–13
4. Azar, A.T., Radwan, A., Vaidyanathan, S. Fractional Order Systems: Optimization, Control, Circuit Realizations and Applications. Academic Press, Cambridge (2018a)

5. Azar, A.T., Radwan, A., Vaidyanathan, S.: Mathematical Techniques of Fractional Order Systems. Elsevier, Amsterdam (2018b)
6. Benchouia, N., Hadjadj, A., Derghal, A., Khochemane, L.: Modeling and validation of pem fuel cell. Revnew. Energy **16**(2), 265–77 (2013)
7. Benchouia, N.E., Elias, H.A., Khochemane, L., Mahmah, B.: Bond graph modeling approach development for fuel cell PEMFC systems. Int. J. Hydrog. Energy **39**(27), 15,224–15,231 (2014)
8. Derbeli, M., Farhat, M., Barambones, O., Sbita, L.: Control of pem fuel cell power system using sliding mode and super-twisting algorithms. Int. J. Hydrog. Energy **42**(13), 8833–8844 (2017)
9. Dixit, T.V., Bankupalli, P.T., Yadav, A., Gupta, S.: Fuel cell power conditioning unit for standalone application with real time validation. Int. J. Hydrog. Energy **43**(31), 14,629–14,637 (2018)
10. El-Sharkh, M., Rahman, A., Alam, M.: Neural networks-based control of active and reactive power of a stand-alone pem fuel cell power plant. J. Power Sources **135**(1–2), 88–94 (2004a)
11. El-Sharkh, M., Rahman, A., Alam, M., Byrne, P., Sakla, A., Thomas, T.: A dynamic model for a stand-alone pem fuel cell power plant for residential applications. J. Power Sources **138**(1–2), 199–204 (2004b)
12. El-Sharkh, M., Rahman, A., Alam, M., Sakla, A., Byrne, P., Thomas, T.: Analysis of active and reactive power control of a stand-alone pem fuel cell power plant. IEEE Trans. Power Syst. **19**(4), 2022–2028 (2004c)
13. Fekik, A., Denoun, H., Azar, A.T., Almustafa, K.M., Almakhles, D., Zaouia, M., Hamida, M.L., Yassa, N.: A self tuning fuzzy-fractional-order p-i controller for speed control of dc motor. In: IEEE 2019 International Conference on Control, Automation and Diagnosis (ICCAD), pp. 167–172 (2019)
14. Feliu-Talegon, D., Feliu-Batlle, V., Tejado, I., Vinagre, B.M., HosseinNia, S.H.: Stable force control and contact transition of a single link flexible robot using a fractional-order controller. ISA Trans. **89**, 139–157 (2019)
15. Luo, F.L., Ye, H.: Advanced dc/dc converters. CRC Press, Boca Raton (2016)
16. Mann, R.F., Amphlett, J.C., Hooper, M.A., Jensen, H.M., Peppley, B.A., Roberge, P.R.: Development and application of a generalised steady-state electrochemical model for a pem fuel cell. J. Power Sources **86**(1–2), 173–180 (2000)
17. Meghni, B., Dib, D., Azar, A.T.: A second-order sliding mode and fuzzy logic control to optimal energy management in wind turbine with battery storage. Neural Comput. Appl. **28**(6), 1417–1434 (2017a)
18. Meghni, B., Dib, D., Azar, A.T., Ghoudelbourk, S., Saadoun, A.: Robust adaptive supervisory fractional order controller for optimal energy management in wind turbine with battery storage. In: Azar, A.T., Vaidyanathan, S., Ouannas, A. (eds.) Fractional Order Control and Synchronization of Chaotic Systems, pp 165–202. Springer International Publishing, Cham (2017b)
19. Meghni, B., Dib, D., Azar, A.T., Saadoun, A.: Effective supervisory controller to extend optimal energy management in hybrid wind turbine under energy and reliability constraints. Int. J. Dynam. Control **6**(1), 369–383 (2018)
20. Mohan, N., Undeland, T.M.: Power Electronics: Converters, Applications, and Design. Wiley, Hoboken (2007)
21. Ntouskas, S., Sarimveis, H., Sopasakis, P.: Model predictive control for offset-free reference tracking of fractional order systems. Control Eng. Pract. **71**, 26–33 (2018)
22. Podlubny, I.: Fractional-Order Systems and Fractional-Order Controllers, vol. 12, Issue 3, pp. 1–18. Institute of Experimental Physics, Slovak Academy of Sciences, Kosice (1994)
23. Ren, H.P., Wang, X., Fan, J.T., Kaynak, O.: Fractional order sliding mode control of a pneumatic position servo system. J. Frankl. Inst. **356**(12), 6160–6174 (2019)
24. Rhouma, A., Bouzouita, B., Bouani, F.: Model predictive control of fractional systems using numerical approximation. In: 2014 World Symposium on Computer Applications & Research (WSCAR), IEEE, pp. 1–6 (2014)

25. Shah, P., Agashe, S.: Review of fractional pid controller. Mechatronics **38**, 29–41 (2016)
26. Sopasakis, P., Sarimveis, H.: Stabilising model predictive control for discrete-time fractional-order systems. Automatica **75**, 24–31 (2017)
27. Suresh, R., Sankaran, G., Joopudi, S., Choudhury, S.R., Narasimhan, S., Rengaswamy, R.: Optimal power distribution control for a network of fuel cell stacks. J. Process Control Effic. Energy Manag.**74**, 88–98 (2019)
28. Tao, W., Min, C., Liu, X., He, Y., Yin, B., Jiang, W.: Parameter sensitivity examination and discussion of pem fuel cell simulation model validation: Part I. Current status of modeling research and model development. J. Power Sources **160**(1), 359–373 (2006)
29. Tepljakov, A.: Fomcon: fractional-order modeling and control toolbox. In: Fractional-order Modeling and Control of Dynamic Systems, pp. 107–129. Springer, Berlin
30. Thangavelautham, J., Strawser, D.D., Dubowsky, S.: The design of long-life, high-efficiency pem fuel cell power supplies for low power sensor networks. Int. J. Hydrog. Energy **42**(31), 20,277–20,296 (2017)
31. Wang, C., Nehrir, M.H., Shaw, S.R.: Dynamic models and model validation for pem fuel cells using electrical circuits. IEEE Trans. Energy Convers. **20**(2), 442–451 (2005)
32. Wang, J., Shao, C., Chen, Y.Q.: Fractional order sliding mode control via disturbance observer for a class of fractional order systems with mismatched disturbance. Mechatronics **53**, 8–19 (2018)
33. Xiong, L., Li, P., Ma, M., Wang, Z., Wang, J.: Output power quality enhancement of pmsg with fractional order sliding mode control. Int. J. Electr. Power Energy Syst. **115**(105), 402 (2020)
34. Yang, B., Wang, J., Sang, Y., Yu, L., Shu, H., Li, S., He, T., Yang, L., Zhang, X., Yu, T.: Applications of supercapacitor energy storage systems in microgrid with distributed generators via passive fractional-order sliding-mode control. Energy **187**(115), 905 (2019)

Battery Pack Construction Scheme Based on UPS System Reliability

Jing Xu, Xiangluan Dong, Dongxu Dai, Changjun Shi, Wenrui Li, and Yi Xiang

Abstract As an important part of the UPS system, the performance of the battery pack has a direct impact on the reliability, security, and safety of the whole system. Therefore, in recent years, the research on the feasibility of supporting battery selection for UPS system operating environment and performance requirements has been gradually expanded and deepened, to achieve UPS uninterrupted power supply and cost reduction. By comparing the performance of acid-proof exhaust type battery and valve-controlled lead-acid battery, this paper introduces the selection principle of the battery pack and expounds the reason why the UPS power system chooses valve-controlled lead-acid battery. An example is given to illustrate how to calculate battery capacity and its selection and configuration, the influence of environmental temperature and safety factor on the calculation result of battery capacity is discussed, and the optimization method of battery capacity calculation and selection for UPS power supply system is put forward.

Keywords Battery pack · Battery selection · Capacity calculation

J. Xu (✉)
State Grid Chaoyang Electric Power Supply Company, State Grid Liaoning Electric Power Supply Co., Ltd., Chaoyang, China
e-mail: 13940395315@163.com

X. Dong · W. Li
Graduate Department, Shenyang Institute of Engineering, Shenyang, China

D. Dai
State Grid Benxi Electric Power Supply Company, State Grid Liaoning Electric Power Supply Co., Ltd., Benxi, China

C. Shi
State Grid Tieling Electric Power Supply Company, State Grid Liaoning Electric Power Supply Co., Ltd., Tieling, China

Y. Xiang
State Grid Dandong Electric Power Supply Company, State Grid Liaoning Electric Power Supply Co., Ltd., Dandong, China

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_65

1 The Series Application of UPS

UPS is composed of a large number of electronic components, power devices, the cooling fan and other electrical devices of power electronic devices. In some emerging cases such as power supply fault, we may use a single UPS power supply with around 100,000 h. The reliability of the power supply system can be greatly improved by using the double machine hot backup redundancy technology. UPS hot backup can be divided into series backup and parallel backup. Figure 1 shows the main circuit of the tandem hot backup operation

Connect the hot backup UPS output voltage to the bypass input of the host UPS. When the UPS host is working normally, it will assume all the load power. When the UPS host fails, it will automatically switch to the bypass state [1]. The UPS standby machine will continue to supply power to the load through the bypass channel of the UPS host. When the power supply is interrupted, UPS is in the working state of battery discharge. Since the UPS host machine bears all the load power, its battery is discharged to the termination voltage first, and then automatically switched to the bypass working state. Backup UPS supplies power for the load until the battery discharge of standby machine ends. The main machine of UPS should be synchronized with the standby machine and the standby machine with the power supply [2]. The static switch in UPS is an important component that affects the reliability of the power supply system. If the static switch fails, both the main and standby UPS cannot supply power for the load. This kind of operation mode is reliable, simple in technology, and low in cost. But because its capacity cannot be effectively used, the relative investment cost is 100% higher than the single machine investment cost, so the cost–performance ratio is not cost-effective for general users [3].

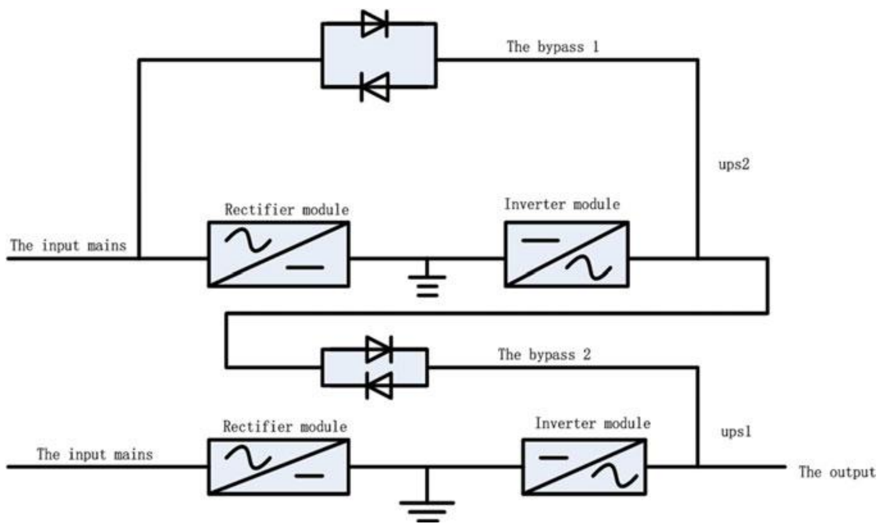


Fig. 1 Schematic diagram of the main circuit of the UPS series hot standby operation

2 Parallel Application of UPS

Figure 2 shows the operation mode of redundant power supply in parallel of two machines. The UPS used for parallel connection of two machines must have the function of parallel connection. The parallel control circuit of two machines in UPS is connected by a parallel signal line to keep the frequency, phase, and amplitude of output voltage of the two machines consistent. This kind of parallel connection is mainly to improve the reliability of the UPS power supply system, rather than for the expansion of the power supply system, so the total load capacity should not exceed the rated output capacity of one of the UPS [4]. In the event of a failure of one of the UPS, the other UPS can carry the full load current. These two redundant, parallel power UPS often operate at low efficiency because their output capacity is less than 50% of the rated capacity. This kind of dual machine mode has high reliability and the parallel operation is achieved by internal microcomputer. The main machine and standby each have a 50% load, and when the main machine is down, the backup device is fully started.

The UPS with parallel function can generally allow three sets, and some even allow six sets to be used in parallel. The conversion efficiency and equipment utilization rate of multiple UPS parallel power supply systems are higher than that of two UPS parallel power supply systems [5]. For example, for a load system with a capacity of 260 kVA, four parallel UPS systems with a rated capacity of 100 kVA may be used to supply power. At this point, there is "1 + 3" redundancy, that is, if one of the three UPS fails, it will not affect the normal power supply to the load; When the power supply system works normally, each UPS bears a load of 65 kVA, and the equipment utilization rate is 65%. Compared with "1 + 1" parallel redundant power supply system, this kind of "3 + 1" parallel redundant power supply system has a higher operation economy.

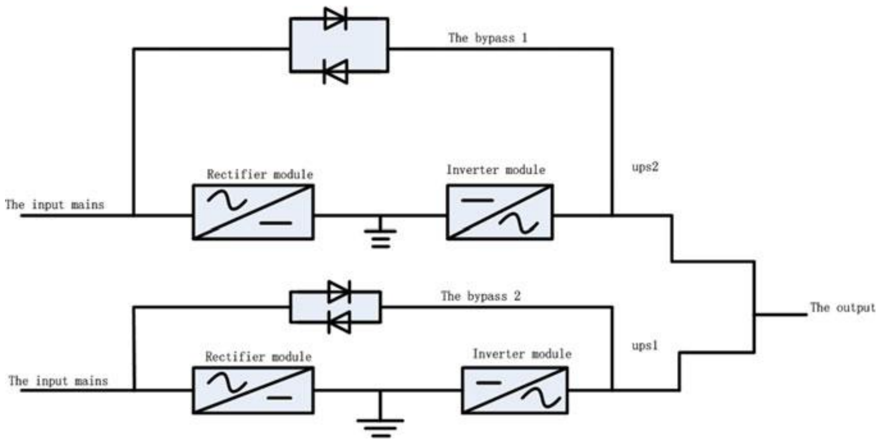


Fig. 2 Main circuit diagram of the parallel redundant operation of two machines

Table 1 Acid-proof exhaust battery and valve-controlled closed lead-acid battery comparison

#	Acid-proof exhaust lead-acid battery		Valve-controlled closed lead-acid battery
1	Discharge capacity	When the capacity is small, it is the same. When the capacity is large, it is weak	When the capacity is small, it is the same. When the capacity is large, it is stronger
2	Thermal runaway	Do not worry about losing control	May cause battery damage
3	Life	15–20	10–12
4	Reliability	Easy to observe the battery internal conditions and easy to detect the situation early	Not easy to observe the internal conditions of the battery and it is difficult to detect the situation early
5	Running performance	Rich experience in nuclear power plant operation	No rich experience in nuclear power plant operation
6	Maintenance	Heavy maintenance workload	Small maintenance workload
7	Decorate	Must be placed upright, covering a large area	Can be placed horizontally or vertically, covering a small area

3 Battery Selection

The electric storage pool has the characteristics of recharging and discharging, making it convenient to use, which makes it irreplaceable in many application fields. There are two kinds of lead-acid battery, and they are in the category of lead-acid battery, in which the valve-controlled sealed lead-acid battery is the main one. Generally speaking, acid-proof exhaust lead-acid battery and valve-controlled lead-acid battery, have higher acid-proof, explosion-proof requirements, at the same time, their operating maintenance is workload [6]. However, due to the anti-acid exhaust lead-acid battery electrolyte, density can be measured at any time and transparent shell. It is easy to monitor the internal situation, so, the early error can occur in time and its reliability has greatly increased. In summary, valve-controlled lead-acid batteries are easy to use and acid-proof exhaust lead-acid batteries are more reliable. Table 1 summarizes the main differences between the two lead-acid batteries.

4 Determination of the Number of Batteries

Due to the UPS power system backup guarantees, the battery pack is directly connected to the safety and reliability of the entire data center power source. Therefore,

Table 2 2 V and 12 V single battery performance comparison table

	2 V	12 V
Design orientation	Long-time, small current, deep discharge	Short-time, high current and shallow discharge
Design life	10	3–6
Lead paste recipe	The long-life formula, long cycle life	High power formula, short cycle life
Over-discharge recovery	Good	Average
Nominal energy/A h	200–3000	≤200

the reasonable choice of battery type is an important link in the UPS power system design. In order to ensure that the UPS power supply can continue to supply power to the load in case of mains power failure, this paper comprehensively considered the equipment configuration system structure, equipment type, standby power supply situation, and maintenance level and selected the discharge time, and the better choice of a shorter discharge time, the worse choice of a longer discharge time. At present, commonly used UPS battery packs are mainly of two types of monomer: 12 V and 2 V. At the beginning of the design, these two battery packs have distinct differences [7]. Table 2 describes the differences between the two batteries in detail.

For example, if the machine room of a small data center is equipped with two parallel redundant UPS, each of which has a rated capacity of 200 kVA, the total capacity of the battery group should meet the requirement of a single UPS running at full load for 0.5 h after power failure. The rated voltage of the UPS battery set is 480 V, and the low voltage warning point of the inverter is 408 V. The number of nodes of each UPS battery set should be calculated [8].

As can be seen from the above example, if the battery with a single voltage of 2 V is selected, the number of power saving of the battery is 48/2 (V/single) = 240, and the termination voltage is 1.75 V; here, the discharge termination voltage is set as 240 (only) * 1.75 V = 420 V > 408 V; If a single battery with a voltage of 12 V is selected, the number of power savings of the battery set is 48/12 (V/monomer) = 40, and the termination voltage is taken as 10.5 V, where the discharge termination voltage is set as 40 (only) * 10.5 V = 420 V > 408 V to meet the design requirements.

As the inverter of UPS is set with low voltage protection, when the number of electricity is a little more, the voltage of a single battery will be a little lower when the inverter is automatically closed at a low voltage, to extend the discharge time and give full play to the battery capacity. But in this way, it is easy to make the battery discharge late single battery voltage is too low, resulting in over-discharge. So, to protect the battery, the number of batteries should be less and the design needs to properly deal with this contradiction. Through the above analysis, it is concluded that the data center UPS system battery pack can meet the design requirements of this paper by selecting 240 2 V single-cell batteries and 40 12 V single-cell batteries, respectively [9].

5 Calculation of Battery Capacity in Engineering

The capacity of the battery group should be configured according to the short-term load, and according to the life of the battery, the long-term development should be properly considered. In the UPS power supply system, each UPS is generally set with a group of batteries. When the capacity is insufficient, it can be connected in parallel. Even if a UPS is connected with two groups of batteries (two groups of batteries are connected in parallel), it is still considered that the UPS is connected with one group of batteries [10].

5.1 *Calculation of Battery Capacity by Theoretical Formula Estimation Method*

The designer can set the discharge hour rate of the total capacity of the battery in the system accordingly [11]. The backup time of the battery in this design is not less than 0.5 h. According to the design of the specification, the total capacity of the battery set should be calculated according to the following formula:

$$Q \geq \frac{KIT}{\eta[1 + \alpha(t - 25)]} \tag{1}$$

where Q is the capacity of the battery, A h; K is the safety factor, taking 1.25; I is the load current, A; T is the number of discharge hours, h; vertical is the discharge capacity coefficient. T is the lowest ambient temperature value of the actual battery location. Home to have a heating device, according to 15 °C, no heating device, according to the 5 °C consideration; Alpha to battery temperature coefficient, 1/°C, when discharge hourly rate of 10 or more, take alpha = 0.006; When 10> discharge hour rate is greater than or equal to 1, take alpha = 0.008; When the discharge hour rate <1, take alpha = 0.01.

With 200 kV, the UPS power as an example to illustrate the process of calculating the battery, the UPS battery system for the data center can choose a different 240 monomer battery section and a 12 volt 40-knot battery, the calculation process of the two cases are the same, this section only choose 2 V monomer battery section 240 battery capacity calculation [12].

Calculating the load current of the discharge current of the UPS power supply using the following formula:

$$I = \frac{1000\, P \cos \alpha}{\eta NE} \tag{2}$$

where I is the discharge current of UPS; P is rated capacity of UPS, KVA; cos is the power factor, which is set as 0.9; vertical is the efficiency of the inverter, which is 0.94

in this design; N is the number of batteries, 2 V, take 240; and E is the termination voltage of discharge in a single storage tank. In this design, it is 1.75 V/monomer.

$$I = \frac{1000\,P\cos\alpha}{\eta NE} = \frac{1000 \times 200 \times 0.9}{0.94 \times 240 \times 1.75} A = 456\,A \tag{3}$$

Calculation of battery capacity: UPS as an example to calculate the battery capacity, UPS is installed in the indoor heating; so, $t = 15$ °C; the discharge time was 0.5 h, so if the discharge hour rate was less than 1 in this example, take = 0.01; the discharge termination voltage was 1.75 V, and take the discharge capacity coefficient = 0.4. Plug into the formula for the capacity calculation coefficient C:

$$C = \frac{KT}{\eta[1 + \alpha(t - 25)]} = \frac{1.25 \times 0.5}{0.4 \times [1 + 0.01 \times (15 - 25)]} = 1.74 \tag{4}$$

By substituting the calculated coefficient C into Eq. (3), the lead-acid battery capacity is obtained as follows: $Q = CI = (567 * 1.74)$ A h = 793.4 A h. In this paper, two redundant UPS in parallel are distributed according to the storage battery capacity at full load, that is, two groups of 400 A h/2 V valve-controlled lead-acid batteries are selected.

5.2 *Calculation of Battery Capacity by Constant Power Algorithm*

The process of calculating battery capacity by constant power method generally adopted by battery manufacturers is summarized as follows:

1. Calculate the total battery power (W) to be configured according to the UPS product nominal power, load efficiency, and inverter efficiency [13, 14].
2. According to the total power supply, standby time, and nominal voltage of the battery pack, the total capacity of the battery pack to be configured is calculated [15].
3. According to the nominal voltage of the UPS product battery pack, calculate the cell number of the battery pack [16].
4. According to the total power of the battery supply and the number of single-cell batteries in the battery pack, calculate the discharge power of the single-cell battery to be configured [17].
5. According to the total capacity of the battery pack, determine the selection scheme of different battery configurations under the condition that the number of configuration groups does not exceed 4 [18].

The above calculation process ignored the impact of temperature on discharge and also did not consider the safety factor, so the calculation of the battery capacity of small, but also the battery manufacturers and designers of the results are inconsistent [19, 20]. Based on the calculation process of battery manufacturers, add the coefficient of the influence of temperature on discharge and the safety coefficient and get the calculation formula of battery capacity is:

$$P = \frac{KS_{ups}\cos\theta}{\eta N[1 + \alpha(t - 25)]} \tag{5}$$

where P battery is the single power of 2 V battery, k W; K is the safety factor, taking 1.25; S_{ups} is the nominal output power of UPS power supply, kVA; As the power factor of the load; vertical is the efficiency of the inverter, take 0.94; N is the number of single battery in series, which is 240 in this design. K for the influence of temperature on the discharge coefficient, this design in indoor heating $t = 15$ °C, the $k = (1 + \text{alpha } (t - 25)] = 0.9$. By substituting the above parameters into Eq. (4), it can be concluded that the monomer power of each 2 V battery is:

$$P = \frac{KS_{ups}\cos\theta}{\eta N[1 + \alpha(t - 25)]} = \frac{1.25 \times 200 \times 0.9}{0.94 \times 240 \times [1 + 0.01 \times (15{-}25)]} W = 1108.2W \tag{6}$$

6 Conclusion

This paper introduces the principle of lead-acid battery and its application in UPS power system. By comparing the acid-proof exhaust lead-acid battery with the valve-controlled lead-acid battery, this paper expounds that the valve-controlled lead-acid battery has the characteristics of small size, good sealing performance, and no maintenance and is more suitable for the UPS power system in the machine room of the data center. Finally, combined with capacity conversion method and current conversion method, this paper discusses the influence of environmental temperature and safety factor on battery capacity and optimizes the method of selecting battery capacity of UPS power supply system.

Acknowledgements This paper is supported by "State grid science and technology project (2019YF-29)."

References

1. Singh, N.B.F., Singh, S., Chandra, A.T.: A review of three-phase improved power quality AC/DC converters. IEEE Trans. Industr. Electron. **51**(3), 641–660 (2014)
2. Bruckner, T.F., Holmes, D.G.S.: Optimal pulse-width modulation for three-level inverters. IEEE Trans. Power Electron. **20**(1), 82–89 (2015)

3. Diego, R.V.F.: A novel load current control method for a leading power factor voltage source PWM rectifier. IEEE Trans. PE **9**(2), 153–159 (2013)
4. Kang, Y.F., Song, T.H.S., Chiu, H.J.T.: Analysis and elimination of voltage imbalance between the split capacitors in half-bridge boost rectifiers. IEEE Trans. Ind. Electron. **49**(3), 5–8 (2012)
5. Raou, M.F.: Average current mode control of a voltage source inverter. J. Electr. Eng. **4**(3), 3–4 (2014)
6. Zang, K.F., Kang, Y.S., Chen, J.T.: Direct repetitive control of SPWM inverter for UPS purpose. IEEE Trans. Power Electron. **18**(3), 784–792 (2014)
7. Hara, S.F., Yamamoto, Y.S., Omata, T.: Repetitive control system: a new type sevro system for periodic exogenous signals. IEEE Trans Autom. Control. **33**(2), 7 (2014)
8. Emadi, A.F., Khaligh, A.S.: Instability in automotive systems: definition modeling, stability, and control of power electronic converters and motor drives. EEE Trans. Veh. Technol. **55**(6), 1112–1125 (2016)
9. Gigore, V.F., Hatonen J.S., Kyyra, J.T.: Dynamics of a buck converter with a constant power load. In: Annual IEEE Power Electronics Specialists Conference, vol. 1(3), pp. 72–78. Fukuoka, IEEE (2017)
10. Singh, J.F., Singh, B.N.S., Chandra, A.T.: A review of three-phase improved power quality AC-DC converters. IEEE Trans. Ind. Electron. **7**(5), 78–80 (2014)
11. Siahaan, A.P.U.: Rabin-Karp elaboration in comparing pattern based on hash data. Int. J. Secur. Its Appl. **12**(2), 59–66 (2018)
12. Basri, A.B., Badron, K., Ismail, A.F., Chanik, A., Jamil, S.F., Salim, H.: Determination of correlation coefficients for RazakSAT received signals. Int. J. Futur. Gener. Commun. Netw. **11**(2), 1–10 (2018)
13. Barotova, S., Drahansky, M., Pernicky, R.: Detection of ridge damages in fingerprint recognition caused by skin diseases. Int. J. Signal Process. Image Process. Pattern Recognit. **9**(11), 125–146 (2016)
14. Sahamijoo, A., Piltan, F., Jaberi, S.M., Sulaiman, N.: Prevent the risk of lung cancer progression based on fuel ratio optimization. Int. J. of u- and e- Serv. Sci. Technol. **8**(2), 45–60 (2015)
15. Peng, R.Y., Chen, J., Li, Y.: Research on users' behaviors in virtual communities from the perspective of information ecology. Int. J. of u- and e- Serv. Sci. Technol. **10**(1), 129–140 (2017)
16. Shuja, M., Mittal, S., Zaman, M.: Decision support system for prognosis of diabetes using non-clinical parameters and data mining techniques. Int. J. Database Theory Appl. **11**(3), 39–48 (2018)
17. Dhakar, J.S., Parmar, M.: A survey on grid based data aggregation in wireless sensor networks. Int. J. Sens. Its Appl. Control. Syst. **4**(1), 11–18 (2016)
18. Kim, J., Lee, S.: Walk-ability evaluation of pedestrian facilities. International Journal of Transportation **6**(1), 75–94 (2018)
19. Bhattacharyya, D.: Garlic expert advisory system using parallel ABC algorithm. Int. J. Disaster Recover. Bus. Contin. **8**, 11–20 (2018)
20. Thakur, S.: Modeling and performance analysis of MEMS based capacitive pressure sensor using different thickness of diaphragm. Int. J. Sens. Its Appl. Control. Syst. **4**(2), 21–30 (2016)

Study on Land Compensation for High-Voltage Transmission Lines in Power Grid Based on Easement

Jingbo Liu, Ou Zhang, Tong Wu, Ping Zhang, Ruyu Zhang, and Jianfeng Lin

Abstract Land compensation for power grid construction is a very important link in the process of power grid construction. In the process of power grid construction, we often encounter land compensation and other related problems that need to be solved. Based on the law currently being implemented in our country, this paper designs a reasonable compensation method for land expropriation for power grid construction to solve the related problems faced by land compensation for power grid high-voltage transmission lines. This method can not only promote the normal construction of the national power grid but also ensure the rights and interests of the broad masses of the people, in particular. It is the interests of the peasants who have been requisitioned for land in construction. It can reduce the cost of electric power enterprises to the greatest extent, at the same time, the profit of electric power enterprises can be improved, so that the construction of electric power infrastructure can be a sustainable development. At the same time, by analyzing a case, this paper illustrates that the land compensation for high-voltage transmission lines based on easement has certain rationality and development prospects, which can alleviate the problem of the sharp increase of compensation costs to a certain extent and solve the contradictions existing in the external environment of power grid construction.

Keywords Easement · High-voltage transmission lines · Land expropriation compensation

1 Introduction

At present, there is no unified standard for land compensation of high-voltage transmission lines in China. At present, it is difficult to coordinate the relationship between different places in the construction of a power grid. Many disputes and conflicts of interest exist in the working process. With the gradual development of China's power

J. Liu (✉) · O. Zhang · T. Wu · P. Zhang · R. Zhang · J. Lin
Economic Technology Research Institute, State Grid Liaoning Electric Power Supply Co. Ltd, Shenyang, China
e-mail: 894376181@qq.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_66

enterprises, the layout of the power grid is constantly optimized and adjusted. With the continuous growth of transmission line length, the number of substations is also increasing. At the same time, the total area occupied by tower and substation construction caused by the construction of high-voltage transmission and distribution lines is expanding. Therefore, we need to find a reasonable way to solve the related problems of land compensation for high-voltage transmission lines in the power grid.

2 Compensation Principle of Easement for Power Grid Construction Land

2.1 *Basic Concepts of Easement*

An easement refers to the right to use other people's real estate by the contract, to improve the benefits of their real estate. Other people's real estate is servitude, and their real estate is servitude. China's Property Law stipulates that when the easement is established, the parties concerned shall conclude the easement contract in written form. If a party requests registration, he may apply to the registration authority for registration of easement. Without registration, he may not confront a bona fide third party [1]. The oblige of servitude shall, by the contract, allow the easement holder to use his land and shall not hinder the easement holder from exercising his rights. The term of the easement shall be agreed upon by the parties concerned, but shall not exceed the remaining term of usufructuary rights such as the right to contractual management of land and the right to use construction land. Where the owner of land enjoys or bears the easement right, the land contractual management right and the right to use the house land shall continue to enjoy or bear the established easement right when the land contractual management right and the right to use the house land are established [2].

2.2 *The Concept and Compensation Method of Easement Royalty*

The easement royalty of power grid construction land refers to the relevant compensation for the easement of power grid construction land, that is, a certain amount of compensation paid by power enterprises to the original landowners to obtain the easement of power grid construction land, the amount of which is mainly estimated by the right ratio method. The ideas for the evaluation of the right ratio method are as follows:

(1) p is the criterion for evaluating land price [3].

(2) Determine and quantify the degree of influence of setting easement of power grid construction land on evaluating land use, i.e., determine the compensation rate r.
(3) The formula for calculating easement royalty of power grid construction land is as follows:

$$v = \sum p * r \tag{1}$$

Here, the compensation rate reflects the impact and limitation of easement of power grid construction land on land use.

2.3 Determination of Compensation Rate

Taking rural land as an example, the compensation rate is expressed by the weighted average value of the product of the use function of transmission lines and the ratio of cable impacts on land.

Principles for determining the compensation rate: The product of the price of farmland and the compensation rate should not exceed the compensation fee of the same area of farmland [4], considering the actual situation of various places, maximizing the income of the legitimate rights of the peasants in servitude, and ensuring that their living standards do not decrease.

Formula for determining compensation rate: Compensation rate R + farmland function weight $A \times$ cable impact ratio Ra + farmland function weight $B \times$ cable impact ratio Rb + farmland function weight C * cable impact ratio Rc $+ \cdots$.

2.4 The Steps of Calculating the Compensation Rate

Determination of Main Functions of Agricultural Land: In this paper, we mainly consider the impact of power grid construction on farming function, carrying function, a landscape function, and asset function, which can be quantified [5].

Weight Determination: There are different emphases on the functional use of agricultural land. Considering its general characteristics, the paper gives a score based on experts'opinions.

Determination of the Ratio of the Impact of Overhead Lines on Agricultural Land Function: The influence of overhead lines on the farming function is negligible, which is set at 0. The main load-bearing functions of farmland are access, irrigation, simple shed, and building buildings and structures. The transmission lines of different voltages affect farmland. The impacts of carrying capacity are different, and the corresponding impact ratios are different. The impacts of overhead lines on landscape function are set to [6]. The impacts of overhead lines on assets function are mainly divided into social security, land value increment, and underground resources (such as minerals), etc. The main impacts of power grid construction are social security, land value increment, and underground resources (such as minerals). When it comes to land value added, it is set at 30%.

Taking the above factors into account, after weighted summation, the overhead compensation rate of farmland under the same voltage level is finally determined [7].

2.5 Compensation Method of Easement Royalty for Power Grid Construction Land

For state-owned land, the cost of compensation should be directly handed over to the local land administrative department. If there is a lease operation on state-owned land, it is also necessary to compensate for the loss of the lease operator.

For collective land, if the land has been contracted, we should make major compensation to the rights holder. If the land has not been contracted, the compensation should be paid to the collective economic organization [8].

The compensation of restricted land for power grid construction is divided into three parts: loss compensation fee, temporary land compensation fee and permanent land compensation fee for auxiliary facilities, which shall be paid at one time when the land for power grid construction is obtained; the transfer fee of land use right for power grid construction can be paid by stages, by years or by years [9].

3 Current Situation of Land Compensation for High-Voltage Transmission Lines

The compensation of easement royalty generally includes the compensation of land under the transmission line and the space right attached to it, but it is special for overhead lines, houses, tower foundation land, and buried pipes. The main forms are as follows [10].

3.1 Overhead Line

The compensation for overhead lines has been included in the corresponding easement royalties, so it does not need to be tested and calculated separately.

3.2 Housing Disposal

It has two situations, including house demolition and reasonable leap. For housing demolition, its price should be added to the other; for cross-building, because the house has special significance and role for the owner, the height of transmission line from the ground should be used as the basis of compensation rate [11]. According to a different height, the distance between the house and the vertical projection of the transmission line, the lines should be measured separately. The corresponding compensation formula for the compensation rate of corridor or line protection area is as follows:

$$v_c = p_1 \times s \times r_c \tag{2}$$

where v_c is the easement royalty in the case of cross-housing;s is the area occupied by the house; p_1 is the corresponding assessment of land price; r_c is the compensation rate of easement in the corresponding region.

Cross Room: Since the compensation of easement royalty includes the compensation of space right of land and its attachments under transmission lines, the compensation for the cross-building of transmission lines can be included in the easement royalty in theory [12]. However, due to the special significance and role of the house for the owner, it is possible to deal with the transmission line across the house accordingly. According to the existing transmission line design regulations in China, 500 kV and above transmission lines are not allowed to cross houses (if they must pass, they must be demolished). Therefore, in the process of easement compensation, it is necessary to distinguish between 220 kV and below transmission lines and 500 kV and above transmission lines.

220 kV and Below Voltage Level Transmission Lines: In the process of compensation for easement of 220 kV and below voltage grade transmission lines, the easement scope refers to the area of the linear path through which the transmission line passes, the width of which is an overhead power line protection area [13]. Specifically, the area of two parallel planes formed by the horizontal extension of the edge of the guideline to the outside and perpendicular to the ground is the area of voltage conductors at all levels in general areas. The extension distance of the edge line is as follows:

$$\begin{aligned} 1-10\,\text{kV}\ 5\,\text{m} \\ 35-110\,\text{kV}\ 10\,\text{m} \\ 154-330\,\text{kV}\ 15\,\text{m} \\ 500\,\text{kV}\ 20\,\text{m} \end{aligned}$$

In densely populated areas such as factories, mines, and towns, the area of overhead power line protection area may be slightly smaller than the above provisions. However, the extension distance of the sidelines of voltage conductors at all levels should not be less than the sum of the maximum calculated sag and the maximum calculated horizontal distance after the wind deviation and the safe distance from the building after the wind deviation [14].

500 kV and Above Voltage Grade Transmission Line: In the process of easement compensation for 500 kV and above voltage level transmission lines, the easement scope is the same as 220 kV and below transmission lines. The difference is that the area formed under 500 kV and above voltage level transmission lines is not allowed to build houses, that is to say, if 500 kV and above voltage level transmission lines pass through. In the housing area, the off-line part must be demolished, and the outside part of the line is divided into restricted land.

3.3 Tower Base and Pole

The compensation for the borrowed land of the tower foundation and the pole still follows the principle of "only making up but not requiring compensation." At the same time, the compensation for the damaged seedlings and ground attachments in the construction of the tower foundation should be considered [15]. The compensation standard can refer to the compensation standard for temporary land use.

3.4 Buried Tube

For buried pipes, the right of an underground passage in easement can be used to compensate, and the standard of compensation can refer to the compensation formula of easement:

$$v_u = p_s \times s \times r_u \tag{3}$$

where v_u is the price of easement compensation; p_s is the evaluation price of the corresponding land; r_u is the rate of easement compensation in the case of buried pipes, taking the depth of buried pipes as the main index [16] (Tables 1 and 2).

Table 1 Compensation rate table of buried pipe easement for transmission lines

Depth of buried pipe	Easement compensation rate (%)
0.8–1.2	40
1.3–1.6	30
1.6–2.0	20
2.0–2.5	10
>2.5	5

Table 2 Compensation standard for buried pipeline easement of transmission line

Depth of buried pipe	Compensation standard for the easement (yuan/mu)			
	Planning area		Outside the planning area	
	Arable land and construction land	Other lands	Arable land and construction land	Other lands
0.8–1.2	16,000	8000	6400	3200
1.3–1.6	12,000	6000	4800	2400
1.6–2.0	8000	4000	3200	1600
2.0–2.5	4000	2000	1600	800
>2.5	2000	1000	800	400

4 Dispute Settlement for Land-Occupied Compensation of Transmission Lines

Even after dealing with the easement compensation of power enterprises, there will still be many rights disputes, and especially in the construction of power line corridors, such incidents occur very frequently. Therefore, after doing a good job of pre-compensation, power companies still need to pay attention to the prevention of land compensation disputes in the future. When dealing with disputes, we should think deeply about whether they violate the law, whether they damage the rights and interests of rights and human rights, and put forward relevant preventive measures to ensure the power and interests of power enterprises [17]. And what we call tort refers to is that the area occupied by the power line corridor interferes with the normal use of other buildings, causing infringement or loss to its rights holders. And it is applicable to the principle of liability for fault, that is, the following three conditions need to be met in order to determine that it constitutes a tort: (1) There are illegal acts in the process of power line layout or use by power companies; (2) the rights and interests of the rights and interests of the disturbed buildings are damaged; and (3) there is a causal relationship between the relevant acts of power companies and the damage of the injured rights and interests.

5 Examples

To better illustrate the compensation method of easement royalty, this paper takes the compensation of a high-voltage transmission line in a certain place as an example and calculates the compensation based on the easement. The project has many tower foundations at the same time [18]. The cost of the project is about 8 million Yuan. However, the cost of compensation for land occupied by tower foundation and for green seedlings accounts for more than one million Yuan, including compensation for long-term land occupied by tower foundation, compensation for road construction and site stacking, compensation for remnant soil cleaning and re-cultivation, compensation for green seedlings constructed by tower overhead line, and compensation for special breeding. In the process of compensation, the tower foundation of transmission line does not need to go through land expropriation formalities, only need to pay a one-time land compensation fee, and according to its nature, it belongs to borrowed use. This compensation calculation method does not take into account the compensation of the easement of the power grid construction land of the relevant oblige [19]. If the compensation is carried out according to the easement (taking the compensation standard of a certain place as a reference), the results are as follows: the price of farmland (Sa): compensation standard for paddy field, dry land, vegetable field, garden land, aquaculture water surface, farmland water utilization land, and construction land. 40,000 Yuan per mu, of which 14,400 Yuan per mu is land compensation fee, 19,800 Yuan per mu is resettlement subsidy (calculated according to per capita arable land is higher or equal to 1 mu, per capita arable land is lower than 1 mu is calculated on the basis of reality), and 20,000 Yuan per mu is other land expropriation compensation fee. Since most of the land involved in the project is the former, the price of agricultural land is calculated at 40,000 Yuan per mu. Because the total length of the project is 8.52 km and the voltage of the transmission line is 110 kV, the average width of the corridor is about 28 m, the total easement area of the project's power grid construction land is 23,850 m^2, that is to say, the total easement area of the project's power grid construction land is 238,560 m^2, 357.82 mu.

The corresponding farmland price Sa is $357.82 \times 40{,}000 = 14{,}312{,}884$ yuan. Compensation Rate (R): Since the voltage of the transmission line is 110 kV, the corresponding compensation rate is 7%. Easement royalties for power grid construction land: $v = s \times r = 14{,}312{,}884 \times 7\% = 1{,}001{,}901$ Yuan. The easement royalty is the purchase price of the easement of the land involved in the construction of the project. After possession of the right, the electric power enterprise can erect, maintain, and operate the transmission line within the space covered by the easement without paying other expenses, but in case of damage to the attachment on the ground of the easement obligor. At that time, corresponding compensation still needs to be made, and the compensation standard refers to the relevant established documents. After adopting the new standard, the total amount of compensation is $1{,}001{,}901 + 922{,}590 + 228{,}827 = 2{,}153{,}318$ yuan, an increase of $2{,}153{,}318 - 1{,}151{,}417 = 1{,}001{,}901$ yuan, an increase of 87.01%. The amount of compensation for the use of easement accounted for 23.01% of the total investment.

6 Conclusion

Generally speaking, the easement right of limited land for power grid construction is suitable for transmission line construction [20]. The establishment of easement can make the owner of land or the owner of land use right in the line channel understand their rights and determine their obligations, and can obtain clear rights after payment of compensation by power enterprises, and their rights and obligations can be effectively protected by relevant laws, at the same time, it can improve the efficiency of protection. The awareness of landowners to protect power equipment in the power grid area can improve the reliability of power grid operation. This way can play a very important role in promoting the development of China's electric power enterprises and has certain rationality and prospects. Therefore, the use of the easement-based compensation model for power grid construction land conforms to the trend of construction according to law and regulations, and the problem of increasing costs and uncontrollable risks can be fundamentally solved.

References

1. Guo, Q., Hou, Z.P.: Research on the compensation model of power grid construction land based on easement. Value Eng. **36**(26), 75–77 (2017)
2. Liu, Zh., Zhou, T., Liu, N., Xian, K.: Design of land compensation mechanism for power grid construction land-based on easement-the Fifth Study on conflict and mutual adaptation mechanism in the external environment of power grid construction. J. Electr. Power **25**(05), 353–359+365 (2016)
3. Liu, Zh., Liu, N., Xian, K.: Design of land compensation model based on easement royalty—research on conflict and mutual adaptation mechanism in the external environment of power grid construction. J. Electr. Power **25**(03), 181–184 (2015)
4. Yan, L.: Research on the compensation model of power grid construction land based on easement. J. Huzhou Norm. Univ. **32**(S1), 69–72 (2016)
5. Zhou, T., Liu, Zh., Wu, J.L., Yan, L.: Research on land expropriation compensation model for power grid construction land based on easement. Land Sci. China **24**(03), 30–35 (2015)
6. Hubbard, B., Qian, H., Patrick, C., Yang, W.: Safety awareness educational topics for the construction of power transmission systems with smart grid technologies. Australas. J. Constr. Econ. Build. **13**(3), 114–127 (2014)
7. Cai, M., Sun, X.: Institutional bindingness, power structure, and land expropriation in China. World Dev. **109**(10), 172–186 (2018)
8. Xie, Y., He, L.: Analysis on causes of rural land expropriation disputes based on Smith's model. Asian Agric. Res. **6**(6), 98–101 (2014)
9. Yi, Ch., Yan, Zh.: Legal knowledge, land expropriation, and agricultural development in rural China. Front. Econ. China **12**(1), 132–166 (2017)
10. Trembecka, A.: Analysis of surveying and legal problems in granting right-of-way and expropriation for the purpose of locating technical infrastructure. Geod. Cartogr. (2080-6736) **65**(1), 95–110 (2016)
11. Rahul, S., Vikas, K., Gulshan, K., Mritunjay Kumar, R., Se-Jung, L.: Analyzing the tradeoff between planning and execution of robotics process automation implementation in IT sector. Int. J. Adv. Sci. Technol. **120**, 49–58 (2018)
12. Md. Abdur, R., Md. Abu, H., Khaled, S., Md. Saeed, S.: Multipartite based test case prioritization using failure history. Int. J. Adv. Sci. Technol. **125**, 25–42 (2019)
13. Kyeong, H., Won-Sung, S.: Wearable computer network design using dual-role WUSB clustering. Int. J. Control. Autom. **8**(3), 127–134 (2015)
14. Tajinder, S., Madhu, K., Triveni Lal, P., Ahsan, C.: Current trends in text mining for social media. Int. J. Grid Distrib. Comput. **10**(6), 11–28 (2017)
15. Chang-Soo, C., Jeong-Man, S., Heeman, L.: Efficient iris recognition method for large scale database. Int. J. Secur. Appl. **9**(3), 99–112 (2015)

16. Rae Hyun, K., Bang Won, S., Jeong Gon, K.: Packet scheduling by buffer management in mac protocol of medical WBAN monitoring environment. Int. J. Futur. Gener. Commun. Netw. **9**(11), 15–26 (2016)
17. Yiqin, C., Xiaoci, Z., Xiaosheng, H.: A road vehicle detection algorithm based on compressive sensing. Int. J. Signal Process. Image Process. Pattern Recognit. **9**(1), 1–10 (2016)
18. Min, Z., Shengbo, Y.: Accuracy estimation of a classifier based on the differences in samples. Int. J. Signal Process. Image Process. Pattern Recognit. **9**(11), 147–160 (2016)
19. Shafaq, S., Mengke, Y., Mahmood, M., Jianqiu, Z.: Teachers' willingness about the use of ICT through M-learning in Pakistan. Int. J. u- e-Serv. Sci. Technol. **10**(1), 71–88 (2017)
20. Olufunke, O.O., Amenawon, I.E.: The impact of some socio-economic factors on academic performance: a fuzzy mining decision support system. Int. J. Database Theory Appl. **10**(6), 71–86 (2017)

Study on the Terrain of Power Network Construction Expropriation and Legal Risk Prevention

Xiaozheng Yang, Ge Li, Songnan Liu, Di Wang, Shuangying Zhang, and Xuefeng Wu

Abstract In recent years, with the doubling of investment in infrastructure construction and key projects in China, the construction of a national power grid has also entered a stage of rapid development. In the face of the situation requires a high standard and the fast pace of construction, demolition, compensation disputes patients, keep the traditional problem has become the main reason for the power grid construction has been blocked, the frequency of the evolution of litigation and disputes are becoming more and more high, how to analyze the legal risks faced by power grid construction, and puts forward some countermeasures to prevent and resolve the legal risk, is the main problem we must consider. Based on the current grid construction land requisition of several related issues, combined with the relevant laws and regulations, this paper discusses and analyzes the relevant legal risks of grid construction land, and puts forward some corresponding solutions to these risks.

Keywords Land for network construction · Legal risk

1 Introduction

At present, many contradictions between power grid construction and land resources utilization hinder the rapid development of power grid construction, including all kinds of overhead lines, towers, and others, involving space rights, easement rights restrict the rights and interests of land users. Especially after the implementation of the Property Law of the People's Republic of China, the people's awareness of safeguarding their rights has been raised, such as what and how to guarantee the social security fees of land-expropriated peasants, the conflicts between the regulations for construction of electric power projects and the regulations for protection of electric power facilities, the unification of the fees collected by governments at all levels

X. Yang (✉) · G. Li · S. Liu · D. Wang · S. Zhang · X. Wu
Economic Technology Research Institute, State Grid Liaoning Electric Power Supply Co. Ltd, Shenyang, China
e-mail: 350376718@qq.com

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9_67

and the actual compensation standards for electric power projects, and so on. The problems will become the legal risk factors of power grid construction. If effective preventive measures are not taken, certain legal problems will inevitably arise. Also, as the technical standards for environmental assessment of power grid projects have not yet been issued, the environmental impact of power transmission and transformation projects on the bottom, whether there is the electromagnetic impact on nearby residents and other issues have brought many negative impacts on the environmental assessment of power grid projects.

2 Some Problems of Land Expropriation in Power Grid Construction Project

2.1 Relevant Problems of Substation Site Land Use

Because the choice of substation site must be based on the requirements of power supply point and overall layout of the system, many planned substation sites cannot be incorporated into the overall land use planning of provinces and municipalities in time, and lack of land use planning. The government agrees that the land used for substation construction is non-construction land, and many of them are basic farmland or some of them occupy basic farmland. At the same time, influenced by realistic factors such as time limit requirement, the phenomenon that power grid project starts first and then completes relevant procedures has been widespread for a long time.

2.2 Relevant Problems of Tower Land Use in Transmission Line Engineering

With the rapid development of social economy, increasing load and increasing demand for electricity, to meet the demand for electricity, many transmission lines not only cross urban areas but also across provinces and regions. On pole and tower land, not only the lines are long, but also many points are set up, and the situation is very complex. For each region, there are differences in the way power grid companies acquire tower land. Some regional power grid companies do not go through the formalities of conversion of construction land or land expropriation when they use tower land below 220 kV [1]. Instead of acquiring land use right, they adopt a way of compensation and permanent land use to solve the problem.

2.3 Relevant Problems of Compensation for Construction Land

It mainly includes two aspects: First, the way of compensation is not standardized enough. It can be said that the compensation agreement should belong to the civil category, and the object of compensation is the land contractor or the rural collective economy. In actual operation, when the construction units carry out tower construction, some construction units are not standardized on the signing of compensation agreements, resulting in a series of hidden dangers. Secondly, the subject of compensation is not clear enough [2]. The compensation for land expropriation arising from the construction land of power grid projects is paid to the government departments at one time by the power grid company and then handled by the local government. However, the reality is not always operated by the norms, which makes the subject of compensation obligation in a vague state. Also, power grid companies should appear in the subject of compensation obligations, but some local power grid companies did not appear in this agreement, which brought trouble to themselves [3]. Third, the compensation standard is not clear enough. The voltage level of the project is different, and the compensation standard is not distinguished, which makes the compensator feel unfair, provokes contradictions easily, and even uses various excuses to obstruct the construction of the project.

2.4 Environmental Problems

Article 90 and other articles of the Property Law directly relate to the adjacent environmental rights and electromagnetic effects. With the rapid development of economy, the corresponding distribution network scale is becoming larger and larger. At the same time, the awareness of protecting the rights of residents and villagers is increasing, and their knowledge and regulations on electromagnetic radiation are gradually understood [4]. Therefore, such environmental problems will inevitably face high risk with the implementation of the Property Law.

3 Legal Risk Analysis of Power Grid Construction

With the establishment and perfection of laws and regulations for the protection of property rights in China, the public's awareness of property rights has gradually improved [5]. The legal relationship of property rights involved in power grid construction is more complex, and the contradictions will become more prominent and sharp. Especially after the implementation of the Property Law, the protection of private property has been strengthened, the threshold of requisition and occupation of land for power grid construction has been raised, and the scope of compensation for land requisition has been expanded. The requisition of land for power grid construction is facing more difficulties and legal risks. Specific performances are in the following aspects.

3.1 Legal Risks in Substation Site Land

According to the Land Management Law and the Measures for Pre-examination and Management of Land Use for Construction Projects, when substation sites are not included in the overall land use planning of provinces and municipalities, the government needs to adjust the overall land use planning according to the principle of "balance between occupancy and compensation," submit approval at different levels, and even hold hearings to carry out the conversion of construction land, which will lead to substations. Construction needs to go through a long period of planning adjustment and approval [6]. But if we do not go through the formalities of conversion of construction land, the construction land of the substation will face the risk of "no evidence" in law. At the same time, with the implementation of the Property Law, in view of the phenomena of starting construction before completing relevant formalities for power grid projects, landowners can obstruct construction at any time for reasons, such as development lease, residents' income, and take power grid enterprises to court for infringement of the rights holder's easement, thus bringing risks to normal and legal operation of enterprises [7].

3.2 Legal Risk of Tower Land Use

According to the above explanations, in solving the problem of tower construction land, although it is to adopt a compensation and permanent use method to solve it, although the opinions of the landowners have been obtained, if compensation is regarded as a paid easement, in the actual operation, it is not treated according to the requirement of the easement set up in the Property Law because as stipulated in the Contract Law, the lease period cannot exceed 20 years, which is in contradiction with the so-called permanent right of use here. At the same time, the relevant documents stipulate that the construction land of poles and towers should also comply with the relevant provisions of the Land Management Law. In addition to the conversion procedures of construction land, the relevant procedures for examination and approval of land expropriation should also be handled to further obtain the right to use the land [8]. Therefore, there is a certain legal risk in the construction of poles and towers if the right to use the land for the construction of poles and towers is not obtained in accordance with the relevant legal provisions, and there is no easement agreement with the relevant person in charge, or a land lease contract with the relevant person in charge, while permanently using other people's land in the construction of poles and towers.

3.3 Legal Risks in Compensation for Power Grid Construction Land

The risk of compensation agreement: Compensation agreement is an important proof of compensation. If there is no written form or the content of the agreement is not standardized, it will bring risks to both sides of the agreement.

Risk of compensation standard fee for land expropriation: After the implementation of the Property Law, except for land compensation fees, resettlement subsidies, attachments on the ground and green seedlings, besides the compensation fees, the specific standards and contents of “arranging the social security fees of the land-expropriated peasants,” “guaranteeing the living conditions of the land-expropriated peasants,” and “guaranteeing the life of the land-expropriated peasants” required by the Property Law have not yet been clearly defined, such as what the social security fees of the land-expropriated peasants include and how to guarantee the living conditions of the land-expropriated. It may become the legal risk of power grid construction [9]. In addition, the traditional way of land expropriation and permanent restriction of “one-time compensation and permanent land use” for pole and tower land use is that the provincial government has not promulgated relevant laws and regulations to clarify the land occupied by pole and tower construction, and the transfer of expropriation has not gone through the examination and approval formalities. As the property right has not been transferred, after a certain period, the owner of property right has reason to request compensation fees based on price factors, land appreciation, and other reasons [10].

3.4 Legal Risks in Environmental Protection

In recent years, the voice of environmental protection has become higher and higher, and many power grid construction projects have been suspended or delayed due to lack of environmental assessment. The Real Right Law also makes clear provisions on environmental protection, such as Article 90 of the Real Right Law, which stipulates that the owners of real property may not discard solid waste in violation of state regulations and discharge harmful substances such as air pollutants, water pollutants, noise, light, and electromagnetic radiation. This provision clarifies the minimum tolerance obligations that should be borne by the adjacent owners of real property [11]. According to this provision, the injured real property right holder has the right to request the infringer to stop the infringement, eliminate the danger, remove the obstruction, and compensate for the loss. Whether the operation of the power grid will produce electromagnetic radiation, whether the electromagnetic radiation will cause harm to the human body, and how much harm [5]. Although the state has not yet issued emission standards for electromagnetic wave radiation, it is generally believed that the operation of the power grid will generate electromagnetic wave radiation, and the higher the voltage level, the greater the electromagnetic radiation.

With the implementation of the Property Law and the enhancement of the legal consciousness of the masses, such disputes are spreading. Especially in substations built in urban centers, such complaints are often the focus of such complaints. For example, the 110 kV Suwu substation in Guangxi power grid is located in the urban center, due to the impact of electromagnetic radiation complaints, the site selection has been delayed for more than three years, and the construction has been started only through the unremitting coordination of local governments and other parties [12].

4 Countermeasures and Suggestions for Actively Dealing with Legal Risks

The Property Law has brought a series of challenges to the power grid construction. Like a power grid enterprise, on the one hand, it is necessary to speed up the construction of power grid to meet the demand of economic and social development for power; on the other hand, it is necessary to actively respond to and take positive measures to adapt to the challenges of the relevant laws and regulations such as the Property Law to the power grid construction and strive to create a new power grid [13]. It is conducive to the social environment of power grid construction and fully demonstrates the social image and social responsibility of power grid enterprises.

4.1 Step by Step, Revise and Improve Electric Power Laws and Regulations

To carry out revision and improvement of electric power laws and regulations, while operating legal business according to laws and regulations, we should carry out the legislative revision work in the field of power management as soon as possible, so as to eliminate the internal inconsistencies in the power legal system, especially the inconsistencies between laws and regulations, technical regulations, and technical standards [14]. Also, we should actively study the conflicts between laws and regulations of power and other laws, and propose cross-industry and cross-field legislative revisions to the national legislature.

4.2 Promoting the Avoidance of Risks in Law Enforcement Level of Power Grid Construction Land

The implementation model of demolition and compensation should be clarified: At present, power grid construction, demolition, and compensation work are dominated by power grid enterprises, which need to negotiate with many demolition compensation objects one by one and spend a lot of time and energy on this link. To complete the task of power grid construction on schedule, power grid enterprises have to make huge concessions in the amount of compensation, which often leads to compensation costs exceeding budgets [15]. To this end, we will promote the re-positioning of the tripartite roles of the government, power grid enterprises, and demolition compensators: The government will become the contractor of demolition compensation from its previous witness status, and the compensation budget will be used by governments at all levels along with the project. Such a compensation implementation mode, led by administrative forces, is an improvement and optimization of the original mode of work, which can ensure that compensation costs can be controlled.

It is necessary to clarify the principle of uniform compensation and the standard of economic compensation: It is clear that the land for substation is allocated according to the law and the land for urban infrastructure; the overhead power line protection area does not impose land expropriation, and relocation and compensation are carried out according to the power design regulations, etc. In this way, the aforementioned disputes and ambiguities in practical operation caused by the conflict of laws or the inconsistency of applicable standards can be easily solved [16].

It is necessary to clarify the coordination mechanism and rights allocation of the management functions of various administrative departments: How to coordinate the construction of power grid with other facilities such as roads, railways, forest areas, municipal administration, waterways, and bridges is an old problem. Local governments should conscientiously maintain the authority and seriousness of planning when formulating relevant policies and systems, strictly abide by the principle of planning first, and the later builders should bear the migration and compensation costs of the built projects. At the same time, preferential policies should be implemented for the related administrative fees involved in power grid construction projects [17]. If crossing grade roads or rivers do not constitute substantial occupancy of the crossed objects, no crossing fee will be charged; if the power line corridor crosses the forest area, the formalities for occupying forest land may be exempted and the fees for occupying forest land in the line corridor will not be charged.

4.3 Pay Equal Attention to Protecting Environment and Strengthening Power Grid Construction and Build Green Power Grid

First of all, we must ensure the legitimacy and compliance of power grid construction projects. Power grid enterprises shall, in accordance with laws and regulations, strive to integrate power grid development planning into local national economic and social development planning, ensure that all links of power grid construction projects are legitimate, complete procedures and complete information, penetrate the design concept of green environmental protection into all links of design, equipment, construction, and operation, and effectively protect the lives and public environment of the public along the line and around the site, so as to avoid it [18]. The phenomenon of boarding tickets after the fact of violation of laws and regulations is avoided. Strengthen the prophase engineering depth of power grid construction project. Processing supporting documents for power grid construction one by one. Construction should abide by laws and regulations, relevant design regulations of electric power, and keep a safe distance from the built buildings, structures, trees, and another real estate. In project bidding and business negotiation, we should combine environmental protection measures, select equipment that meets environmental protection requirements, select low noise equipment, adopt noise control measures such as sound absorption and insulation, and strictly control the noise level of the transformer, reactor, and other main sound sources. Substation equipment is gradually developing indoors and underground. In principle, it is far away from residential buildings, hospitals, schools, and other environmentally sensitive areas. It is necessary to balance soil erosion and earthwork and properly solve the contradiction between electromagnetic environment, noise, radio interference, landscape coordination and meeting the needs of production and public life [19]. The compact and centralized transmission lines are widely used, the paths are integrated, and the towers with land saving and multiple parallel erections on the same tower are selected as far as possible to avoid a large number of a house demolition or spanning.

4.4 Improving the Level of Legal Risk Prevention and Control in Enterprises

Power grid enterprises should establish legal awareness, operate, and build according to law, and do not touch the legal bottom line. Strengthen enterprise risk management, make decisions according to law, operate according to law, and safeguard legitimate rights and interests according to law. Power grid enterprises should timely revise and improve the system, expand the legal risk prevention from simple litigation safeguard and contract management to the compliance examination of enterprise behavior, sort out and optimize the business process of power grid construction projects, analyze the links that often occur disputes and are prone to risks, and study the key and difficult points involved in policies, laws and external relations in power grid construction and operation [20]. Exhibition of legal argumentation takes precautions, effectively resolves legal risks, and creates a good internal and external environment for the healthy and orderly operation of power grid enterprises.

5 Conclusions

The construction of a power grid is related to people's livelihood and should be paid attention and supported by the whole society. We should seriously deal with the compensation problems related to land expropriation and occupation in power grid construction and how to avoid legal risks, to create a good environment for power grid construction.

References

1. Xuesheng, W.: Research and discussion on legal risk prevention of power grid construction. Power Enterp. Manag. **13**(1), 27–31 (2015)
2. Ying, G.: Study on legal risk management of power grid construction projects. North China Electr. Power Univ. **15**(01), 12–13 (2014)
3. Qi, W.: Causes of land expropriation dispute and prevention of legal risks. Enterp. Manag. **30**(4), 134–135 (2017)
4. Xu, Ch.: Study on the construction of the risk management system of power grid project. South China Univ. Technol. **13**(02), 11–12 (2016)
5. Xia, L.: Legal risk prevention of power grid construction land. China Electr. Power Enterp. Manag. **12**(7), 68–70 (2014)
6. Xiaolin, G.: Land expropriation and rural conflicts in China. China Q. **166**(10), 422 (2015)
7. Sandberg, H.: Land Title settlement and expropriation. J. Isr. Hist. **23**(2), 216–231 (2016)
8. George, K.: Managing expropriation risks in the energy sector: steps for foreign investors to minimise their exposure and maximise prospects for recovery when takings occur. J. Energy Nat. Resour. Law **23**(1), 36–59 (2015)
9. Chaves, A., Terry, Bahill A.: Comparison of risk analysis approaches and a case study of the risk of incorporating solar photovoltaic systems into a commercial electric power grid. Syst. Eng. **17**(1), 89–111 (2014)
10. Javanbakht, P., Mohagheghi, S.: A risk-averse security-constrained optimal power flow for a power grid subject to hurricanes. Electr. Power Syst. Res. **116**(10), 408–418 (2014)
11. Chen, T., Tang, Z.: Speaker recognition of noisy short utterance based on speech frame quality discrimination and three-stage classification model. Int. J. Control Autom. **8**(3), 135–146 (2015)
12. Dutta, S., Manideep, B.C., Basha, S.M., Caytiles, R.D., Iyengar, N., Ch, S.N.: Classification of diabetic retinopathy images by using deep learning models. Int. J. Grid Distrib. Comput. **11**(1), 99–106 (2018)
13. Tomar, A.S., Kumar, S., Chaurasiya, S.K., Shrivastava, V.: Novel approach for E-passport authentication using Elgamal cryptosystem. Int. J. Secur. Appl. **12**(2), 19–28 (2018)
14. Alamin, S.M., Khalifa, Othman, O.: Developing a robust framework to reduce the size of a recorded video surveillance systems. Int. J. Fut. Gen. Commun. Netw. **11**(1), 37–46 (2018)
15. Fu, Q., Xu, S., Wu, A.: Moving object detection with high precision via marked watershed and image morphology. Int. J. Signal Process. Image Process. Pattern Recogn. **9**(11), 71–82 (2016)
16. Mandal, A.K., Kahar, M.N.M.: Variant of LSB steganography algorithm for hiding information in RGB images. Int. J. Signal Process. Image Process. Pattern Recogn. **10**(1), 35–48 (2017)
17. Li, X., Ma, J.: Retail channels choice under competition: platform-sell or traditional resell. Int. J. u- and e- Serv. Sci. Technol. **11**(1), 1–16 (2018)

18. Harwood, L.C., Doerzaph, Z.R.: Visualization of the effect of topography on connected vehicle communications using LiDAR-derived models and interactive mapping techniques. Int. J. Transp. **6**(1), 15–28 (2018)
19. Faruk, A.G.M.: Systematic literature review: disaster recovery and business continuity plan for core banking solution (CBS) project. Int. J. Disaster Recovery Bus. Contin. **8**, 29–44 (2017)
20. Dhakar, J.S., Parmar, M.: A survey on grid based data aggregation in wireless sensor networks. Int. J. Sens. Appl. Control Syst. **4**(1), 11–18 (2016)

Author Index

A. E. Hassanien et al. (eds.), *Advanced Machine Learning Technologies and Applications*, Advances in Intelligent Systems and Computing 1141,
https://doi.org/10.1007/978-981-15-3383-9