

CONVERGENCE

INTERNATIONAL JOURNAL OF ICTAK

A Multidisciplinary Journal of Engineering, Technology & Employability



**INFORMATION &
COMMUNICATION TECHNOLOGY
ACADEMY OF KERALA**
ICTAK

A Govt. of India supported, Govt. of Kerala partnered Social Enterprise

VOLUME 06

ISSUE 01

DEC 2021



Culture eats strategy for breakfast...

~ Peter Drucker

Editorial Board

Chief Patron

Dr. Sandhta Chintala
Vice President, NASSCOM &
Executive Director, Sector Skills Council, NASSCOM

Chief Editor

Dr. Manoj A.S.
Head - Program, Govt. Operating Unit
ICT Academy of Kerala

Editorial Advisory Board

Dr. Rajasree M.S.
Vice Chancellor
APJ Abdul Kalam Technological University
Thiruvananthapuram

Prof. Mithileshwar Jha
Professor - Marketing (Retired)
Indian Institute of Management, Bengaluru

Dr. A. Damodaran
Chairperson, Economics & Social Science Area
Chair Professor IPR Chair on IP Management (MHRD)
Indian Institute of Management, Bengaluru

Prof. L.S. Murthy
Professor - Production & Operations Management
Indian Institute of Management, Bengaluru

Dr. Prahlad Vadakkepat
Associate Professor
Electrical & Computer Engineering
National University of Singapore (NUS)

Contents

1. A Study of Cyber Security and Challenges in the Distributed Data Environment	5
2. Women Empowerment: EX.ON.ER.ATE (To be free from the weight of your burdens.), A Handful for Women	13
3. Blockchain and Internet of Things in Education Field	23
4. Indirect Free - Space Optical Communications for Vehicular Networks	33
5. Monitoring Methods in Online Health Science Education- A KUHS Experience	39
6. A Hybrid Approach for Detection of Pneumonia from X-Ray Images	43
7. Multilingual Social Media Cyberbullying Detection	51
8. Flood Prediction and Analysis using Machine learning	57
9. A Comparative Review of Various Approaches for Malayalam Text Summarization	63

Introduction

“Nothing stinks like a pile of unpublished writing.” - Sylvia Plath

ICT Academy of Kerala is euphoric to present before our sagacious writers, scholars, academicians, industry professional, students and other stakeholders about the 6th edition of our revered International Journal - “Convergence” - A multidisciplinary Journal which aims to provide trailblazing ideas on Engineering, Technology & Employability were faculty members, research scholars, industry professionals can bestow upon their ingenious and prolific ideas in the form of articles, empirical studies, case studies and book reviews.

Most of the articles included in this edition of the journal were preponderantly from the scholars who have presented their papers in the ICTAK International Conclave on Skills, Engineering and Technology under the marque ‘ICSET 2021’ on the theme ‘Leading the New Normal’. The conclave stands out as a platform to explore the multiple facets of emerging technologies and recent trends. The conferee is intended to understand the speed of change happening in Engineering, Technology, Lifestyle, Innovation & Entrepreneurship and also in Business in New Normal and also to understand the new Normal Challenges and Opportunities in various domains. The conference also tried to present a birds eye view to understand the changes happening in the field of work, education, lifestyle and business. ICSET 2021 has presented a holistic picture about “the transformation of entire systems of production, management, and governance” in New Normal. ICSET finally helped to groom professionals and technocrats in tune with the New Normal Opportunities and prospects.

Based on the above theme surfeit of discussions and contemplations happened with regard to the conference theme in the form of speaker sessions, panel discussions by industry experts, scholars and paper presentation by academicians and research scholars on the conference theme. The selected papers were included in this edition of our Journal - “Convergence”, which will give a candid understanding with regard to the subject matter of the conference.

We wish you a fecundating and very illuminating reading experience!

Dr. Manoj A.S.

Chief Editor - Convergence

A Study of Cyber Security and Challenges in the Distributed Data Environment

Mr. Jibin N.¹, Ms. Nayana Habeeb²

1. Assistant Professor, Department of Computer Applications, MES AIMAT, Marampally - 683105 | jibintcr@gmail.com
2. Assistant Professor, Department of Computer Applications, MES AIMAT, Marampally - 683105 | nayanaaharis2@gmail.com

Abstract

Securing the data in the cyber world is one of the biggest challenges in today's world. Cyber-crimes are the first thing that comes to our mind when we think about the cyber security. The cyber threats are increasing day by day as the number of persons in the internet is increasing largely in this pandemic situation. So Cyber security is one of the most important areas that everyone in cyber space should consider. In this research paper, the authors discuss about the meaning and importance of cyber security, various types of attacks in cyber security, the relevance of critical thinking in the field of cyber security. This paper also explains the various aspects in the cyber security and the factors that affect the cyber security domain. This research work discusses about developing a novel model for digital forensic analysis of heterogeneous big data and it will help investigator to identify the threat. The proposed model is composed of four steps: application analysis, operating system analysis, transaction analysis and forming an opinion.

Keywords: *cyber security, cybercrime, cyber ethics, social media, cloud computing, critical thinking*

1. Introduction to Cyber Security

According to NIST, Information Security is the act of providing confidentiality, integrity and availability and protecting unauthorized access to information assets. The principles confidentiality, integrity and availability are known as CIA Triads. In ^[1] Nan Sun and Jun Zhang, 2019 discussed about Confidentiality ensures privacy, it ensures data is accessible by authorized persons only and it is achieved by employing data encryption method. Ensuring data consistency and accuracy over

its entire life cycle is known as the integrity of data. In ^[2] Schatz, 2019 explained what happened in the cyber incident. During the transit of data through local network it should not be changed, and steps should be taken to ensure that the data is not altered by any unauthorized person or subject makes any changes to our data. Any unauthorized changes to data can be identified by using hash value for the data. The hash value of original data and downloaded data should be same, which ensure that integrity of data

Availability principle ensure that the data is available 24 by 7, 365 days. Business continuity plans, disaster recovery, redundancy techniques are used for ensuring data availability.



Fig 1.0: *CIA Triad*

2. Related Work

In [3] Schatz et.al 2019 discussed about the Vulnerability, it can be defined as something that can be exploited, like flaw, error loop-hole or oversight, to violate the system security policy. An event, natural or man-made, that able to cause negative impact to an organization is called Threat. Exploit means a defined way to breach into security of an IT system through the vulnerability. And finally the probability of an event or that event could actually happen is called risk.

2.1 Security threats

In [4] T. W. Edgar and D. O. Manz, 2017 explained that security threat can be classified into two main categories, human factors and natural factors. The natural factors include lighting, environment disasters, floods etc. are important to consider, for ensuring business continuity plan. The human factors can be classified to internal and external factors. The threats former employees and current employees come under internal factors, and these threats are very important since they have access to internal resources, and they also know how the organization works and how internals are executed. In [6] S. A. Elnagdy, M. Qiu and K. Gai, 2016, explained that the threats from malicious software, hackers or crackers are included in external factors. External threats are happened by exploiting the vulnerabilities. They try to find a way to get into the organization and make damages to the IT assets. All these attacks are called human factor attacks because they either interact with human or developed by humans.

2.2 Vulnerability assessment

In [7] S. Ziegler explained the process of identifying, analysing, ranking and classifying the vulnerabilities are known as the vulnerability assessment. The vulnerabilities are classified according to specific criteria like whether the vulnerabilities can be exploitable or not, null vulnerability and many other parameters are also considered. The assessment will identify such threats and report to the admin, so that admin can take necessary action to rectify the problem.

2.3 Security attack definition

In [8] the K. Davis et.al, 2018, explained that in real scenario the cyber security architecture

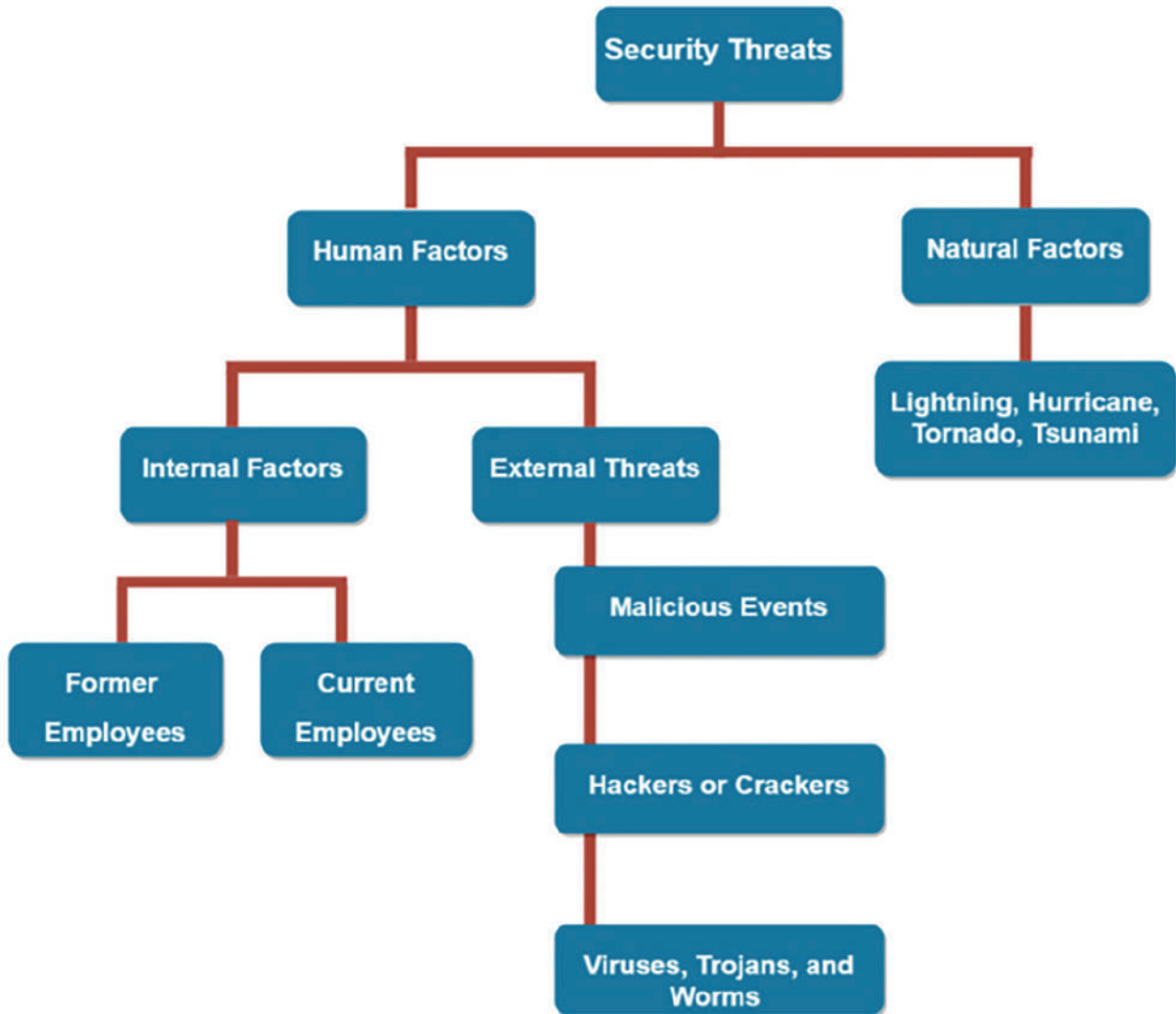


Fig 2.0: *Security Threats*

is very complex to implement since a huge volume of heterogeneous data are updated every minute in the cyber world and we have a lots of connected devices which makes the system highly complex. We need to protect. Not only we need to protect servers but also we need to protect our computer, our documents , our tablet, our smartphones smart watches and all the connected devices that

we carry, the information we share, should be protected. In [9] A. Tsirtsis et.al, 2018, discussed that there are lots of latest technology trends are emerging we need to know those in order to implement the security architecture. When we implement a security program we need to evaluate, identify, understand the risk and the threats that we are dealing with also it is required to assess

monitor and control those risks and threats. The term information security of IT security can be defined as the practise of defending computer servers, mobile devices, electronic systems, networks and most importantly the data from malicious attacks. The security system that we implement should be always correct and the attackers need to be correct only once.

3. Beyond Technology: Critical Thinking in Cybersecurity

In [10] A. Gawanmeh and A. Alomari, 2019 discussed about critical thinking in cyber security. In Cybersecurity the critical thinking is very important part. When we discuss about Cybersecurity the immediate thought is its very technical field like operating systems, networking, and the very technical things only. Critical thinking is organized, focused thinking towards the goal. Critical thinking should be applied in the cyber security field because of the constantly changing fast passed environment. The critical thinking skill strength us to think and act in situations where there are no clear answers, and there are no specific procedures in place. This helps to make good decisions.

In [11] Diogenes & Ozkaya, 2018, discussed said that we need to attain some key skills for critical thinking they are challenge assumptions, consider alternatives, evaluate data, identify key drivers, understand context. The assumptions that we made are usually based on our past experience, thoughts, evidence or personality. So it is required to challenge the assumptions, gather more data in a systematic, disciplined manner. The second skill you need to have consider the different alternative explanations by answering the 6W's, who/what/when/where/why/how, this can

be achieved by conducting brainstorm sessions. The next key skill is to evaluate the data and analyse data against multiple hypothesis and to see how well it fits. The next step is to identify and analyse the key drivers that affect the data, some external and internal drivers the affect the data are technology, regulatory authorities, government, the employees etc. The next key skill is to understand the context, which is the operational environment or the environment where you

4. Types of Actors in Cyber Crime

In [11] Diogenes & Ozkaya, 2018, also discussed about the types of actors can be classified into four. The first among them is hackers; they will try to hack into company's security system and try to extract the database

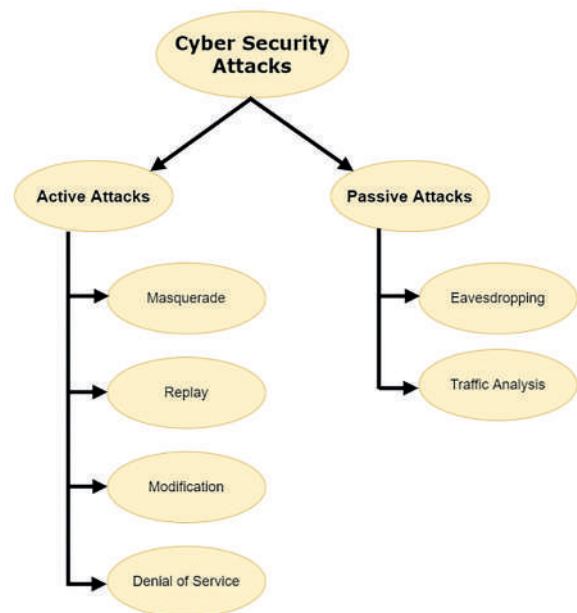


Fig 3.0: *Cyber Security Attacks*

to obtain the intellectual properties. Hackers can be paid or not. The second category is

internal users, these could be intentional or not. This category is very important since they are internal to the company and they know the internal workings of the company. Proper security policies should be imposed in order to avoid the threat from this category. Hacktivism is the next category, its hacking as the form to send a message through their activities to gain visibility for a cause. The last category is Governments and obviously their intention is not monitory but they want to monitor what happening in the cyber are and they want to spy. So the motives of the actors can be, just play, demonstrate their capacity, to gain money, Hacktivism, protest etc. [10].

5. Classification of Security Attack

In [12] Joe Haggerman, 2018 discussed about Security attacks can be classified mainly into two types; active attacks and passive attacks. Eavesdropping is one of the main categories in passive attacks, here the intruder or attacker sees the message from and understands the message. The second category of passive attack is traffic analysis. It is harder to detect the passive attacks since the message is unaltered. The confidentiality is violated in this type attacks. Active attacks on the other hand changes the original message and creation of false story and it falls into four different categories they are Masquerade, replay, modification of the messages, and denial of service. In [13] P. Bednar, 2019, discussed about masquerade, the original message masked from the receiver. In reply attack, intruder reads the message and sends a false message to the receiver. The third category is modification of messages, here the messages is modified and send to receiver. The last category is denial of service, which denies services to users. It is very difficult to defend attacks

globally because of diversity of attacks.

5.1 Security services

In [14] William Stallings. 2007 explained about the security services are employed to protect specific kind attacks in the system. The security services implemented using security mechanisms. Some examples are X.800 security service ensures adequate security to data transfers. The X.800 style of security ensures surety by providing authentication, Data confidentiality; data Integrity and availability, RFC 2828 provide specific kind of protection to system resources. Access management techniques are also employed. The access management ensure the effective access restrictions to organizations data. The security team could restrict access by groups by time frame or by specific dates, or by location that ensures better security.

6. OWASP Project

The non-profit organization Open Web Application Security Project dedicated to web application security. Their best-known project is OWASP Top 10. The core principle of OWASP is that anyone can freely access the Materials, which include the documentation tools, videos and forums, on their website and to improve the web application security. OWASP Top 10 is regularly updated report that explains 10 most critical security concerns for web application security. The report is put together by security expert from all over the world. It's an awareness document and they recommend all companies to integrate the report into their process to reduce risks.

6.1 Common security threats

In [16] Abbott et.al, 2018, Computer security threats are uncompromisingly increasing.

The attackers are constantly evolved to find new means to annoy steal and harm. Some common threats are discussed here. Malicious program that duplicates itself and infects the files and programs of the system are called virus. This is one of the most important category in the cyber threats. The we have malwares the name derived from malicious software, which causes serious damage to the system. Once the files are infected with malware, it can damage your files. Unwanted messages that causes security threats, are called spam. Phishing, botnets, spywares, ransom wares are examples of other threats. These threats should be properly managed.

7. Firewalls

Firewalls are a protection mechanism that allows some packets to pass and others to block from larger internet and helps to protect from attacks. Mainly there are two particular types of attacks SYN flooding at TCP connections attacks and preserve things so busy that there's no resource left in real connection. Apart from this there is violation of security policy attack which denies the service and modifies data illegally. The firewall allows only authorized access to the network. Firewalls can be classified as application level, packet filtering and XML firewalls. In packet filtering firewalls the decisions are made based on a couple of parameters on the packet, whether to forward the packet or not. From where the packet is coming from and 'where to' the packet is going to are those parameters. Application level firewalls filters packets on application data as well as on IP/TCP/UPD fields, We have another category XML Gateway which allows XML traffic passes through a conventional firewall without inspection across all normal web ports[17].

8. Antivirus/Antimalware

Antivirus and Antimalware are specialized programs that will detect, prevent and destroy the computer virus or malware. This will scan the system and search for malicious definitions of programs and based on the scan it will take actions. The viruses definitions will be constantly get updated by the vendors so that new virus threats can be easily identified. There many antivirus and antimalware software solutions are available [18].

9. Frameworks and Best Practices for Cyber Security

In order to prevent these types of attacks we need to have frameworks baseline and to follow some best practices, these are called effective Cybersecurity strategy. COBIT is good example of cyber security framework. COBIT is the abbreviation for Control Objectives for Information and Related Technology created by ISACA (Information Systems Audit and Control Association) for IT governance and management. Information Technology Infrastructure Library, ITIL is another framework with a set of best practices for delivering efficient security services. If you are not following these baselines and best practices it will not necessarily harm your business but if you follow surely it will add some additional security to your business. Cyber criminals use more sophisticated tools and techniques for cyber-attacks, this is one of the main reasons for developing various frameworks to achieve robust cyber security programs. The framework contains processes, practices, and technologies can use to secure network and computer systems. SO IEC 27001/ISO 27002, NIST Cybersecurity Framework, IASME Governance, NIST 800-53 etc. are some examples of cyber security frameworks. [19]

10. Penetration Testing

In [19] Reed, T., Abbott, R, et.al discussed about Penetration Testing or Pen testing is practice of testing a computer system, network applications, and software applications to find any security vulnerabilities that an attacker could use. The objective of this testing is to find the security weakness before the attackers can identify them and exploit them. There will be legal agreements at different levels before conducting the penetration testing [19].

11. Digital Forensics

Digital forensics or computer forensics is a branch of forensics science that includes identification, recovery, investigation, validation and presentation of facts regarding a cybercrime. The committer of a crime will bring something to the crime scene and leave something from it. And these are use as forensics evidence. When someone commits crime he will bring something into the scene and he will leave something and these two can be used as evidence. We have hardware tools as well as software tools for digital forensics that help to effectively analyse the crime scene.

12. Tiered Approach for Cyber Forensics in the Distributed Data Environment

Ten percent of the main problem in the current scenario is the legal compliance such as jurisdiction of the data servers located at various continents. The legal formality to enable the review process itself is different in

various countries. This research paper proposes to develop a novel model for a computer forensic investigation of mottled big data, and it will help the examiner to identify the threat. Application analysis, the operating systems analysis, transaction analysis and finally forming an opinion are the major four phases of the proposed system process. The proposed model is independent of operating system and the application. A stamp of all the transactions in the system is kept and it contains the details of all the activities in that transaction. This stamp of transaction can be further can be analysed in case of any cybercrime is occurred. The first phase of the model is application analysis, will analyse the application such as browser data and we will get an initial understanding about the event. The next we will perform operating system this will help to predict the cyber incident. The transaction management will analyse what is actually happened in the with the system by the transaction. An opinion is formed application, operating system and the transaction. The stamp of operations can be used as an evidence of cybercrime that happened.

13. Conclusion

In this highly connected world the cyber security is very important area to consider. Due to the COVID 19 pandemic the number of online users has increased enormously and day by day the relevance of cyber security is increasing. In this research paper the authors



Fig 4.0: Tiered Approach for Cyber Forensics

discussed about various cyber security concerns that should be considered. The paper discussed about various cyber-attacks and the classification and motives of the actors involved in the crime cyber activities. The paper discusses about various defence mechanisms in the cyber domain like firewalls, antivirus antimalware etc. There is no perfect solution for all cyber-attacks because of the diversity of attacks. The cyber forensics techniques are used to analyse once the crime is happened. Finally the paper proposes a cyber-forensic model.

14. References

1. Nan Sun, Jun Zhang, Senior Member, "Data Driven Cyber Security Incident Prediction: A Survey", IEEE Communications Survey & Tutorial, Volume: 21, Issue: 2, Second quarter 2019.
2. Schatz, Daniel; Bashroush, Rabih; Wall, Julie (2017). "Towards a More Representative Definition of Cyber Security". Journal of Digital Forensics, Security and Law. 12 (2) ,2019. ISSN 1558-7215.
3. Stevens, Tim (11 June 2018). "Global Cybersecurity: New Directions in Theory and Methods" (PDF). Politics and Governance. 6 (2): 1–4. Doi:10.17645/pag.v6i2.1569.
4. T. W. Edgar and D. O. Manz, "Research methods for cyber security", Syngress, 2017.
5. G. Kessler, J. Craiger and J. Haass, "A Taxonomy Framework for Maritime Cybersecurity: A Demonstration Using the Automatic Identification System", TransNav Int. J. Mar. Navig. Saf. Sea Transp., pp. 429-437, Sep. 2018.
6. S. A. Elnagdy, M. Qiu and K. Gai, "Understanding Taxonomy of Cyber Risks for Cybersecurity Insurance of Financial Industry in Cloud Computing", 2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud), pp. 295-300, 2016.
7. S. Ziegler, "Internet of Things Cybersecurity Paradigm Shift Threat Matrix and Practical Taxonomy" in Internet of Things Security and Data Protection, Cham: Springer International Publishing, pp. 1-7, 2019.
8. K. Davis, Y. Levy and B. Delak, "Towards a Development of Cybersecurity Risk-Responsibility Taxonomy of Small Enterprises for Data Breach Risk Mitigation", AMCIS 2018 Proc., Aug. 2018.
9. A. Tsirtsis, N. Tsapatsoulis, M. Stamatelatos, K. Papadamou and M. Sirivianos, "Cyber security risks for minors: A taxonomy and a software architecture", 2016 11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), pp. 93-99, 2016.
10. A. Gawanmeh and A. Alomari, "Taxonomy Analysis of Security Aspects in Cyber Physical Systems Applications", 2018 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 1-6, 2018.
11. Diogenes & Ozkaya (2018). Cybersecurity--Attack and Defense Strategies. Packt Publishing. ISBN 978-1-78847-529-7.
12. Joe Haggerman, Michael Mylrea, Sri Nikhil Gupta Gouriseti, Andrew Nicholls, "Buildings Cyber security Framework", PNNL-EERE Working Draft (Forthcoming) 2017
13. P. Bednar, A contextual integration of individual and organizational learning perspectives as part of IS analysis, Informing Science 3 (3) (2000) 145–156
14. William Stallings, Lawrie Brown Computer Security: Principles and Practice August 2007, SBN:978-0-13-600424-0
15. D'Amico, A., Whitley, K., Tesone, D., O'Brien, B., & Roth, E. (2005). Achieving cyber defence situational awareness: A cognitive task analysis of information assurance analysts. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, SAGE Publications, 229-233
16. Abbott, R.G., McClain, J., Anderson, B., Nauer, K., Silva, A. & Forsythe, C. (2015)., Log analysis of cyber security training exercises. Proceedings of the Applied Human Factors and Ergonomics Conference, Las Vegas, NV.
17. E.A. Lee, Cyber physical systems: Design challenges, in: 2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC), May2008, pp.363–369.
18. Reed, T., Abbott, R., Anderson, B., Nauer, K. & Forsythe, C. (2014). Simulation of workflow and threat characteristics for cyber security incident response teams. Proceedings of the 2014 International Annual Meeting of the Human Factors and Ergonomics Society, Chicago, IL
19. Stevens-Adams, S., Carbajal, A., Silva, A., Nauer, K., Anderson, B., Reed, T. & Forsythe, C. (2013). Enhanced training for cyber situational awareness, Proceedings of the Human-Computer Interaction International Conference, Las Vegas, NV

Women Empowerment: EX.ON.ER.ATE (To be free from the weight of your burdens.), A Handful for Women

Afnita K.A.¹

1. Master of Computer Applications (MCA), First Year | M.E.S Advanced Institute of Management and Technologies | Afnithaafni25@gmail.com

Abstract

The day India gained independence in 1947 was a time when both men and women gained independence. Whether or not women actually have freedom? NO, because they were discriminated against on the basis of gender discrimination, it is still a problem facing women in this century. However, women face inequality on Education, Employment, Freedom to make their own decisions, Control over the cost of wages, Freedom of dress, and even children face this inequality too from their homes. All these are detrimental to their progress and uplift. In India women are considered impure, inferior and servants to serve husband, children and family. People here believe that her dreams of flying should not be fulfilled and that they are all detrimental to the culture. Women face more difficulties in rural areas than in urban areas. Lack of education, following cultures, superstitions, and fear of society are some reasons for the decline in women's empowerment. We need more women empowerment to make every woman in country aware of their rights, powers and security to be free. The feminization we women want must come more than the current feminization, because many of our sisters are victim of mental, sexual harassments.

This paper analyse the problems faced by women in society today and even in their own homes, solutions to help them and schemas implemented by Government of India for women.

Keywords: *women empowerment, overcome women problems, support women*

1. Introduction

The ancients considered women to be a goddess and a gift from God. But today, society considered women as second class . Even though India is a developing country , society has not changed attitude towards women. From unborn child to the elderly

women are abused. Worldwide 27% of women aged 15-49 years who have been subjected to some form of physical/or sexual violence from their partners.[10] From a report about 53 % children reported experiencing different kinds of abuse like sexual abuse, inappropriate touching, and forced to nude

photography. [13] [11] Between 2005 and 2012, 55 % are homicide victims were killed by family members.

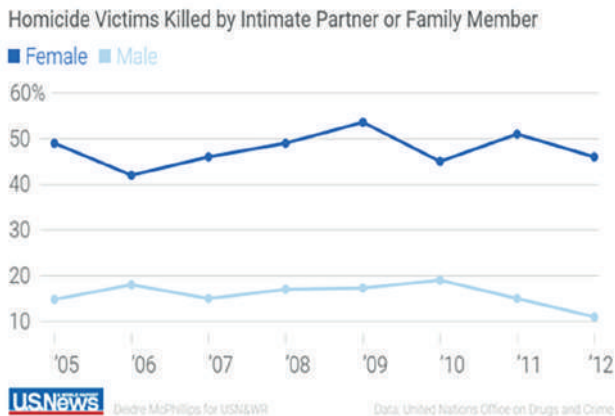


Fig 1.0: *Homicide Victims Killed by Intimate Partner or Family Members*

Who is responsible for this? Of course, it's our fault, we must first change the perspective of each of us. Women empowerment must first change the perspective of each women. Understand that you are not a slave to anyone and that you have the same abilities and freedom as men. With the empowerment of women, many women have come to the forefront of society by overcoming their shortcomings in recognizing their potential. Historically, we have many predecessors who have transcended gender equality, overcame protests and obstacles and worked hard for their rights and aspirations. Mother Teresa, Kalpana Chawla, Anjali Gupta all these are examples for us.

In a way, women's empowerment is a human right or women's empowerment is related to human rights. It includes the right of women to make their own decisions and to influence social change for others and for themselves.

We can create a peaceful atmosphere by changing the way society views on women and it will bring peace to the whole world. Feminist empowerment in the West has also been associated with movements throughout history. The right to vote in the early twentieth century began with the sexual revolution of the 1960s and feminism in the 1990s. [9] Protecting women or enforcing justice for them is an issue that needs to be discussed globally. That is why the whole world is becoming aware of and working for women's empowerment and its need. International Women's Empowerment Day plays a role in this.

UN Global Compact and UN Women created 7 principles for women. [15]

1. Establish high-level corporate leadership for gender equality.
2. Treat all women and men fairly at work-respect and support human rights and non-discrimination.
3. Ensure the health, safety and well-being of all women and men workers.
4. Promote education, training and professional development for women.
5. Implement enterprise development, supply chain and marketing practices that empower women.
6. Promote equality through community initiative and advocacy.
7. Measure and publicly report on progress to achieve gender equality.

Women face all stages of life from childhood to old age with great difficulty and suffering. They face everything from torture to murder to suicide. All this is due to the perception that she is a women, she does not speak out in front of society and she is a slave to men.

Currently the worst situation for women to live

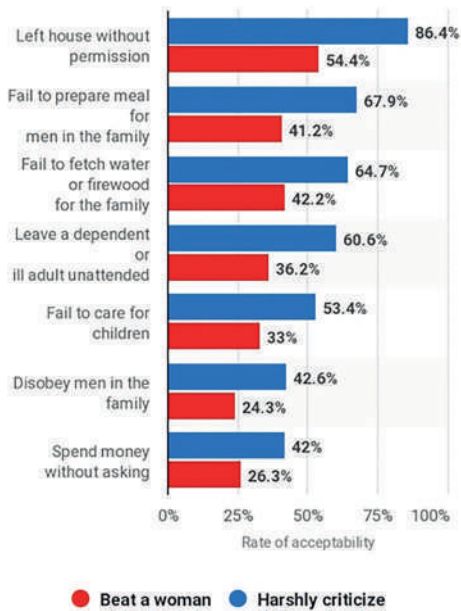


Fig 2.0: *Attitude towards acceptability of violence against women in India in 2018*

in is in India. The worst record of sexual violence, harassment from cultural and traditional practices, and human trafficking. In other words, India has become the worst country in the world for protection of women. These are the findings of a global perception poll conducted by the Thomson Reuters Foundation.

Rank	2018	2011
1	India	Afghanistan
2	Afghanistan	Democratic Republic of Congo
3	Syria	Pakistan
4	Somalia	India
5	Saudi Arabia	Somalia

Fig 3.0: *Most Dangerous Countries for Women 2018, 2011*

A survey of 558 women experts was conducted in each country to assess women's safety. The bad situation here defeated India. India was fourth in 2011, and India ranked in the top five in every polls. [8]

There are 37 million more men than women in India, of which 27% of girls are married before the age of 18. This is the highest rate in the world. "India is still fighting the deep-rooted patriarchal mind set, which sees women as inferior in the world's biggest democracy" said by Monique Villa, Chief executive officer, Thomson Reuters. The death of eight-year-old ASIFA in Jammu and Kashmir alone was enough to understand the global sexual violence against women and children in India. The murder of that child who was gang-raped was a big issue globally. India responds by imposing death penalty on child rapists. This was believed by all to some extent to prevent sexual assault but none of us can forget the rape and murder of student on a bus in Delhi. It was two moments that shook the world's conscience. There are many more examples. These two examples show the world how safe women and children are in India. So it can be said that women's empowerment has little to do with women's security, in other words, women's empowerment needs to focus more on women's safety.

If a woman comes forward with courage, she can help others who are suffering in life, It helps them to stand upright in front of the society. Changes in society can only take place if we first need to change the attitudes of each and every woman. I can't agree with the view that girls are cheaper than boys, instead of pursuing education in school, she is often equipped to provide for her family. In a way, women are the ones cheating on women,

aren't they? The kitchen and the family are a women's world has been injected into a girl from her childhood. Every year 12 million girls get married before reaching the age of majority and they have no opinion of their own.

2. Problem Statements

2.1. From infants to adults are victims of physical, mental and sexual abuse.

Only a few have expressed a desire to let the outside world know of their difficulties. They rely on the police and medical law. There are lot of women who melt themselves into fear of being looked down upon by society, culture

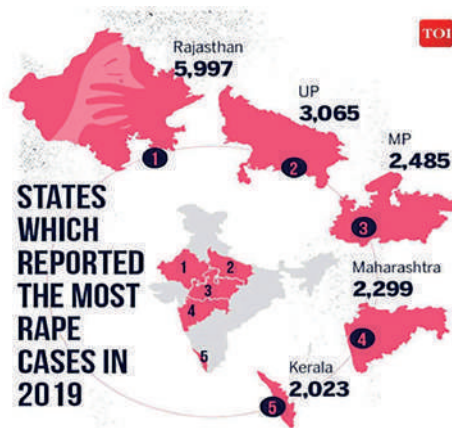


Fig 4.0: States which Reported the Most Rape Cases In 2019

, stress at home and blaming shame.

2.2. Every girl in her teens should know menstruation and sex.

In today's society children have no knowledge about sex and its necessity. They learn such things in many other wrong ways.

2.3. Family issues

From the survey of International Institute of Population Science-Mumbai among people in the age of 15 - 49, 30% women suffer physical assault by husband. 40% think man can beat up wife and 30% believe husband can assault wife, notice that only 12% of Kerala women travel alone.

2.4. Dowry system

Dowry is a ritual that has survived to this century. The dowry system is a system that has been discontinued as Sati has been discontinued in the past. India reports the highest total number of dowry deaths (8391) in 2010, that is 1.4 deaths per 100000 women. [14]

2.5. Incompatibilities in marriage

If life partners get to know and understand each other, it will be the perfect marriage.

2.6. Mental and physical health

Mental difficulties and delusions are much more common and different in women than in men. The mental health of the two is very different. Girls with physical difficulties, Girls who have been sexually abused, women who are married at a young age, marital difficulties, Health problems, Alcohol use, increases the risk of suicide or self-harm. The way society looks at mental health causes women to recognize their difficulties.

- Depression, anxiety and stress are 2-3 times more common in women than in men.
- The World Health Organization estimates that about 7.5% of the population is addicted to certain mental disorders, which is expected to affect 20% of Indians by the end of the year.
- 56 million Indians suffer from depression

and 38 million suffer from anxiety. [6]

2.7. Sexual offenses

Sexual violence in India has increased in recent times. In India, an average of 87% cases are reported daily in 2019. As of September 2020, 405861 crimes have been reported. This is an increase of 7% over 2018. According to the



Fig 5.0: Crimes against Women in India

National Crime Records Bureau, there were 378236 cases of crimes against women in the country in 2018. 32033 rape cases have been reported in 2019. [1]

2.8. Domestic violence

Domestic violence is a problem that many women around us experience. In 1983, the government introduced Section 498-A to prevent women from being abused by relatives or husbands. This category includes harassment of a woman for causing mental or physical harm to her body or forcing her

to commit illegal offenses [3]. Today, marital abuse and father and other relatives abusing girls are on the rise. In 2019, 32.25 lakh crimes (62.6%) were registered under the Indian Penal Code (IPS) [2]

- Cruelty committed by husband and relatives (30.9%). [11]
- Assault on women with intent to outrage her modesty (21.8%) [1]
- Kidnapping and abduction of women (17.9%). [1]
- In 2019, total 51.56 lakh crimes were registered in India. Compared to 2018, it has increased by about 1.6%. In 2018, 50.7 lakh crimes were registered. [2]

2.9. Domestic slavery

Unnecessary coercive imposition on women or someone to perform domestic services in an unacceptable situation is termed as Domestic violence. Its victims are at the feet of their owners. In the case of women, they take on all the responsibilities, from cooking, cleaning the house and environment and caring for children. They work for about 10 to 15 hours. If they go out and work during this time, they get a small salary. This is due to the misconception in the society that women are only for housekeeping and housing purposes. In most households, men are unwilling to help women with their household chores. Men are drawn back by the thought of what others will think.

2.10. Forced marriage and forced abortion.

Forced marriage may involve the physical, mental, sexual and emotional stress of some women. This is a kind of human rights violation. Indian laws give certain rights to the child or a woman and forced marriage

violates those rights. Forced marriage of children under 18 is in some ways against child rights. Causing them to be abused and physically and mentally exhausted. It also significantly affects the health of the child. And it can also be considered as rape. The education of girls through forced marriage also deprives them of their desires. Also abuse of domestic and sexual abuse is possible. Being alone in the rape without the consent of the girl child leads to unwanted pregnancy and abortion.

2.11. Unemployment on its own

The majority of women in India suffer from unemployment. The number of women in paid jobs has dropped significantly. Unemployment and lack of self-employment make

Nowadays, a woman is not safe inside and outside of the home, her body is her enemy. The saddest thing is that our country is not safe at all when it comes to women’s security. According to the National Crime Records Bureau, the risk of raping a woman or a girl has increased by 44%. Society blames women for going out at night, wear modern clothes and going to work at night. But what do little kids do wrong? They are not safe in society either. The patriarchal attitude is the number one reason for the increase in crime. This is because women are seen as second class and men give high quality and value. Failures in punitive procedures, men have a clear ignorance about sex and menstruation, all of this threatens the safety of women.

3. Solutions

What we women need first is a good community, a system that understands women’s issue and works for women. The women are the speaker and she is the listener. This community is made up of people who are at the top of society and at all levels of society. She Power-Exonerate is an online collaborative platform that provides exclusive services for women with an aim to empower them in our country. The online platform provide many services such as legal services, women health related services, and an online platform exclusively for homemade products, also various services to support women entrepreneurs. The platform provides an online platform to collaborate for talented successful women from all walks of life, doctors, psychiatric care experts, police officers, women entrepreneurs, lawyers, social workers etc. The platform also has to features such as instant messaging, video conferencing, chat room etc. Women must possess self-worth, confidence,

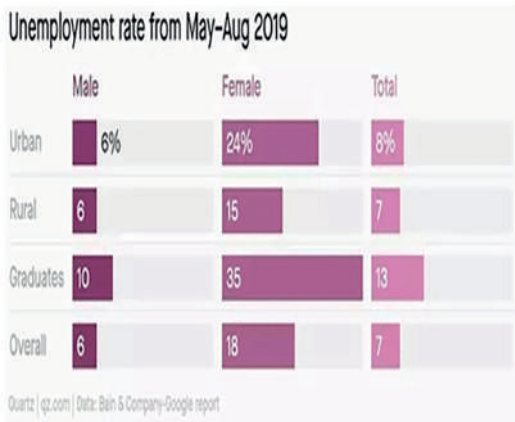


Fig 6.0: Unemployment Rate

women more vulnerable. She has to live as a slave to men because their needs are done by men. This condition is more common in married women who do not have a job or income, cause great hardship and marital problems for some.

2.12. Security problems

and freedom to acquire their needs and goals. Classification based on gender, is unreasonable and having no worth. They are paid less, expected to cook and are restricted by family members. To overcome these and to become independent, we need women Empowerment. "She power-Exonerate" is an online portal to empower women. This community will stand up for their safety and for the downtrodden in society. It will be a supportive hand for those who aspire for elevation. 24*7 this community must be available. People joining this community should be connected to each other. A lot of women or girls are victims of a lot of hardships. They may discuss their difficulties with others without anyone knowing. With the services available to us, we can reduce the mental and physical stress and many other difficulties of women to some extent. Women could address their problems and could find own solutions through this platform.

3.1. Online market for homemade products (buy, sell, includes small scale and large-scale entrepreneurs).

This is a service that can help a lot of women a platform for sell their products. Items made by women as a community of their own like paper bag, mask, food items, small initiatives through kudumbasree, crafts, paintings, bottle arts etc. It also helps youtubers to sell their products also. A lot of women make a living by selling this. This will also help the seamstress. Users or sellers can join when they join this community. In addition, both will receive other services in this community.

3.2. To find jobs and provide new job opportunities.

A lot women have given up their desires and dreams for family. But she wants to stand on her on two feet. This community is paving the way for that too. There is a lot of success entrepreneurs in this community, through this community they can find workers to reduce their workload or a women can apply for job based on her educational qualification and work experience. It is also very useful for small entrepreneurs, because it is useful to get large orders for bakers, farmers, tailors etc. They can also share about job vacancies.

3.3. Provide legal help (right advise, action plan and awareness)

A lot of women and girls face a lot of mental and physical hardships from their home and society. This service will be very useful for them. Because we can point our difficulties directly to the police and other authorities without the help and knowledge of family and others. Today there are facilities like women' cell but most women are afraid to use them or they will be the ones who find it difficult to get to the station. Through this community we get the help of women 'police and women's cell. It is equally useful for those who experience abuse in their husband's home. Every woman can learn about the legal aspects of the law. Legal aid is easily accessible to women of all legal backgrounds.

3.4. Support for entrepreneurs.

Even before the Covid period, many women had started self-employment. Many women have started small business to overcome the crisis of the Covid period and to secure the family security. This community helps to some extend to help them and meet their needs as needed to make the project a complete success. There are a lot of women in this

community who have proven their skills as self-employed. They help new entrepreneurs with their doubts, marketing of their products and making a profit through the venture. Educates new entrepreneurs about marketing and their fluctuations. This helps to bring women to the forefront of society to some extent and boost their confidence.

3.5. Live tracking and emergency message button(for safety)

Safety is one of the most important needs of women. This service will be useful for women night travel. The emergency message button is a way for women to pass information to the police or other authorities as direct when they are in trouble. Live tracking system means that police can track the correct location, and vehicle information, and other police officers and authorities track lively when women travel at night.

3.6. Health & nutrition.

The health, nutrition and diet are something that women give very little importance to. Women need to eat the foods that their bodies need for each period, it affects their mental and physical health. Through this community doctors provide advice on nutrition for women and their regulation.

3.7. Direct benefits (included other individual platforms like RMK (rashtriya mahila kosh) etc.

The Government of India has a lot of facilities and a lot of help services for help women and girls. But most of us women and girls have no idea about them.

- Rashtriya Mahila Kosh- make credit available for lower income women in

India-1993

- Mother and Child tracking system-monitor the health care system to ensure that all mothers and their children have access to a range of services-2009
- Rajiv Gandhi scheme for empowerment of adolescent girls-2012
- Pradhan Mantri Matritv Vandana Yojana-maternity benefits for pregnant and lactating women age 19 and over for their first two live births-2010.

The services provided by the Government for women and girls and its details reach out to women as directly through this community• Beti Bachao,Beti Padao -Save the girl child -22 January 2015. [22]

3.8. Mental health support

A lot of women around us are experiencing psychological distress. Mental illness is more prevalent in young children and women or some may find it difficult to talk openly about their mental problems. The inability to go to the hospital or the way society views mental illness may be the reason why most women do not give much importance to mental health. But they do not know that it will affect their physical condition in future. In this community female mental care specialist, psychologist, psychiatrist and counselling services are available. It provides this kind of assistance to the women who work on this committee, their difficulties can be adequately identified and medications can be taken. This community also facilitates direct doctor visits.

3.9. Gynecologists services, care of babies and their diets.

Gynecology is an area where a lot of women are reluctant. Their physical difficulties,

hormonal changes, sexual doubts, pregnancy doubts all are need a help of a gynecology. These specialists also provide information on things from early pregnancy to late pregnancy and diet.

As well as childcare this committee provides information on the diet of infants and the activities required for their growth. These includes the foods that babies need to eat at each stage of growth, their activities, weight and height.

4. Conclusion

Although women empowerment and its services are still in place, we women still need more women empowerment and its services. Even today, many women and children in society are victims of various forms of exploitation. Women in distress should be supported and brought to the forefront of society. They should be brought to the top of the society. When women lead a safe and productive life they can reach their full potential. Unemployment and the economic crisis in the country can be overcome by contributing their skills to the employment sector. When women live in happy and joyful conditions it gets to the children as well. They can fuel a sustainable economy and bring great benefits to society and humanity. It needs education. For a girl to succeed at the social level, education must increase. Women's empowerment includes economic growth and sustainable social transformation for the nation.

5. References

1. <https://thewire.in/women/average-87-rape-cases-daily-over-7-rise-in-crimes-against-women-in-2019-ncrb-data>.
2. <https://factly.in/data-number-of-registered-cognizable-crimes-increase-marginally-in-2019/>
3. <https://blog.ipleaders.in/understand-section-498a-domestic-violence/>
4. <https://qz.com/india/1805246/women-are-worst-hit-by-indias-unemployment-crisis/>
5. https://m.timesofindia.com/india/india-sees-88-rape-cases-a-day-but-conviction-rate-below-30/amp_articleshow/78526440.cms
6. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4539863/>
7. <https://wcd.nic.in/womendevelopment/national-policy-women-empowerment>
8. https://www.business-standard.com/article/current-affairs/femicide-rapes-culture-of-violence-make-india-the-worst-country-for-women-118062700112_1.html
9. <https://www.worldvision.com.au/womens-empowerment/>
10. <https://www.who.int/news-room/fact-sheets/detail/violence-against-women>
11. <https://journals.sagepub.com/doi/full/10.1177/0253717620951391>
12. <https://www.usnews.com/news/best-countries/articles/2016-10-20/violence-against-women-in-5-charts>
13. <https://m.timesofindia.com/india/over-53-children-face-sexual-abuse-survey/articleshow/1881344.cms>
14. https://en.m.wikipedia.org/wiki/Dowry_death
15. <https://www.unglobalcompact.org/take-action/action/womens-principles>
16. https://www.researchgate.net/publication/23795416_Women_Empowerment_in_India/link/5a0841beaca272ed279fa917/download
17. https://en.m.wikipedia.org/wiki/Women's_empowerment

18. <https://www.unfpa.org/pcm/node/9551>
19. <https://m.timesofindia.com/topic/Women-Empowerment/ampdefault>
20. <https://ieeexplore.ieee.org/document/7906859>
21. <https://ieeexplore.ieee.org/document/8285555>
22. <https://www.goodreturns.in/amphtml/classroom/2018/02/7-indian-government-schemes-women-empowerment-680804.html#referrer=https://www.google.com&csi=022>.

Blockchain and Internet of Things in Education Field

Sneha Alen Santhosh¹ , Sundev Nalees²

1. snehaalendec2000@gmail.com | 6238064602

Abstract

During the last couple of years the extension of the internet into the real world, also referred to as the Internet of Things (IoT), was positively affected by an ongoing digitalization. Furthermore, one of the most active IoT domains is the education system. However, this thesis proposes building of an innovative blockchain-based architecture across the Internet of Things (IoT) platform for the education system in order to enhance communication efficiency within the 5 G network. Wireless networking would have been the main research area allowing people to communicate without using the wires. It was established at the start of the Internet by retrieving the web pages to connect from one computer to another computer. Moreover, high-speed, intelligent, powerful networks with numerous contemporary technologies, such as low power consumption, and so on, appear to be available in today's world to connect among each other. One of the complex tasks throughout the area of mobile communications would be to design a new virtualization framework based on blockchain across the Internet of Things architecture. The goal of this research is to connect a new study for an educational system that contains blockchain to the internet of things or keeping things cryptographically secure on the internet. This research combines with its improved blockchain and IoT to create an efficient interaction system between students, teachers, employers, developers, facilitators and accreditors on the Internet. This specified framework is detailed research's great estimation.

Keywords: *blockchain, internet, education, internet of things*

1. Introduction and Motivation

For many decades the education sector has experienced continuous technological innovations such as Artificial Intelligence (AI), Virtual Reality (VR), Internet of Things, Blockchain, and wearable devices are being used to improve the education sector's ability.

Blockchain is the technology that appeared with the creation of Bitcoin, which is behind the other cryptocurrencies and allows its decentralization. The Blockchain development of "bitcoin" can also offer advantages in the educational field. Based on the analysis of a series of studies, the paper confirm that the

relationship of the “blockchain” with education is in an embryonic phase while offering a series of recommendations to encourage the development of this technology. The proposed study is a move forward in the area of IoT in 5G diverse systems in which the authors propose a unique blockchain-based IoT structure of interacting different peoples and connected devices in the 5G network. That study result would be to introduce a new structure of communications on the IoT. The suggested study utilizes the required study’s appropriate as well as effective simulation and could be introduced through an IoT structure. It seems the whole universe is now becoming completely reliant on mobility facilities as well as

wireless technology. The Blockchain throughout the IoT has become a novel innovation that behaves on a decentralized, distributed, public as well as a real-time database to collect operations among IoT endpoints . The blockchain is indeed a sequence of blocks where every block is connected to the prior blocks. Every block must have the cryptographically secure key, prior block hash, as well as its information. The Blockchain operations will be the fundamental modules that had to transmit information among IoT endpoints. The IoT access points seem to be different kinds of natural however smart devices with integrated detectors, sensors, systems as well as worthy of interacting with several other IoT

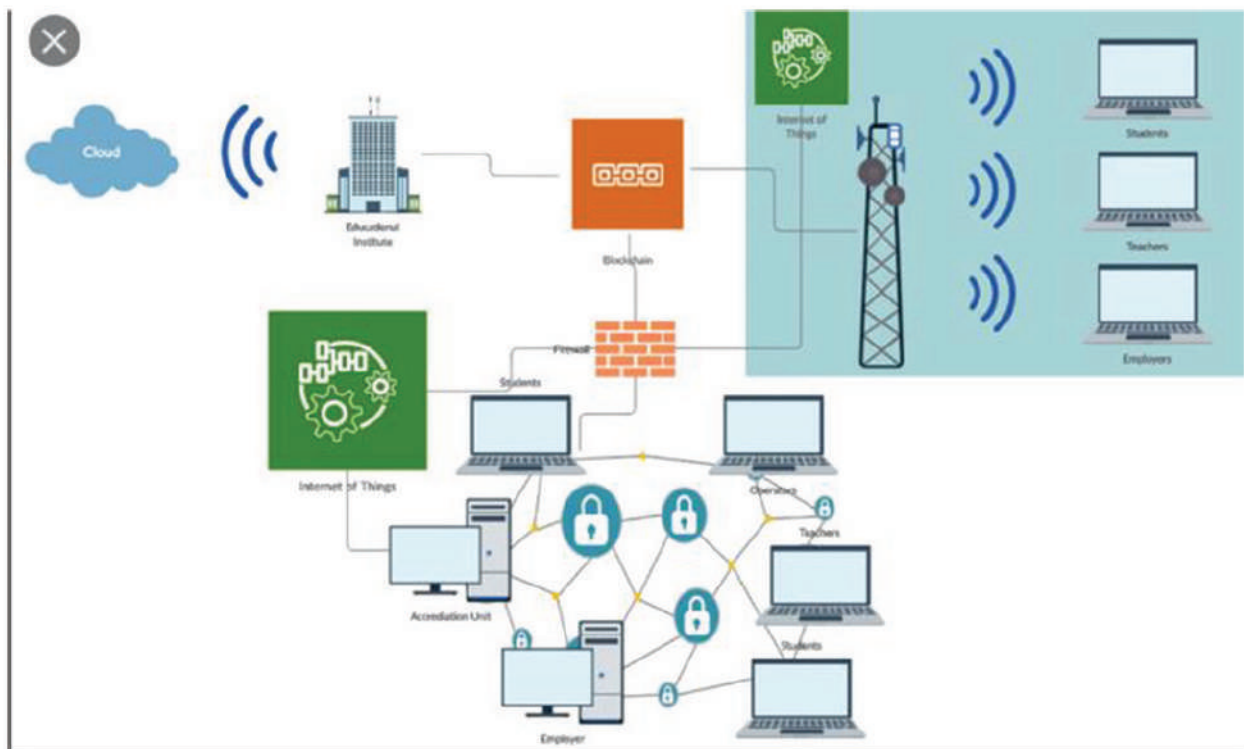


Fig 1.0: Blockchain-IoT Framework for the Educational System.

endpoints. Blockchain's role in IoT would be to have a mechanism for handling protected information records by IoT endpoints. BC seems to be a safe innovation that could be used openly as well as publicly. The Internet of things enables this technology to enable asymmetric cryptography among IoT endpoints in such a diverse system. Blockchain transactions might be monitored as well as traversed through everything accessed to interact throughout the IoT. BC might well enhance interaction protection. Exchange information from one specification to the next are using wireless communication begins in the form of the radio network packets. A machine was able to connect with another machine of the same specification.

Figure 1 represents the framework where the teachers, students, operators, employers, and accreditation unit members can communicate using the blockchain-IoT framework. Blockchain-IoT framework for the educational system is such a decentralized strategy in which the IoT users have been directly linked. It is therefore much more complicated to be using the conventional current security strategies in the interaction among IoT users. BC is an innovation that provides security in transactions among IoT users. This offers a decentralized, distributed as well as publicly available mutual ledger to collect blocks information which is stored or confirmed in such an IoT system. Its information stored throughout the distributed ledger is immediately attempted to use peer-to-peer configuration. The cryptographic blockchain technology is an innovation at which IoT endpoints handle the transactions in the type of such a block in the blockchain.

Blockchain in the education industry eliminate

the barriers in fee payments, report creation, connectivity with parents and students in different countries, etc. The following requirements are needed to have a blockchain in higher education:

1.1 Issuing credentials and academical certifications

Conventionally, through a big load of digital paperwork and a series of procedures, qualifying a student for graduation was done either manually or digitally. This added tremendous workload overhead to the administration to store, monitor and verify student's academic credentials. And at times, there is a risk of getting those stolen or destroyed too. With blockchain, certifying a student can be accomplished instantly. In the blockchain, the entire data gets stored and protected in a smart contract or distributed ledger according to your choice. Each of these blockchain-based solutions are known for its real-time updating of data and it also provides tamper-proofing. Thus, the certificates are digitally protected too.

1.2 Modern pedagogy

Every learner does have its learning skills. It is important to build new educational frameworks to meet each student's needs.

1.3 Publishing of study materials and academic content

Researchers, academic authors, publishers and educators can write their own unique content, without any worry of copyright issues and content ownership. They can even track data for validity and originality using the P2P network provided by blockchain. Top of it, with blockchain technology, authors, publishers

and researchers can publish their content on the global market instantly without having to go through the long process of editing, reviewing and approvals. However, blockchain provides its own reviewing team of whomsoever is interested to review and are a part of the P2P network. The notable fact here is that, the review process takes much lesser time with blockchain than the time taken to traditionally approve a publication. Blockchain in education and higher studies protects all content from digital piracy by providing proof of authenticity to each and every writer, author, publisher and researcher.

1.4 Student loans

Spending on education convinces learners to choose a profession of preference. Learner's incentive services have designed to recognize the talents of the learners.

1.5 Maintaining immutable security infrastructure and identity

With blockchain, education institution can add n number of cameras, sensors within the campus, vehicles to monitor, control and track environmental activity in real-time with instant updates at every hour. It helps to ensure physical safety of students, tutors, administrators and digital safety of educational equipment, assets and data.

Blockchain stores the identity of everyone with public or private key encryption depending upon your preference and this gives an increase to anonymity, privacy to the identity holder. The objective of this study is to develop a communication structure for the educational system and to provide a trustworthy, safe and quick link among connected users on the Internet of Things using

blockchain technology

2. Why Blockchain and IoT in Education?

Packets are transmitted in cloud using a block. Every node has a hash code and connection to exchange information. Every connection to exchange information contains a lot of blocks with its previous blocks hash code and data which are connected with security through cryptography techniques. Since blockchain nodes are like the linked list node in the data structure, blockchain is a complex data structure. The blocks are distributing in a decentralized system using the point-to-point topological network. When a new block is created then it moves to the network and visit every connected node and checks its authentication. If it is valid then it will connect to the blockchain and its hash will generate only once. This newly generated block stores the hash of the previous block and connects to the chain.

The performance of the proposed system is evaluated through different experiments. Firstly, Creation of thousands of blocks with a fixed size by using the open-source software (Node.js) have been done. Thereafter the IoT network created has been connected with cloud and, created fog and blockchain. This suggested structure depicted the use of Fog computing with IoT gadgets onto the edge of the network using blockchain technology to communicate, exchange and share information amongst the IoT endpoints in the educational sector. Transactions in the suggested structure were transferred in the point-to-point network topology. There are some unique IoT endpoints termed as Miners throughout the system. These are usually used to verify network transactions. Whether

the transactions are confirmed so they are transformed into the blocks or decided to add to the existing Blockchain or transmitted to the channel. Its miners play a vital role in adapting the creation of a new block in the blockchain. Throughout this study, we used many examinations to determine the structure. A hashing algorithm is introduced using the IROHA tool of Hyper ledger. A Docker-compose is built on the device. A Hyperledger IROHA tool involves several services like decentralized Hyperledger, Proof of Work (PoW) algorithms, P2P system, and so on. Sumeragi in Hyperledger IROHA methodology is introduced in Blockchain. An IROHA Android and iOS packages need the facility with the blockchain to communicate with the IoT endpoints. According to the Sumeragi algorithm, IoT nodes have requested the transactions or follows the basic steps:

Step 1: Transmitting: The members confirm, arrange or sign the transactions or send data to the system.

Step 2: Authentication as well as going to sign: This validates, requests or signs the transaction or broadcasts to the peer-to-peer channel's approved IoT endpoint.

Step 3: Committed: Dedicate regarding signing.

Throughout the incident of server failure, the automated system contributes a move named error detection. Also, the algorithm works with the existing server to monitor the inconsistencies. Consider the newly created block node that is referenced by two previously created blocks. The node is placed between these two blocks. Consider that the block is verified by the miners, the time is represented the verification time calculated by the Poisson process

. When a new block is created then it moves to the network and visit every connected node and check its authentication. If it is valid then it will connect to the blockchain and its hash will generate only once. This newly generated block stores the hash of the previous block and connects to the chain. Blockchains should impact the higher education system such as business schools that are more urgent or far-reaching than the other widely mentioned innovations like artificial intelligence and robotics. Although these innovations provide significant learning and convening ability, the blockchains could significantly by changing the way the school of business conducts its practice. This study is proposed to build up an educational system depending on blockchain and IoT. Building a trusted educational credit and grading framework could be of benefit to the blockchain technology. Proof of concept is introduced using Ark blockchain open-source framework. This platform suggested addresses learners and institutions from a worldwide homogeneous viewpoint. Learners benefit from either a single or transparent perspective of their finished classes, whereas the model has access to state-of-the-art information irrespective of the academic background of a learner. Many recipients of the proposed change are prospective employers, who are able to immediately verify the student information. This solution implemented is focused on decentralized peer-to-peer network mechanism. This really exchanges the higher education grading system to an effective, streamlined, pervasive model based on blockchain from current global-world physiological documents or conventional electronic documents. This is a process to ultimately adapt into a homogeneous, streamlined, and worldwide pervasive credit and grading system for higher

education.

3. Blockchain and IoT in Higher Education

Blockchains are beneficial if data needs to be shared by various parties without trust. A blockchain is a means of sending data in a completely automated and secure manner, as falsification is virtually impossible. A blockchain is a time-stamped sequence of an unchanging information record that is attempted by a cluster of computer systems not operated by any single user. Not conventional database blockchains are distributed, and everybody has to accept to make any changes to manage it. The biggest achievement behind this study has been building a structure for the educational system to communicate using blockchain technology and IoT on the Internet. This framework becomes primarily suitable for the users where information becomes regularly conveyed to the connected devices on the internet network. In addition, we used a retransmission strategy, variable packet length, or populated congestion situation to improve the proposed system. A plan of action for this study has been defined. For performance onto the Internet of things, the execution of the IoT-BC structure to interact safely between IoT devices will be programmed. The whole study would be implemented as a three-layer structure, such layers becoming Fog, Blockchain, and IoT layers. This research supports wireless communication to develop an IoT-Blockchain framework between several users on the IoT.

An IoT-Blockchain framework does have the following elements:

- a. Smart Devices
- b. Network

- c. Development tools
- d. Blockchain

Figure 2 shows the blocks in a blockchain. Blockchain Technology in IoT provides a mechanism of processing protected information files by IoT endpoints. BT transactions might be monitored or examined through someone verified to interact throughout the

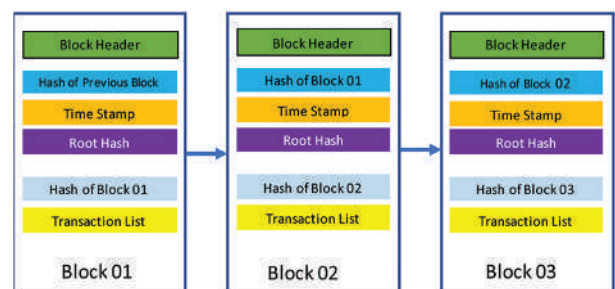


Fig 2.0: Blockchain's Blocks

IoT system.

An IoT-BC could enhance interaction security. BC's major advantages with IoT are:

1. Building trust between IoT public endpoints or minifying the possibility of accidents.
2. Minifying transportation costs by communicating directly without third parties.
3. Speeding up transactions in real-time.

The visualization servers contained efficient services in the fog, the intermediaries are 3rd-party servers that could also store secured information. A token offered by the approved Blockchain repository to the smart device does have the power to connect the model, notification key from the use of a key computer, collect information from the fog

Several smart contracts released in the approved Blockchain repository through the visualization servers, intermediaries or fog proprietors. IoT's portable device discovers the smart contracts from within the permitted Blockchain repository. Its permitted Blockchain repository produces the token to IoT's smart device. A smart device requests the keys in middleware from a key server as well as directs the token with the query. A primary server verifies the token from such an approved blockchain repository or produces a key for each smart device as well as a reaction down to a device. Once again, the IoT smart device is permitted to retrieve information from the cloud. The IoT nodes might be a student, a teacher, an employee, an operator or an accreditation unit member. They all

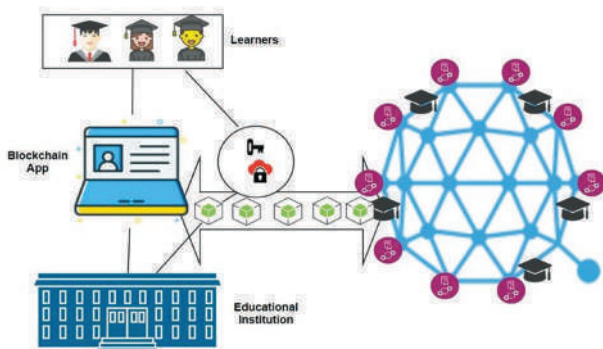


Fig 3.0: *Blockchain Based Educational Systems*

have smart devices for communication within the system (Figure 3).

4. Functions of Blockchain and IoT

A blockchain with IoT functions together with its goals that could be summarized as follows:

4.1 Decentralization structure:

Internet of things as well as BC, both

approaches would be identical. This structure eliminates the centralized approach or even provides the facility for a decentralized architecture. This enhances the aggregate control system probability of failure or efficiency.

4.2 Protection:

Transactions among endpoints users were often protected throughout the BC. This is a different approach to secure interactions. The BC enables IoT users to interact reliably with one another.

4.3 Identifier:

Both connected devices have always been distinctively recognized with such a cryptographic signature in IoT. Every block of BC is often distinctively recognized. Therefore, BC is a trustworthy innovation that offers distinctively recognized information which is accessed throughout a distributed ledger.

4.4 Accuracy:

Internet of things endpoints in BC has become capable of accessing the information being passed on the Internet. Its information has been accurate when it is confirmed by the miners once joining BC. Just confirmed blocks could indeed join the BC.

4.5 Independent:

All IoT endpoint users have become available to interact with any computer cluster with a decentralized framework

4.6 Optimization:

IoT users could interact in high-availability, a decentralized intelligent network that communicates to the destination device in real-time or transaction data. The Internet of

things enables the linked physical objects to exchange their data throughout the diverse system. It could be separated into the following points.

4.6.1 Smart devices:

The Internet of things assigns the distinctive identification number to each connected device in the system. These devices could exchange information among the IoT endpoints.

4.6.2 Routers:

The routers seem to be the machines that operate among physical objects as well as the cloud to make sure that now the link is maintained and therefore that protection has been granted to the network system.

4.6.3 Building a network:

It would be used to manage the flow of data and to maintain the quickest route amongst the other IoT endpoints.

4.6.4 Cloud:

It would be used to deposit or determine information. A repository in Blockchain is said to have characteristics including a distributed security model, restricted access, higher public access, bottom to top confidentiality or exchangeable identifications, whilst in a central database, the characteristics have become a centralized trust model, low security, low public access, high confidentiality but also un-transferable identifications. A blockchain is now more sophisticated than that of the centrally controlled storage in the above characteristics.

5. Advantages of Blockchain and IOT

A Blockchain-IoT interoperability strategy does have many extraordinary possibilities in

the educational system. This unlocks the two new windows next to each other. Several of the possibilities would be characterized as follows:

5.1 Constructing the trust among parties:

Due to its highly advanced features, the BC-IoT strategy would develop a relationship amongst the multiple smart devices. Just confirmed gadgets could interact in the system, as well as the miners would first confirm every block of the transaction, then they should join the Blockchain.

5.2 Lower costs:

Such a strategy would lower costs as it interacts effectively without the third party. This removes all third-party nodes among the recipient as well as the sender. This strategy offers full correspondence.

5.3 Minimize time:

It reduces a huge amount of time. This strategy minimizes the transaction time from weeks to seconds. For example, in the traditional approach, when students want to take admission in the university, he pays a lot of time to fill forms and wait for the reply from the university administration.

5.4 Security and privacy:

It offers privacy and security to users, devices, and data.

5.5 Public services:

This approach offers services to the public sector. Authenticated users could connect and exchange information.

5.6 Investment management:

Such an approach safely exchanges money without the need for a third party. This offers quick, safe as well as financial information facilities. This lowered the cost or time of transition.

5.7 Risk management:

Such an approach plays a vital role in examining or reducing the chance of resource or transaction failure.

5.8 Educational institution activities:

Blockchains also can change certain roles at educational institutions. The blockchain would become a great format for granting qualifications and experience, that would remove the institution administrator's role as the site of delivery. This would also reduce a stream of institution profits, this might reduce the costs or allow those services to be distributed anywhere.

6. Usecases

In higher education, blockchain has several applications, but this technology has not entered yet the public consciousness. It's just that, there is no lack of startup companies and major players intending to make that possible. But here are some instances of blockchain often used in the educational system around the world. Learning is becoming the essential part of our life. This is the base for innovations, too. The Blockchain and IoT are the aspect of such innovation which the educational institutions can profit greatly from. Most of the educational resources are available with the advent of technology in smart devices. The following are some Blockchain applications used for educational purpose.

6.1 EdgeCoin – educational digital token stable coin

EdgeCoin provides Blockchain as a service (BaaS) to positively disrupt the educational model of educational institutions. They use Interplanetary file system (IPFS) to digitize paperwork data and store them with the concept of proof. Also, EdgeCoin provides various smart solutions for higher education with the help of blockchain that enable users to create attractive dashboards with secured digital identity and security. It not only saves time and controls fraudulent digital piracy but also conducts global marketing for content with blockchain.

6.2 APII

Confirms the academic qualifications of potential students and teachers. This application expects to maximize everybody's ability and has functionalities on its website and mobile applications that are doing this. Individual people can maximize their capability and enterprises can minimize their charges and administration while quickly finding learners and skills. Find easy direct connections to your web pages, fingerprint verification, professional verification records and a nanoaccreditation generate mechanism utilizing QR coding.

6.3 Sony global education

A website that maintains a consistent record of student information and educational accomplishments. It introduces a new transcript and score management service via a digital service called Blockchain technology. It demonstrates how well the future of establishing and handling transcripts and high-security information in education will become a

blockchain mechanism. This is also demonstrated the use of blockchain for the installation of the “Next Generation of School ICT Environment” for the Japan’s Ministry of Internal Affairs and Communications.

6.4 Parchment

It is a forum for assessing academic qualifications, processing programs and generating certificates. Learners can discuss their academic performance with prospective employers with ease. It is the most broadly accepted electronic credential service that allows learners, academic institutions and employers to apply for, verify and share their credential in an easy and secure environment. Its framework has helped millions of people and thousands of colleges and universities transfer more than 30 million transcripts and other certificates worldwide.

6.5 ODEM

A platform that links teachers, students, and professionals to relevant courses and services. This is a blockchain-powered framework in which learners can improve their educational progress effectively with academics and the potential employers, and they also often establishing useful, energetic syllabuses. By using smart contracts to automatic payment, ODEM is easy to handle and much more cost effective to learn. Entering thousands of ODEM students from all over the world and get adequately compensated low-cost, student-driven market, allowing them to focus on what they do strongest: motivating curious students who enjoy learning

technologies are utilizing. The proposed research shows a combination of the IoT and blockchain for the educational system. The Blockchain is used to create a hyper-distributed public authentic ledger to record the transactions. The research opened a new opportunity in this area. The framework is implemented using a different set of IoT nodes and tested. This study can be a valuable framework to improve the communication security and efficiency in the educational system. This framework is appropriate for providing communication security where huge data is transmitted in a heterogeneous environment in the future. We have tested the system in different scenarios such as memory and processor usage in the integrated system and its impact on the performance of overall the system. We found that the proposed framework, not only increases the throughput but also the direct connection among IoT nodes are established which makes the system more stable. The outcomes of this research established a new IoT framework with blockchain technology for the existing educational system. In the future, researchers can enhance this research and apply it for whole the kingdom.

7. Conclusion and Future Scope

This study links the blockchain and IoT for the educational system, where IoT and blockchain

Indirect Free - Space Optical Communications for Vehicular Networks

Jijina Johnson¹

1. jijinajohnson99@gmail.com | +91 7012281835

Abstract

FSO is Free Space Optics and it has a line-of-sight technology that uses lasers to provide optical bandwidth connections to obtain desired outputs. FSO is an optical communication technique that propagate the light in free space means air, outer space, vacuum, or something similar to wirelessly transmit data for telecommunication and computer networking. This system can be used for communication purpose in hours and in lesser economy.

Key words: *free space optics, vehicular networks, optical communication*

1. Introduction

Working of FSO is similar to OFC (optical fiber cable) networks but the only difference is that the optical beams are sent through free air instead of OFC cores that is glass fiber.

FSO system consists of an optical transceiver at both ends to provide full duplex (bi-directional) capability. Vehicular networking is an enabling technology that provide traffic safety, traffic efficiency and autonomous cars

Parameters	FSO	Optical fiber	Microwave radio	Coaxial cable
Installation	Moderate	Difficult	Difficult	Moderate
Data rate	Gbps	Independent	Mbps	Mbps
Security	Good	Very good	Poor	Good
Connectivity	P2P, P2MP short and long reach	P2P, P2MP short and long reach	P2P short reach	Multidrop short reach
Maintenance	Low	Low	Low	Moderate
Spectrum license	Not required	Required	Required	Required

Fig 1.0: Comparison of FSO, Optical Fiber, Microwave Radio, and Coaxial Cable

management. Due to the wireless nature of the vehicle-to-vehicle communication network, a route may suddenly break. Designing of cross layer protocols, which span between transport and routing layers, can be beneficial in vehicular networks that support real time and multiple applications. Data packets carrying traffic safety and traffic efficiency information usually have higher significance and forwarded faster than other packets.

2. Vehicular Networks

Vehicular networks are mainly composed of vehicle nodes, which behave quite differently from other wireless nodes. Compared to common mobile nodes, vehicles may move at quite a high speed. This causes the topology of a vehicular network to change frequently. Depending on the traffic density, which can be very high in the case

of a traffic jam, or very low, as in suburban traffic. At either extreme the network may frequently disconnect. The network scale could be large in dense, urban areas, such as city centers, highways, and at entrances to big cities. The vehicular networks often have a new type of communication, which addresses the geographical areas where packets need to be forwarded. Vehicular networks differ from other types of mobile ad-hoc networks in which nodes move in a random way. Vehicles are constrained by road topology and layout, by the requirement to obey road signs and traffic lights, and by responding to other moving vehicles, leading to predictability in terms of their mobility. A common characteristic of nodes in vehicular networks is that nodes have ample energy and computing power. Since nodes are cars instead of small handheld devices. Vehicular networks are

usually operated in two typical communication environments. In highway traffic scenarios, the environment is relatively simple and straightforward whereas in city conditions it becomes much more complex. The streets in a city are often separated by buildings, trees, and other obstacles; therefore, there is not always a direct line of communication in the direction of intended data communication.

3. ID-FSOC

Indirect Free Space Optical Communication consist of transmitter, receiver, data packets, diffuse reflectors. This ID-FSOC system provide transmitter to receiver connection within transceivers. Transceivers are placed in the vehicles for the vehicle to vehicle communication. Similarly the vehicle to transmission media with wirelessly. The data packets are used for give safety for drivers and awareness of the traffic safety. Diffuse reflectors are used for reduce complexity of this system. Data delivery is a crucial task in vehicular networks since current applications require the cooperation of each and every vehicle. Current autonomous vehicles relay on embedded sensors to obtain a local vision of the environment to take decisions. Construction of a global vision requires the exchange of information through network communication. Meanwhile next-generation commercial systems will be capable of single fiber Vehicular Networks offer a range of opportunities for urban monitoring and data sharing on various aspects of the traffic. Vehicular networks do not have common constraints of WSNs, such as energy, bandwidth, and memory constraints, which allows for more accurate sensing and a larger amount of data to be collected. The need of humans to be connected has led to rapid growth in

wireless communications technology and non safety. Safety applications include spreading an alarm or warning with the aim of avoiding danger and reducing risk. The use of free space optics is particularly interesting when we perceive that the majority of customers does not possess access to fibers as well as fiber installation is expensive and demands long time. Moreover, right-of-way costs, difficulties in obtaining government licenses for new fiber installation etc. are further problems that has turned FSO into the option of choice for short reach applications. FSO uses lasers, or light pulses, to send packetized data in the terahertz (THz) spectrum range. Air, or fiber, is the transport medium. This means that urban businesses needing fast data and Internet access have a significantly lower-cost

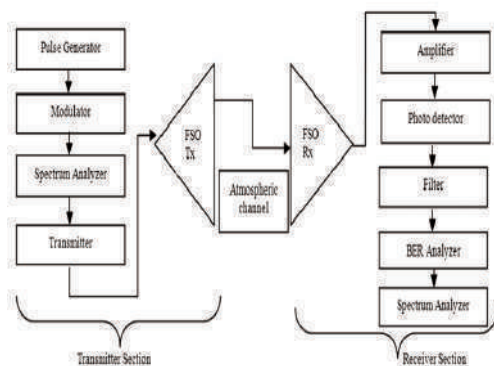


Fig 3.0: FSOC Block Diagram

option.

The transmitter part consists of a modulator, Laser Driver, Laser, Transmitter Antenna. The modulator modulate the data after laser driver. This laser driver help to provide a current instead of a voltage to the laser diode. In receiver part consist of clock recovery, amplifier, photodetector, filter, receiving antenna. The receiver generates a clock from

an approximate frequency reference and also recovery done with the help of clock and data recovery circuit is necessary to extract the data transmitted by the transmitter from the corrupted received signal and also to recover the accompany clock timing information at the receiver side of the communication system. Photodetector is a sensor that can convert the photon energy of light into electrical signal. Then signal is transferred to the filter that filter the unwanted impurities of the received signal. Receiver antenna sends the required output signal to the vehicular network.

4. Advantages of FSO

1. It has very low initial investment.
2. There is relatively high bandwidth.
3. It is a secure system because of line of sight operation and so no security system up
4. Free space optics is a flexible network that delivers better speed than broadband.
5. High data rate can be obtained which is comparable to the optical fiber cable's data rate but error rate is very low and the extremely narrow laser beam enables having unlimited number of FSO links which can be installed in a specific area .
6. Transmission of optical beam is done in air. Hence, transmission is having speed of light.
7. There is immunity to radio frequency interference.
8. Electromagnetic and radio-magnetic interference cannot affect the transmission in FSO link.
9. It is a straight forward deployment system. There is no need for spectrum license or frequency coordination

- between users as it is required in radio and microwave systems previously
10. FSO offers dense spatial reuse.
 11. Low power usage per transmitted bit is merit of FSO system.
 12. It has flexible rollouts.
 13. Installation is very easy and it takes less than 30 minutes to install FSO technology is implemented using a laser device. These laser devices or terminals can be mounted on rooftops, Corners of buidings or even inside offices behind windows. The high data rate and large information throughput available with laser communications are many times greater than in radio frequency (RF) systems. the laser had potential for the transfer of data at extremely high rates, specific advancements were needed in component performance and systems engineering, particularly for space-qualified hardware. Optical networks are widely regarded as the ultimate solution to the bandwidth needs of future communication systems. Optical fiber links deployed between nodes are capable to carry terabits of information but the electronic switching at the nodes limit the bandwidth of

PARAMETERS	LED	LASER
Working Principle	Electro-lumiance	Stimulated Emission
Full Form	Light Emitting Diode	Light Amplification by Stimulated Emission of Radiation
Response	Slow Response	Fast response in comparison to LED
Driving Current	It ranges from 50 to 100 mA.	It ranges from 5 to 40 mA.
Nature of Emitted Light	Incoherent and consists of various colours.	Coherent and Monochromatic.
Junction Area during Manufacturing	Wide Junction Area	Narrow and small Junction

Fig 4.0: Comparison between LED and Laser

a network. Optical switches at the nodes will overcome this limitation. With their improved efficiency and lower costs, Optical switches.

5. Data Packets

An IPv4 addresses a 32- bit binary value, which can be displayed as four decimal digits. The IPv4 address space offers about 4.3 billion addresses. Only 3.7 billion addresses can only be assigned out of 4.3 billion address. An IPv4 datagram is a variable-length packet comprised of a header (20 bytes) and data (up to 65,536 along with header). The header

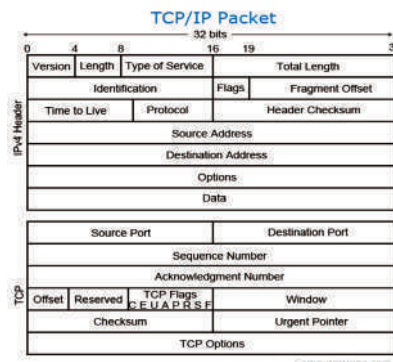


Fig 5.1: TCP IP Packet

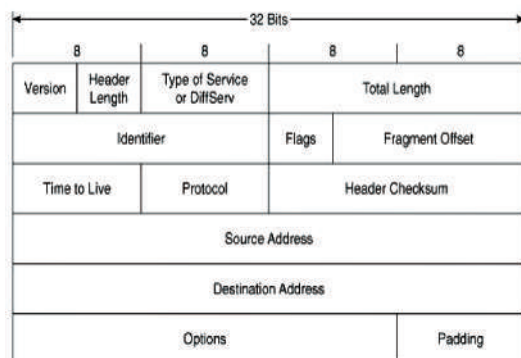


Fig 5.2: IPv4 Address

contains information essential to routing and delivery.

Base Header

- **Version:** It defines the version number of IP, i.e., in this case, it is 4 with a binary value of 0100.
- **Header length (HLEN):** It represents the length of the header in multiple of four bytes.
- **Service type:** It determines how datagram should be handled and includes individual bits such as level of throughput, reliability, and delay.
- **Total length:** It signifies the entire length of the IP datagram.
- **Identification:** This field is used in fragmentation. A datagram is divided when it passes through different networks to match the network frame size. At that time each fragment is determined with a sequence number in this field.
- **Flags:** The bits in the flags field handles fragmentation and identifies the first, middle or last fragment, etc.
- **Fragmentation offset:** It's a pointer that represents the offset of the data in the original datagram.
- **Time to live:** It defines the number of hops a datagram can travel before it is rejected. In simple words, it specifies the duration for which a datagram remains on the internet.
- **Protocol:** The protocol field specifies which upper layer protocol data are encapsulated in the datagram (TCP, UDP, ICMP, etc.).
- **Header checksum:** This is a 16-bit field confirm the integrity of the header values, not the rest of the packet.
- **Source address:** It's a four-byte internet

address which identifies the source of the datagram.

- **Destination address:** This is a 4-byte field which identifies the final destination.
- **Options:** This provides more functionality to the IP datagram. Furthermore can carry fields like control routing, timing, management, and alignment. IPv4 is a two-level address structure (net id and host id) classified into five categories (A, B, C, D, and E).

An IPv6 address is a 128-bit binary value, which can be displayed as 32 hexadecimal digits. Colons isolate entries in a sequence of 16-bit Hexadecimal fields. It provides 3.4 x 10³⁸ IP addresses. This version of IP addressing is designed to fulfill the needs of exhausting IP's and providing sufficient addresses for future Internet growth requirements.

As IPv4 uses two-level address structure where the use of address space is insufficient. That was the reason for proposing the IPv6, to overcome the deficiencies IPv4. The format and the length of the IP addresses were changed along with the packet format and protocols were also modified. Each packet consists of a mandatory base header succeeded by the payload. The payload includes two parts namely optional extension headers and data from an upper layer. The base header consumes 40 bytes, inversely the extension headers and data from the top layer usually hold up to 65,535 bytes of information.

Base Header

- **Version:** This four-bit field specifies the version of the IP, i.e., 6 in this case.
- **Priority:** It defines the priority of the packet concerning traffic congestion.
- **Flow label:** The reason for designing this

protocol is to facilitate with special controlling for a certain flow of data.

- Payload length: It defines the total length of the IP datagram excepting the base header.
- Next header: It's an eight-bit field describe the header that trails the base header in the datagram. The next header is one of the optional extension headers which IP uses or the header for an upper-layer protocol such as UDP or TCP.
- Hop limit: This eight-bit hop limit field assist with the same functions at the TTL field in IPv4.
- Source address: It is a 16 bytes internet address identifies the source of the datagram.
- Destination address: This is 16-byte internet address that generally describes the final destination of the datagram.

6. Conclusion

This Indirect Diffused Free Space Optical Communications for vehicular networks using data packets is very useful to passengers and drivers of the vehicle. It provide traffic safety, passengers needs and safety, vehicle to vehicle communication and vehicle to transmission system. There is no suffering of multipath dispersion. Unwanted impurity signals are removed by using filters. FSO is

a communication system where free space acts as medium between transceivers and they should be in LOS for successful transmission of optical signal. Medium can be air, outer space, or vacuum. FSO communication link is currently in use for many services at many places. Outdoor wireless access: it can be used by wireless service providers for communication and it requires no license to use the FSO as it is required in case of microwave bands.

7. Reference

1. Y. Kaymak, R. Rojas-Cessa, J. Feng, N. Ansari, M. Zhou, and T. Zhang, "A survey on acquisition, tracking, and pointing mechanisms for mobile free-space optical communications," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 1104–1123, 2018.
2. Diffuse reflectors - gigahertz-optik. Last Access: 12/22/2018. [Online]. Available: <http://www.gmp.ch/htmlarea/pdf/DiffuseReflectors.pdf>
3. H. Willebrand and B. S. Ghuman, *Free space optics: enabling optical connectivity in today's networks*. SAMS publishing, 2002.
4. S. Hranilovic, *Wireless optical communication systems*. Springer Science & Business Media, 2006.
5. D. Cornwell, "Space-based laser communications break threshold," *Optics and Photonics News*, vol. 27, no. 5, pp. 24–31, 2016.
6. G. Yun and M. Kavehrad, "Indoor infrared wireless communications using spot diffusing and fly eye receivers," *Canadian Journal of Electrical and Computer Engineering*, vol. 18, no. 4, pp. 151–157, 1993.

Monitoring Methods in Online Health Science Education- A KUHS Experience

Unnikrishnan Vadakkoot Vasudevan¹

1. Professor of Physiology, Dept. of Medical Education, Govt of Kerala & Former Dean (Academic), Kerala University of Health Sciences.

Abstract

Monitoring is pivotal for any academic activity. It comprises of three components, namely, monitoring of Content, Process and Outcomes. These are monitored at 3 levels (Multi-level Monitoring): Institutional, Faculty and University, respectively by the Institutional Academic Committee, Institutional Technical Committee for Online Education (both under the Principal) and the University, (through Faculty Deans and Dean [Academic]). Periodic Reports are submitted to the University through an Electronic Format. This serves as a three tier Multi-level Monitoring system which is found to be an effective approach especially in the emerging flux in conventional academic monitoring systems.

Key words: *online education, health sciences, multi-level academic monitoring systems*

1. Introduction

Kerala University of Health Sciences affiliates 313 institutions in seven streams, namely Modern Medicine, Ayurveda, Homeopathy, Nursing, Pharmacy and Allied Health Sciences. It has a state-wide jurisdiction.

During the Covid-19 pandemic, conventional classroom and laboratory and clinical teaching learning became impracticable and world-over, the universities and educational institutions switched over to online systems. The KUHS followed suit. However since the calamity unfurled in an unpredictably fast manner, the stakeholders were unprepared for such drastic changes the effectiveness could not be

assessed, at least during initial few months. The main issues were the lack of uniformity in (1) quality of content, even though syllabus was in place (2) method of delivery which included the lack of a common platform for content delivery and (3) monitoring the outcomes at critical levels. These issues were approached using a three-pronged strategy of Academic Monitoring of the content, process and outcomes.

2. Methodology

The status of the existing facilities and practices were first assessed through an extensive survey among the stakeholders who included Teachers, Students, Members of the

Academic Bodies like Boards of Studies, Academic Council and Senate, Principals and Managements. The results were analyzed and preliminary conclusions regarding the core issues were drawn, which include:

1. Lack of uniformity in content and its delivery among institutions.
2. Lack of a common platform for content delivery
3. Issues on non-availability of bandwidth and cost effectiveness.
4. Acceptance of new technology by students and teachers.
5. No effective monitoring system to address these issues.

The KUHS addressed each of these issues by instituting an Online Academic Monitoring System, which ensures the following:

- a. **Monitoring the Quality of Content:** This was done at the institutional level. The Principal / Head of Institution is the person responsible for ensuring the delivery of content in the Online Platform, to whom the faculty reports the progress, like the subjects taken, attendance, timing, duration, assessment reports etc. He/she prepares a report in a standard

format and reports to the Faculty Dean of the concerned stream.

- b. **Monitoring the Method of Delivery:** There were several options available with variable features. These included, among others, G-Suite, Zoom, Whatsapp, Webex and Moodle. However these platforms were widely varying in their specifications like the number of participants permissible, bandwidth requirements, cost, conferencing / interaction facilities, privacy issues and ease of use. Further, each faculty Member used the platform he/she was comfortable with. The Principal / Head of Institution is directed to monitor the whole content delivery system.
- c. **Monitoring the Outcomes at critical levels:** The reports of (a) and (b) are reported weekly to the Deans of the respective faculties. After analysis, they submit a consolidated report to the Dean (Academic) of KUHS. These are assessed by the University and appropriate actions taken.

Usage of digital dashboard:

The institutions are encouraged to use the Digital Dashboard of the KUHS Institutional

Parameter	Person Responsible	I Level Monitoring	II Level Monitoring	III Level Monitoring
Content	Teacher	Principal	Faculty Dean	Dean [Academic]
Delivery	Teacher & Technical Section	Principal	Faculty Dean	Dean [Academic]
Outcomes	Institution	Principal	Faculty Dean	Dean [Academic]

Tab 1.0: Summary of Academic Monitoring System in KUHS

Academic Committee of the College, which is the dedicated and secure communication channel to the University.

3. Outcomes

Actions Initiated consequent to inputs from the monitoring process

1. Formulation of an Online Education Policy for the University.
2. Adoption of MOODLE as the preferred Learning Management System
3. Instituting training of teachers through ToT mode and Self-Learning Mode in MOODLE.
4. Sensitization workshops to members of Academic and Governing Bodies of the University and Principals of all affiliated institutions.

5. Sensitization workshops to technical members in charge of Online Education Programs of affiliated institutions.
6. Training Modules and Teaching Materials [Notes and Videos] for the training made available to all stakeholders free of cost.
7. Keeping a communication channel open for potential flaws or gaps in the system.

4. Conclusion

Monitoring the Online Educational Activities is paramount in ensuring the quality of education. The content quality, mode of delivery and outcomes are measured using a three-tier system which ensures compliance and overall effectiveness.

Communications to be addressed to:
Prof. Dr. V.V. Unnikrishnan
drvvuk01@gmail.com; 9447407042

A Hybrid Approach for Detection of Pneumonia from X-Ray Images

Maheshwaran T. Parameswaran¹, Shreyas S. Kumar², Pankaj Kumar G.³
Shemin Anto⁴, S. Krishna Bhat⁵, Narayanan V. Eswar⁶

1. Dept of EEE, Federal Institute of Science And Technology, Kerala, India | mahestpm@gmail.com
2. Dept of CSE, Federal Institute of Science And Technology, Kerala, India | shreyassreekripa@gmail.com
3. Dept of CSE, Federal Institute of Science And Technology, Kerala, India | pankaj@fisat.ac.in
4. Dept of CSE, Federal Institute of Science And Technology, Kerala, India | sheminanto@gmail.com
5. Dept of CSE, Federal Institute of Science And Technology, Kerala, India | memotoskbhat@gmail.com
6. Dept of CSE, Federal Institute of Science And Technology, Kerala, India | narayananveswar@gmail.com

Abstract

The paper describes our work on comparing various prevalent image masking and classification techniques for detection and classification of pneumonia from chest X-Ray images. The imbalanced dataset used consists of two basic classes, Healthy Lungs and Pneumonia affected Lungs X-Rays. Pneumonia affected Lungs X-Ray images can be classified further into two subcategories, Lungs affected by Bacterial Pneumonia and lungs affected by viral pneumonia. The work proposes a hybrid model combining different prevalent Machine Learning models like Sequential CNN Classifiers, Resnet classifiers, AlexNets, DenseNets, and classical machine learning algorithms like Support Vector Machines and KNNs and analyses the optimal architecture for 1. Binary Classification between Healthy and Disease affected Lungs, 2. Classification between Bacterial and Viral Pneumonia, 3. Multi-Class classification between Healthy, Bacterial and Viral Pneumonia. Depending on the training and validation accuracies, the most promising models will be further acted upon for hyperparameter tuning to optimise the models and improve the overall accuracy, precision and recall. Medical professionals can utilise the model to aid in the faster identification of potential patients, increasing their recovery chance.

Keywords: *convolutional neural networks, ensemble classifier, pneumonia detection*

1. Introduction

Pneumonia, which can be sometimes life-threatening, is a disease affecting the lungs characterized by infection, which results in inflammation of air sacs. It may become a risk factor in cases where the affected people

belong to the categories such as infants, adults over age 65, those who suffer from other diseases such as asthma, diabetes and pulmonary diseases, having weak immune systems.

In current pandemic situations like COVID-19,

early detection of diseases like pneumonia needs extreme attention as COVID-19 has adverse effects on people with respiratory diseases. In such scenarios, an automated system that can easily classify pneumonia affected lungs helps the medical professionals save much time and provide the patients with medical assistance faster.

Here we discuss an ML-based system that can precisely predict pneumonia affected lungs. The dataset fig 1 contains images of Pneumo-

that correspond to different categories of the disease, has been proposed. There is a weight associated with each model, which will be considered for calculating the final result.

When the system is provided with a new x-ray image, each model identifies whether the person has lungs affected with pneumonia or not individually. After the identification, the individual models' results are combined using a weighted voting scheme to make the final prediction.



Fig 1.0: *Illustrative Examples of Chest X-Rays in Patients with Pneumonia*

nia patients and normal patients. The normal chest X-ray depicts clear lungs without any areas of abnormal opacification in the image. Bacterial pneumonia (middle) typically exhibits a focal lobar consolidation, in this case in the right upper lobe (white arrows), whereas viral pneumonia (right) manifests with a more diffuse "interstitial" pattern in both lungs. Several approaches have been tried for detecting pneumonia. The dataset consisted of X-rays of 3875 pneumonia patients and 1341 normal patients.

A hybrid system that uses a voting scheme among multiple models, which have been already trained with various available datasets

2. Related Works

Tawsifur Rahman et al. mentions that the fatality rate of Pneumonia in children is 1.4 million every year, which is 18% of the total children's death rate of less than 5 years [1]. The leading cause for this is incorrect diagnostics, leading to wrong medication, which is the major cause. The training dataset consisted of 5247 images of resolution 400*2000 pixels. The study used MATLAB to train and evaluate the different algorithms. These images were resized and normalised according to the pre-trained model standards, after which the images were visualised to find layers that activated the neural network. They found that most convolutional neural networks (AlexNet,

DenseNet201, ReseNet8) learn to detect features like colour and edges for the first layer and the build-up on other layers increases.

The performance of each of the neural networks was found. After the analysis, the classification of accuracy, precision, and recall of normal and Pneumonia images, bacterial, and viral pneumonia images, and normal, bacterial, and viral Pneumonia were 98%, 97%, and 99%; 95%, 95% and 96%; and 93.3%, 93.7% and 93.2%, respectively.

Dimpy Varshni et al.; developed a method for the detection of Pneumonia using a densely connected neural network(denseNet-169)[2]. It was a three-stage process that included preprocessing, feature extraction and classification of the data set. In the preprocessing model, the images are downsized from 1024*1024 to 224*224. This downsizing is down to reduce the complexity of the computation. A pre-trained DenseNet-169 was then used for feature extraction.

The DenseNet-169 is the most productive way of dealing with image recognition because of the peculiarity in the convolution and the pooling layers, which helps in the extraction and the mapping of the images. And finally, the extracted features are fed to the SVM classifier. The gamma values obtained were high. Ilyas Sirazitdinov et al. used an alternative method for the detection of Pneumonia as it is very difficult for the doctor to decide if the lung problem is caused due to Pneumonia or other preexisting lung problem [3]. For the training of the classifier, an existing RSNA dataset was used.

The proposed methodology is based on an ensemble of RetianNet and Mask R-CNN. As Pneumonia manifests in the small region

of the lung, the detection is challenging. To overcome this, the author uses FPN principle to generate multi-scale feature maps with quality information than default methods. For further enhancing FPN residual model is used as a base. After training the model for the identification of the region, post-processing is done. In the post-processing, the overlapping bounding boxes obtained during the prediction stages are removed using the non-maximum suppression (NMS) algorithm. The next is the tuning of the sigmoidal activation function; the higher the function's value, the better it is. One of the major difficulties faced due to ensembling is the difficulty in combining due to the different confident level scale. To tackle this, the author picked a greedy strategy model. During the process of training and testing, it has been found that there was a decrease in the loss as the number of epochs increased. This pattern continued till the 20 epochs, and the decrease in the loss became very low.

Vikash Chouhan et al. elaborates on the advancements made in the increasing accuracy of the deep learning models to reach human-level accuracy[4]. Even though the model has not become so advanced to replace the doctors, it has still made ways to many medical fields as a support system that reduces the time necessary for diagnosis. He adds that of the 1.7 million that get Pneumonia, about 50000 adults die in the US alone; moreover, World Health Organisation (WHO) has reported the majority of death in children below five is caused due to Pneumonia which is estimated to be 1.4 million. To tackle this development of cheaper and accurate methods has to be developed. In this paper, a transfer learning approach using five different neural network models, namely AlexNet,

DenseNet121, ResNet18, InceptionV3, GoogLeNet. At first, data preprocessing is done as the multiple models are used, and overfitting was possible, as preventive step noises were added. After this, the images were resized to 224*224, and augmentation techniques were used to augment images with varying intensity. After which, the data set is passed to the five neural networking systems and is fed to the ensemble classifier. The classifier takes the output from the model and considers the one with the majority vote as the output. The created model outperformed the individual model, and the author further added that the accuracy could be further improved by increasing the size of the data set.

3. Architecture

Voting is a fairly common ensemble of ML models designed to give higher accuracy of predictions than the individual models can perform independently. Weighted voting involves the assignment of weights to each of the models based on various factors, such as how well a model has fared in its testing/validation phase. Bagging or bootstrap aggregating uses bootstrapping to generate new training sets from a single training set D by sampling with replacement, which are then used to train multitudes of models. Once the models are trained, either voting (for classification) or averaging of output (for regression) is used to determine the final prediction.

Our proposed algorithm uses a hybrid method of bagging and weighted voting. The concept is that the use of small datasets for a single model is highly impractical. This is because a small or incomplete dataset can result in the underfitting of a model. We propose a form of weighted voting ensemble which shows good results when the dataset is highly limited in

its quantity. From the dataset that we used [7], we generated synthetic data. Synthetic data has proved to be a highly stable way of expanding or increasing dataset size, provided that the generation replicates naturally occurring data well enough. Toda et al.[5] describes the training of a segmentation neural network, using synthetic datasets generated from a limited set of pictures of crop seed phenotypes, using a mask RCNN, and shows that such synthesised datasets are sufficient to train a model. Since synthetic data responds differently to different models, our proposed method takes advantage of this fact.

Different models or networks are trained using this combination of both synthetic and original data. Weights are assigned based on their performance in the individual testing phase. An average product of weights and

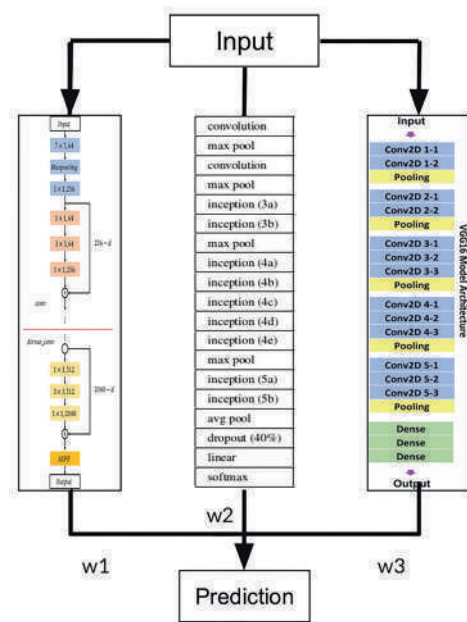


Fig 2.0: Dataflow Diagram of a General Ensemble Learning Model in Machine Learning

predictions of individual models is calculated, called the voting aggregate. This is compared with a voting threshold. This threshold defines the minimum required average aggregate from the voting process for the final prediction to be positive. In other words, if the voting aggregate is greater than the voting threshold, the prediction is declared positive, else it is declared negative.

For our experiment, we have used three networks: the InceptionV3, the VGG16 and the ResNet50V2. We have used these for the prediction of pneumonic lungs from the x-rays dataset [7]. From the individual performance in the testing and validation phases of each of the models, the InceptionV3 has been assigned the highest weight of 2.5 while the other models have been assigned 0.25 each. The synthetic data was generated from the dataset using the Keras image preprocessing module.

4.. Experimental Results



Fig 3.0: Training Data Distribution

Figures 7 through 10 show the accuracy and loss graphs for all the ResNet50V2,



Fig 4.0: Testing Data Distribution

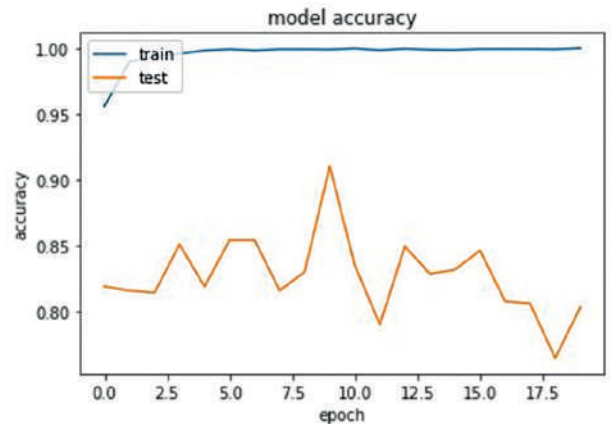


Fig 5.0: Model Accuracy for the ResNet50V2 Network

InceptionV3 and VGG16 networks. Figures 3 and 4 respectively, show the data class distributions in the training and testing phases. The models themselves do not show any individual superiority. The ROC curve shows that individually, the InceptionV3 network has the best results.

From the Figures 11 (Confusion matrix) and 12 (ROC curve), we can see that the hybrid ensemble has the best tradeoff between TP

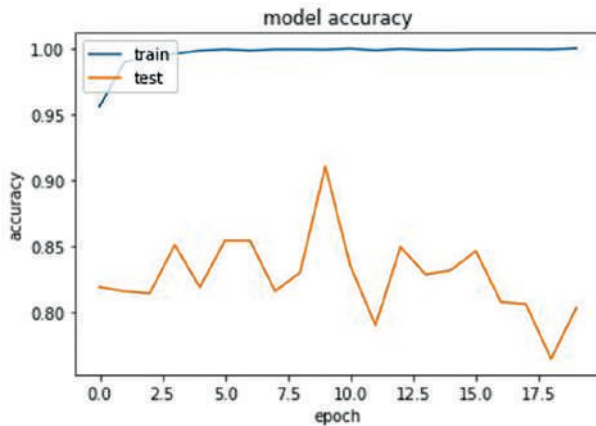


Fig 6.0: Model Loss for the ResNet50V2 Network

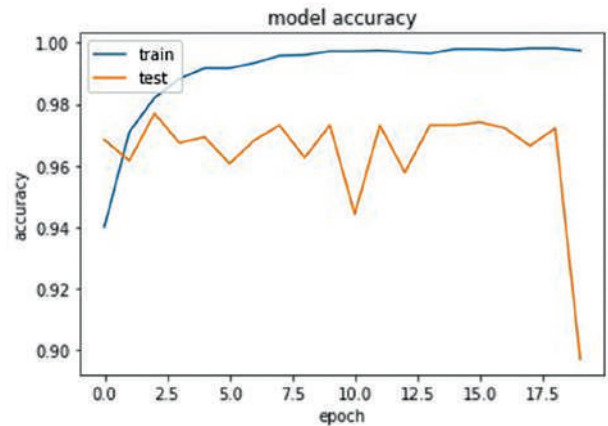


Fig 8.0: Model Accuracy for the InceptionV3 network

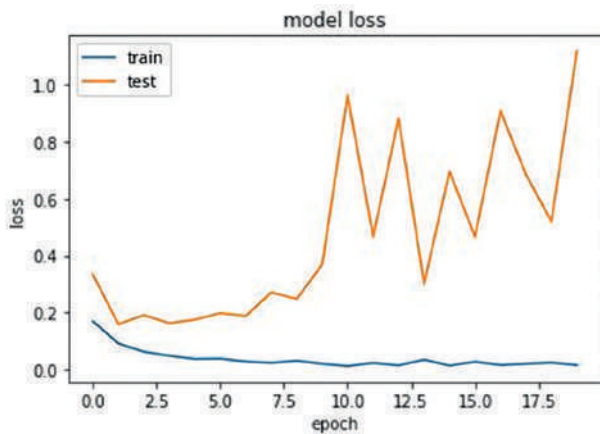


Fig 7.0: Model Loss for the InceptionV3 network

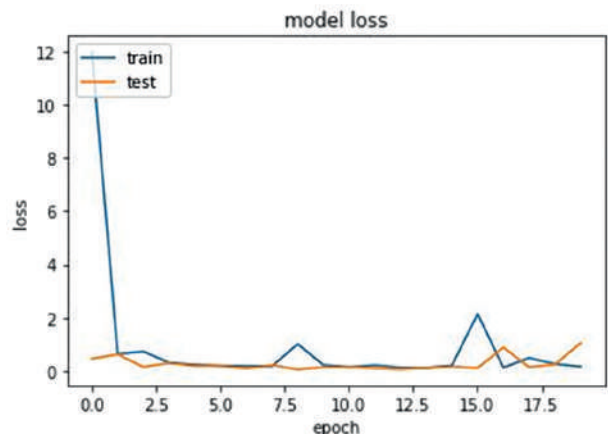


Fig 9.0: Model Loss for the VGG16 Network

rate and FP rate (true positives and false positives). The precision and recall values can be calculated from the confusion matrix and is determined to be 0.871 and 0.815, respectively. The ROC curve also proves that it is closer to the ideal model than any individual models.

5. Conclusion

The paper describes a hybrid ensemble approach for detecting pneumonic lungs from X-Ray images using 3 different models. The

experimental results showed that the final hybrid is considerably better than the individual models, and can be used for any models regardless.

6. Acknowledgment

Our work would not have been possible without the constant support of the Cognitive Computing and Research Center (CCRC), FISAT. Hence we express our heartfelt gratitude to all the associated CCRC members for

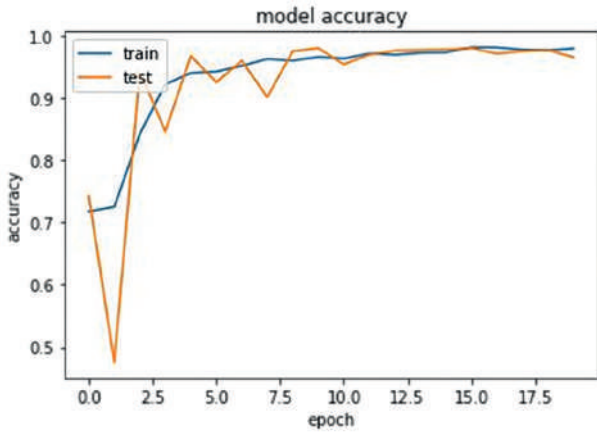


Fig 10.0: Model Accuracy for the VGG16 Network

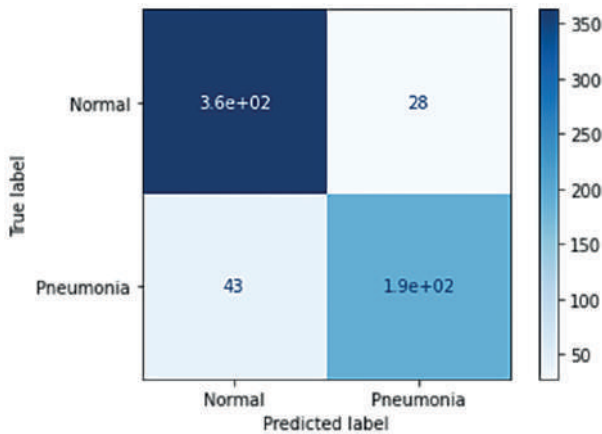


Fig 11.0: Confusion Matrix for the Final Hybrid Ensemble Model

their support as well.

7. References

1. Rahman, T., Chowdhury, M.E., Khandakar, A., Islam, K.R., Islam, K.F., Mahbub, Z.B., Kadir, M.A. and Kashem, S., 2020. Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. *Applied Sciences*, 10(9), p.3233
2. Varshni, D., Thakral, K., Agarwal, L., Nijhawan, R. and Mittal, A., 2019, February. Pneumonia detection using CNN based feature extraction. In 2019 IEEE International Conference

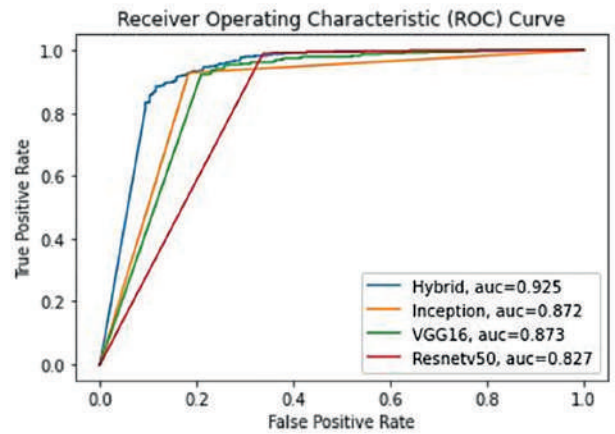


Fig 12.0: Collective ROC Curves of Individual Models and the Final Hybrid Ensemble

on Electrical, Computer and Communication Technologies (ICECCT) (pp. 1-7). IEEE

3. Sirazitdinov, I., Kholiavchenko, M., Mustafaev, T., Yixuan, Y., Kuleev, R. and Ibragimov, B., 2019. Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. *Computers & Electrical Engineering*, 78, pp.388-399
4. Chouhan, V., Singh, S.K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., Damaševičius, R. and De Albuquerque, V.H.C., 2020. A novel transfer learning based approach for pneumonia detection in chest X-ray images. *Applied Sciences*, 10(2), p.559
5. Toda, Y., Okura, F., Ito, J., Okada, S., Kinoshita, T., Tsuji, H. and Saisho, D., 2020. Training instance segmentation neural network with synthetic datasets for crop seed phenotyping. *Communications biology*, 3(1), pp.1-12
6. Jaderberg, M., Simonyan, K., Vedaldi, A. and Zisserman, A., 2014. Synthetic data and artificial neural networks for natural scene text recognition. *arXiv preprint arXiv:1406.2227*
7. Chest X-Ray Images (Pneumonia): <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>
8. Wang, S. and Yao, X., 2009, March. Diversity analysis on imbalanced data sets by using ensemble models. In 2009 IEEE symposium on computational intelligence and data mining (pp. 324-331). IEEE.
9. Surdeanu, M. and Manning, C.D., 2010, June. Ensemble

- models for dependency parsing: cheap and good?. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (pp. 649-652).
10. Mohan, S., Thirumalai, C. and Srivastava, G., 2019. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, 7, pp.81542-81554.
 11. Liu, T., Fan, W. and Wu, C., 2019. A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. *Artificial intelligence in medicine*, 101, p.101723.

Multilingual Social Media Cyberbullying Detection

Joel Jorly¹, Shiv Shankar Nath^{2,3}, Nahid M.A.³
Pankaj Kumar G.⁴, Paul Elias Sojan⁵

1. Computer Science and Engineering, Federal Institute of Science & Technology, Angamaly, Kerala | jorlyjoel@gmail.com
2. Computer Science and Engineering, Federal Institute of Science & Technology, Angamaly, Kerala | shivshankarnath@gmail.com
3. Computer Science and Engineering, Federal Institute of Science & Technology, Angamaly, Kerala | nahidma7@gmail.com
4. Computer Science and Engineering, Federal Institute of Science & Technology, Angamaly, Kerala | pankaj@fisat.ac.in
5. Computer Science and Engineering, Federal Institute of Science & Technology, Angamaly, Kerala | paul.eliassojan@gmail.com

Abstract

Cyberbullying or cyber harassment is a form of bullying or harassment using electronic means. It has become increasingly common, especially among teenagers, as the digital sphere has expanded and technology has advanced. This is a solution capable of identifying cyberbullying made on social media in Malayalam or Manglish by analysing a particular post on social media. The data which we used are collected from Instagram Comments, and then the data is manually labelled. The labelled data is used to find the rate of cyberbullying and the rate of support from various comments, and accuracy is calculated with various machine learning algorithms.

Keywords: *cyberbullying, sentiment analysis, malayalam text processing*

1. Introduction

Cyberbullying or cyber harassment is a form of bullying or harassment using electronic means. In this day and age, where the internet is booming, social media has become a huge part of our lives. However, social media has made it hard for many young people to accept themselves for who they are, and as a result, it forces many of them to change themselves mentally and physically. They believe that these steps are necessary to be accepted by the lifestyle and standards that social media has set. Cyberbullying can affect anyone and can even cost someone their life. It is crucial to eradicating it because, through

this, hundreds of lives can be saved.

This work aims to classify the comments under a particular Social Media Post and classify it as either a Support comment or a Bullying comment [5]. Thereby, the percentage of bullying in the comment section can be deduced accurately. In this era where the worldwide web has become an integral part of our lives, this project has massive ethical and social relevance.

Cyberbullying is indeed a real problem in today's society. Statistics have shown that :

1. Over 36.5 percent of people feel that they have been cyberbullied in their lifetime.

2. 17.4 percent reported that they had experienced cyberbullying at some point in the last 30 days.
3. 60 percent of teenagers have experienced some sort of cyberbullying.
4. 70 percent of teenagers have experienced people spreading rumors about them online.
5. 87 percent of teenagers have seen cyberbullying occurring online.

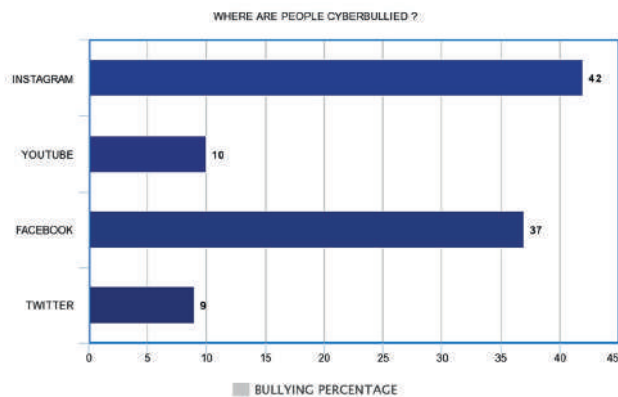


Fig 1.0

A high priority has been given to the Indian regional languages. The primary focus is on the Malayalam social media community. There has indeed been a steady increase in the rate of cyberbullying through regional languages. This has mainly been prominent due to the rise in the usage of special multi-lingual keyboards. Thus, besides English, this project also targets regional languages like Malayalam.

Moreover, research has also proved that cyberbullying has a wide range of adverse effects on the victim. Some of these include :

1. Experiencing sadness, anger and frustration for prolonged periods.

2. Losing interest in various activities.
3. Withdrawal from other individuals.
4. Changes in an individual's sleep patterns.
5. Sudden deviation in the individual's appetite and eating habits.
6. Lack of motivation and a tired demeanour.
7. Depression and negative thoughts.

2. Related Works

Soumya et al. [7] proposed a preprocessing technique that retrieved tweets contain hyperlinks, punctuation's, special characters, etc., these have been removed using regular expressions in python language. After that, the 3184 tweets are manually verified and assigned with positive and negative sentiments. Among the 3184 tweets. 1586 tweets are positive sentiment oriented, and 1598 tweets are negative sentiment oriented sentences.

Malayalam is a highly agglutinative language, and preprocessing is challenging. A hybrid approach is done to find the positive, negative or neutral opinions from users' writings. This work will give the sentence's polarity, and each word is tagged with certain rules with over 30,000 tokens [9].

Similar work is done in the Malayalam language using a rule-based approach by tokenization of the Malayalam words based on sandhi rules. Based on the sentiments, the most frequently occurring words have been assigned the sentiments, which are used for analyzing the words with the help of negation rules [13].

Itisha et al. [4] proposed a normalization approach on tweets, such as removing stop-words, URLs, username, punctuation and normalizing elongated misspelt words, and informal acronyms, negation handling and

emoticon replacement.

Zhao et al. [8][6] put forward various text pre-processing methods on Twitter sentiment analysis such as Replacing negative mentions, Removing URL links in the corpus, Reverting words that contain repeated letters, Removing numbers and stop words and expanding acronyms to their original words.

3. Problem Statement

There were many problems that hindered the accurate analysis and classification of the comments. These included the presence of the following:

A. Emoji in comments

The presence of components such as expressive Emojis, special symbols, punctuations and user handles appearing in the comments can create noise. Thereby, these undesirable components will subsequently hinder accu-



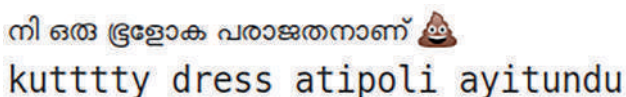
എല്ലാവർക്കും ഇഷ്ടപ്പെടും ഉണ്ടല്ലോ അതിനെ ഇങ്ങനെ മോശം രീതിയിൽ താളി കെട്ടരുത് 😊
polichu 🍕 muthe super aayittundu..... @anawara
You r such a kind person!! 🙌🙌🙌🙌🙌🙌🙌

Fig 2.0: *Emoji in comments*

rate detection and classification.

B. Spelling mistakes

Multiple comments may contain spelling mistakes. As a result, the presence of spelling



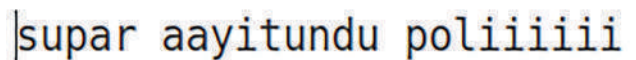
നി ഒരു ഭൂളോക പരാജയനാണ് 🙄
kuttty dress atipoli ayitundu

Fig 3.0: *Spelling mistakes*

errors in these sentences can lead to misclassification. This will adversely affect the overall quality of the classifier.

C. Repetition of end characters

A good percentage of the words in the comments may consist of repeated tailing characters. In other words, this involves the repetition of a single character back-to-back multiple times. One example of a word with



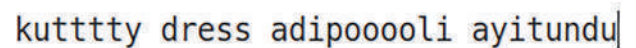
supar aayitundu poliiiiiii

Fig 4.0: *Repetition of end characters*

repeated tailing characters is displayed in the figure below. Thereby, the presence of these tailing characters can lead to misclassification.

D. Repetition of intermediate characters

Similarly, a significant fraction of the Social Media comments may consist of words with



kuttty dress adipoooooli ayitundu

Fig 5.0: *Repetition of intermediate characters*

repeating intermediate characters. One example of a word with a repeated intermediate character is displayed in the figure below. Hence, this can also severely affect the accuracy of the detection and classification.

E. Occurrence of intermediate English words

Lastly, some Malayalam or Manglish comments may contain intermediate English

kutttty dress adipoooooli ayitundu

Fig 6.0: Occurrence of intermediate English words

words. However, once again, the presence of these words will prove to be an undesirable component. Thereby, the overall accuracy and, subsequently, the classifier's strength will be affected by this hindrance.

4. Architecture

This work presents a novel solution towards predicting the cyberbullying rate and support rate based on comments made in Malayalam or Manglish under a particular social media post. Comments in social media are captured and undergo a series of steps and classified into Support and Bullying comments by the decision engine. A social media monitoring utility to monitor social media platforms continuously. After passing through a series of preprocessing steps, the inputs are classified as either bullying comments or support comments. The dataset for training was created by manual effort spanning hundreds of hours

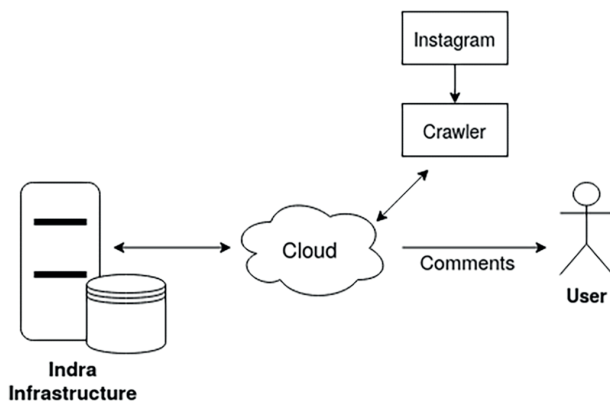


Fig 7.0

[5].

The Indra Cloud infrastructure has been used in order to implement this system [10]. Initially, a VM (Virtual Machine) instance is created in the infrastructure [11]. A Crawler is initialized within this instance. This Crawler enables the system to extract the comments from Instagram for classification [2][3]. These comments then undergo various steps, such as filtering and preprocessing. Thereby, they are finally classified as either support comments or bullying comments.

5. Methodology

A. Solution 1

The emojis and punctuation can be filtered by comparing them with the Unicode for emojis and punctuations. If a match is found, then we can remove it. Account mention and tagging can be found out by checking if the first letter of a word is '@', '#' and then removing the symbol along with the rest of the word.

B. Solution 2

The majority of the comments on social media may consist of spelling mistakes. These can hinder accurate classification. Hence, spelling mistakes can be corrected using the Levenshtein distance algorithm [1][12], which finds the distance between the word and words in the dictionary and replaces it with the word with the least distance.

C. Solution 3

The majority of the comments on social media may also contain words with repeating tailing characters. Repeating Tailing letters can be removed by using regular expressions. Here, repetition of the tailing letters is detected and removed.

D. Solution 4

Similarly, the presence of words with repeated intermediate characters also exist. Intermediate character repetition can be detected using regular expression. Through this, the repetitions are subsequently removed. Hence, more uniform comments are obtained. This can eventually help in attaining superior classification results.

E. Solution 5

There may also be the presence of Malayalam comments with intermediate English words. An Intermediate English word in a Malayalam sentence can be translated to Malayalam. Thereby, linguistically uniform comments can be obtained.

6. Scope and Future Enhancement

With the advent of the internet and social media platforms, connecting and sharing parts of our lives with other people has never been easier.

This is a project which has massive scope in ensuring a better internet experience for humans. The world wide web is a deep ocean filled with all types of metaphorical “flora” and “fauna”. In this day and age, where the internet has become an integral part of our lives, this project has massive untapped potential.

A. Scope of the project

1. Ability to aid in the minimization of Cyberbullying through accurate detection of comments implying “Cyberbullying”.
2. Can aid in obtaining the demographic of online bullies. This can once again help streamline the pursuit to thwart online bullying.
3. It can ultimately help bring a halt to

depression among individuals, especially teenagers, fuelling rising suicide rates.

B. Future enhancement

1. A system can be devised to accurately convert the English words to Manglish words that appear in between. This can be carried out in such a way that the overall structure of the sentence is preserved.
2. We can track the demographics of those users who are spreading a lot of cyberbullying comments. Moreover, scores can also be assigned for each user based on how often they make cyberbullying comments.
3. The system can accurately detect personal threats. Moreover, it can also aid in the elimination of such threats.

7. Conclusion

In a nutshell, for this system, a language-based method has been used for detecting cyberbullying. By recording the percentage of support and bullying words within a post, an accuracy of 79 percent was obtained on the test dataset. With the help of fine-tuning and by adopting a larger corpus, the method's accuracy can be significantly improved. To conclude, this system is essentially a Multilingual-Based Social Media Cyberbullying Detection System. This model accepts Social Media post URLs from the user and uses a trained model to predict the nature of the Social Media comments under this post. In other words, each comment is classified as either a “Support” comment or a “Bullying” comment. All the testing and training of the classifier are done using manually created datasets. The comments which were initially extracted go through a series of refining

steps like Filtering and Preprocessing. This further involves various smaller stages such as spell-checking, tokenization and so on. Moreover, methods such as Dimensionality Reduction and Hyperparameter tuning have been incorporated to improve the accuracy. We believe that our project can go a long way in pulling the plug on cyberbullying in Social Media.

8. Acknowledgment

Our work would not have been possible without the constant support of the Cognitive Computing and Research Center (CCRC), FISAT. Hence we express our heartfelt gratitude to all the associated CCRC members for their support as well.

9. References

1. Zhang, Shengnan & Hu, Yan & Bian, Guangrong. (2017) , "Research on string similarity algorithm based on Levenshtein Distance", 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), (2017)
2. Singh, Rahul & Roy, Sudipta. (2018), "Web Scraping Of Social Networks And Websites Using Scrapy And Selenium", 2018 IEEE International Conference on New Trends in Engineering & Technology, (2018)
3. N. S. Purohit, A. B. Angadi, M. Bhat and K. C. Gull, "Crawling through web to extract the data from Social networking site - Twitter," 2015 National Conference on Parallel Computing Technologies (PARCOMPTECH), Bengaluru, India, (2015)
4. I. Gupta and N. Joshi, "Tweet normalization: A knowledge based approach," 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), Dubai, United Arab Emirates, (2017)
5. E. V. Altay and B. Alatas, "Detection of Cyberbullying in Social Networks Using Machine Learning Methods," 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), Ankara, Turkey, (2018)
6. Z. Jianqiang, "Pre-processing Boosting Twitter Sentiment Analysis?," 2015 IEEE International Conference on Smart City/ SocialCom/SustainCom (SmartCity), Chengdu, China, (2015)
7. Soumya S and Pramod K.V, "Sentiment Analysis Of Malayalam Tweets Using Machine Learning Techniques", 2020 The Korean Institute Of Communication And Information Sciences(KICS), 2020
8. Z. Jianqiang and G. Xiaolin, "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis," in IEEE Access, vol. 5, (2017)
9. Nair Deepu S., et al. "Sentiment analysis of malayalam film review using machine learning techniques", 2015 International Conference on Advances in Computing, Communications and Informatics, ICACCI, IEEE (2015)
10. Kumar, G. P., Zachariah, C. A., Umesh, D. R., & Kumar, M. A. (2021). CloudStore: A Framework for Developing Progressive Streaming Games. In *Intelligence in Big Data Technologies—Beyond the Hype* (pp. 291-296). Springer, Singapore.
11. Abijith, P., Kumar, K. A., Allen, R. S., Vijayakumar, D., Paul, A., & Kumar, G.P., "Intelligent Cloud Load Balancing Using Elephant Herd Optimization. In *Intelligence in Big Data Technologies—Beyond the Hype*" Springer, Singapore, 2021
12. Shama Rani, Jaiteg Singh "Enhancing Levenshtein's Edit Distance Algorithm for Evaluating Document Similarity", International Conference on Computing, Analytics and Networks (ICAN) 2017
13. Nair Deepu S., Jayan Jisha P., Sherly Elizabeth "Sentiment-sentiment extraction for malayalam", 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI, IEEE (2014)

Flood Prediction and Analysis Using Machine learning

Athul Jayakumar¹, Raimon T. Simon², Amal A.M.³, Ogyen Thoga⁴

1. Computer Science and Engineering, Holy Grace Academy of Engineering, Mala, Kerala | athul27k@gmail.com
2. Computer Science and Engineering, Holy Grace Academy of Engineering, Mala, Kerala | raimonrts@gmail.com
3. Computer Science and Engineering, Holy Grace Academy of Engineering, Mala, Kerala | amalappachath2000@gmail.com
4. Computer Science and Engineering, Holy Grace Academy of Engineering, Mala, Kerala | ogyen99@gmail.com

Abstract

The changing patterns and behaviors of river water levels that may lead to flooding are an interesting and practical research area. They are configured to mitigate economic and societal implications brought about by floods. Stochastic Gradient Descent Optimization Algorithm and Back Propagation Algorithm are machine learning algorithms suitable for predicting changes in levels of river water, thus detection of flooding possibilities. The two algorithms employ similar hydrological and flood resource variables such as precipitation amount, river inflow, peak gust, seasonal flow, flood frequency, and other relevant flood prediction variables. In the process of predicting floods, the water level is the most important hydrological research aspect. Prediction using machine learning algorithms is effective due to its ability to utilize data from various sources and classify and regress it into flood and non-flood classes. This paper gives insight into mechanism of the two algorithms in perspective of flood estimation.

Keywords: *flood prediction, time series, stochastic gradient descent optimization algorithm, back propagation algorithm*

1. Introduction

Throughout the history, it has been observed that floods are one of the major disasters ever reported and can cause severe personal injuries along with the property damages. Escalation in flooding events is indeed a dilemma through recent years, as innumerable causalities are caused by them every year. One of the main reasons has been that with drastic population growth, the human settlements are getting nearer to water bodies. Thus the

infrastructure and lives associated get badly damaged and disturbed. With river overflows and flash floods. Moreover, climatic change has many consequences as surge in frequency of rainfalls potentially enhance the rate of flood in. As people have started to live in close proximity of rivers, precautions and safety measures are required to reduce the mortality and other harms associated with it. One of those measures can be the real-time flood forecasting which a challenging task is at the

same time. It combines the efforts to identify subject risk areas, modelling of relations between flooding regions, securing probabilities of flood, thresholds, and setting alarming conditions. The parametric and causal factors that embed uncertainties in flood predictions are a very perplexing task, for scientists involved in study of it. The forecasting errors or misconstructions can lead to increase damages as cost factor. Therefore, estimation of uncertainties and integrating them for development of a prediction scheme can provide efficient warning systems. Computational algorithms as neural networks have been majorly used to estimate flood in threatened area of a river and its effect outside of the specific area: for e.g. the upstream river flow or discharge is very helpful in finding downstream flows which are not equipped or lack measurements. Various researchers have reported different approaches towards forecast of water flood. Most prominent have been microwave imagery analysis: that uses land scanning through emissions received from satellite systems and on bases of parameters for an instance: Emissivity Polarization Index (EPI) threshold or Gravity Recovery and Climate Experiment (Grace) estimates extent of floods and ground level water estimations respectively. Other techniques include wireless sensors based flood forecasting that provide robust communication channelling for flood warning systems. In an extensive aspect hydrological modelling of flood prior to estimation require decision support systems that calculate flood level assessment, topographical categorization, catchment characteristics, resources management, data availability and abilities to run a flood model. Apart from compiling the complete information, the lead-time requirement

and uncertainties for flood warning be properly evaluated and adjusted as per requirement. These are the fundamental aspects that need to be addressed for flood modelling for wide range of applications. Machine learning (ML) is a sub-field of Artificial Intelligence (AI) whose applications as part of algorithmic and heuristic approach tend to understand patterns in certain dataset through intuitive training. Not more than decades ago prediction and forecasting was part of fiction, but as truth is more magical, the improvement and research in machine learning made the field emerge from laboratories into practical applications and industries. Today, health-care associated detections, Natural language processing, fraud detection, face recognition, stock trends analysing, natural hazard (flood, earthquake) estimation all are attributed to ML algorithms. The learning behaviour as anticipated by Arthur Samuel in 1959 defined Machine Learning as it gives "Computers the ability to learn without being explicitly programmed". Conversely in flood prognosis, using such techniques allow scientists and relevant flood authorities to simulate the occurrence and the expected magnitude of the impending flood. The statistical modelling of the flood problem with basic linear regression helps in predicting the behaviour of a system by programming an equation for fitting the data in a successive manner. However, all hydrological and metrological influxes systems are non-linear in nature. That means that they cannot be logically or smoothly interpreted from one stage to another. Even if they are assumed to be linear, this estimation will lead to compromise majorly on accuracy and error detection. With computational advancement and algorithms enhancements, Machine Learning (ML) has emerged as preferable

instrument to delve with non-linear systems and explore automatically generated predictions of flash floods too for instance. In this paper we are going to discuss how prediction strength of ML algorithms Stochastic Gradient Descent Optimization Algorithm and Back Propagation Algorithm.

2. Literature Survey

First, In [1] this paper proposed an IOT based water monitoring system that measure water level in real time. The prototype is based on idea that the level of water can be very important parameter when it comes to the flood occurrences especially in disaster prone area. A water level sensor is used to detect the desired parameter and if the water level reaches the parameter the signal will be freed in real time to social network like Twitter. A cloud server was configured as data repository. The measurements of water level are displayed in remote dashboard. The proposed solution with integrated sensory system that allows inner monitoring of water quality. Alerts and relevant data are transmitted over the internet to a cloud server and can be received by user terminal owned by consumer. The outcome of water measurement is displayed in web based remote dashboard. Presents a neuron-fuzzy controller based on flood monitoring system using wireless sensor network. The distributed sensor nodes used IEEE 802.15.4 protocol, to collect sensor information such as water level data from the river. The Sensor information is sent to distributed alerts center via Arduino microcontroller and X Bee Transceiver. At the distributed alert centre, X Bee transceiver and Raspberry pi microcomputer are used to generate flood alert based on sensor information and to detect flood data and this data are

stored in database. This is not cost effective system. And performance also weak as compared to our system.

In [2] this paper has proposed an effective approach to flood forecasting based on the data-driven method. The LSTM neural network model was constructed and carefully assessed to forecast one-day, two-day, and three-day flood flow at the Hoa Binh Station on the Da River. Contrary to requiring a variety of input data such as land-use and topography for rainfall-runoff simulation, the developed model uses only the measured data information available at the target station and the upstream meteorological and hydrological stations to forecast the flow rate at the target station for multi-step output. The LSTM model has learned long-term dependencies between sequential data series and demonstrated reliable performance in flood forecasting. Two scenarios have been considered to assess the model's performance in flow rate forecasting and the influence of input data characteristics on the model's flood forecasting capability. The model was validated and tested by applying criteria such as NSE value, RMSE value, as well as relative error values in forecasting the maximum flow rate using the independent data sets. For validation and testing phases, both scenarios illustrated equally good performance. Additionally, there is no substantial difference in the simulation results for all three discharge forecast cases. However, when considering the flood peak forecasting factor, the second scenario indicated slightly better-forecasted results.

In [3] this paper design early flood detection and control monitoring system studied, which should record and give proper warnings in real time. This paper presents a

technique which use image acquisition unit for image capturing in real time and transfer the data via ZigBee and GSM networks to remote host, which will finalize a decision by comparing with existing data. It will lead to automatic generation of warnings. The result of this system is obtained from the real time graph of SIFT algorithm in the flood monitoring and detection system built by Net Beans java software. From this graph the rank of calamity can identify earlier. If the yellow curve "L" rank is above the normal rank then it will change the status of the left hand side table to "Danger". This will in turn change the expert opinion status to "Danger" after adept decision. Then the system will alert us with a beep sound of alarm. Thereby we can evacuate the nearby peoples earlier. This system has the advantage of low power, user friendly programming, economic, real time in nature. This system has application in the area dam resident evacuation, forest fire detection, air pollution detection, temperature monitoring, water purity monitoring. Future scope of this paper makes it more relevant in nature.

In [4] this paper Flood detection system has floods victims within particular area. Four machine-learning algorithms were utilized to classify flood data. Random Forest algorithm provides a significant improvement in terms of accuracy over other classification with 99.5 % accuracy in comparison to 97.7% using Bagging algorithm. Moreover, Decision Tree achieved 94.6% accuracy, which is slightly lower than Bagging. While, Hyper Pipes algorithm obtained 89.8%, which is considered the lowest accuracy percentage acquired among other classifiers. Been designed for immediate notification to the local authorities. It determined the current water level using sensor network, which provides notification via SMS

and web base public network through GSM modem. SMS and web base public network are valuable alert communication tools that can distribute the information to the user Instant .

In [5] this paper the Video Surveillance system for Flood Disasters provides the most accurate flood warning information to disaster relief units and citizens, in order to reduce the negative impacts of the flood disasters. It also has other advantages, such as it avoids the common needs of a standard water-level ruler, flexibility in selecting location, and relatively a large field of view, when compared with other water-level measurement techniques. On the other hand while using sensors the problem arises with the selection of right sensor for measuring water level which may largely depends on the site and the data required by flood models. However, these works do not discuss the dissemination of flood information to the general public. Thus an application named Flood Alert is developed as an Alert Messenger using Android and web based Java in order to provide warning and alert to people during flood instant.

In [6] this paper Thus the smart flood disaster prediction system is relevant in terms of actual deployment and reliability with real time monitoring and updating of environmental parameters and prediction of flood as compared to existing approaches. The integrated approach combines the scalability of IoT and reliability of artificial neural networks to handle data provided by a sensor network and by effective communication between these two components, an early prediction of flood is done. After the experimentation, it has been proved that Leven berg Marquardt training algorithm with NARX network gives

better results and provides real time flood prediction with step-ahead alert. The prediction chart indicate that the event of flood is predicted one time-step ahead and warns the communities at risk using Thing Tweet in simple to understand language with the use of Internet. Thus by harnessing latest technological disruptions such as IOT and machine learning, big data, predictive analytics along with social media and mobility allows effective emergency & disaster management for smart nations.

In [7] this paper based on the development of a smart flood monitoring system using ultrasonic sensor with NodeMCU and Blynk application which also involving buzzer, LED and LCD display. By using the Blynk application, the information can easily be accessed by the person-in-charge. Transmission of flood information begins at first NodeMCU then transmitted to the second NodeMCU using the Blynk application. Early notification on flood monitoring can be seen through the LED and buzzer. The database of this system is based on the history graph in Blynk application and also in excel format form. During the outdoor test, it shows that the level of the water is in safety level due to the reading ranging from 66.1 cm to 75.2 cm. This water level of the lake is in safe mode to the nearer residential area due to the hot weather condition.

3. Acknowledgment

We would like to convey our sincere gratitude to teachers for their assistance throughout the procedure. The successful completion of this paper is only because of their proper guidance. We would also like to thank our friends, family, and all others who supported us in our work.

4. Abbreviations

Abbreviation	Expansion
AI	Artificial Intelligence
SVR	Support Vector Regression
IoT	Internet of Things
ML	Machine Learning
GSM	Global System for Mobile Communication
LSTM	Long Short-Term Memory
EPI	Emissivity Polarization Index

Fig 1.0: Abbreviations

5. References

1. A Data Science Methodology Based on Machine Learning Algorithms for Flood Severity Prediction- Mohammed Khalaf, Abir Jaafar Hussain, Dhiya Al-Jumeily, Thar Baker, Robert Keight, Paulo Lisboa, Paul Fergus, Ala S. Al Kafri.
2. A Real Time Solution to Flood Monitoring System using IoT and Wireless Sensor Networks- Sonali Patil, Jija Pisal, Aishwarya Patil, Siddhi Ingavale, Prajakta Ayarekar, Prof. Mrs. Shagupta Mulla.
3. Application of Long Short-Term Memory Neural Network for Flood Forecasting- Xuan-Hien Le, Hung Viet Ho, Giha Lee, Sungho Jung.
4. Real-time WSN Based Early Flood Detection and Control Monitoring System- Tibin Mathew Thekkil, Dr.N.Prabakaran.
5. Flood prediction using flow and depth measurement with artificial neural network in canals- Nikhil Binoy C, Arjun N, Keerthi C, Sreerag S, Ashwin H Nair.
6. Flood Alerts System with Android Application- Mohamad Nazrin Napih, Ismail Ahmedy, Mohd Yamani Idna Idris, Md Asri Ngadi.
7. Advance Flood Detection and Notification System based on Sensor Technology and Machine Learning Algorithm- Mohammed Khalaf, Abir Jaafar Hussain, Dhiya Al-Jumeily, Paul Fergus, Ibrahim Olatunji Idowu

A Comparative Review of Various Approaches for Malayalam Text Summarization

Jomiya Joju C.¹, Rameela Ravindran K.²

1. MSc Scholar, Department of Computer Science, St. Thomas' College (Autonomous), Thrissur, Kerala | jomiyaajuc@gmail.com
2. Assistant Professor, Department of Computer Science, St. Thomas' College (Autonomous), Thrissur, Kerala | k.rameela@gmail.com

Abstract

Text summarization is a way to condense the large amount of information into a concise form by the process of selection of important information and discarding irrelevant and redundant information. With the huge amount of data available over the internet, the relevance of text summarization has increases. Summarization helps to gain required information in less time. It's helpful for information retrieval systems, question answering systems, medical domain and news domain etc. Even though a lot of text summarization systems have been developed for summarizing documents in various languages, there is no such well performing system for Malayalam. This paper is an attempt to summarize and present the view of Malayalam text summarization from different aspects through Maximal Marginal Relevance (MMR), Graph Reduction approach using MST, Graph Theoretical approach etc. This paper represents a comparative study of these three papers in the basis of accuracy.

Keywords: *maximal marginal relevance, graph reduction, minimum spanning tree, text summarization*

1. Introduction

Text summarization is a process of extracting or collecting important information from original document and presents as a consolidated summary. It is one of the most important chores of Natural Language Processing as it helpful for information retrieval systems, question answering systems, medical domain, news domain and data mining applications etc. Summarization is reduced content by size, so preservation of time in understanding

the content is achieved. Reduced content is easily understandable by the reader. A good summary system preserves the overall content of document while keeping redundancy to a minimum. Summarization tools may also search for headings and other markers of subtopics in order to identify the key points of a document. Microsoft Words

AutoSummarize function is a simple example of text summarization.

Malayalam is the official language of Kerala and there are around 33 million people who speak Malayalam. There is a vast amount of online data available in Malayalam. It demands creating a tool that can be used to explore digital information presented in Malayalam language.

The main issues faced in Malayalam text summarization are due to the following facts:

1. Malayalam is morphologically rich - Recognizing root of words is hard and it requires morphological analyzers to perform it.
2. Absence of Capitalization feature - Capitalization is not found in Indian languages.
3. Ambiguity - The vagueness between common and proper nouns makes it hard to make sense of the words.
4. Lack of Resources - There are no automated tools available in Indian languages to perform preprocessing tasks required for summarization such as POS Tagging, Chunking, NER Classifier and Parser etc.
5. Lack of annotated data - There is no easy availability and access for annotated corpus in the community. This paper is an attempt to compare various approaches for Malayalam text summarisation.

2. Various Summarization Methods

Text summarization approaches can be broadly divided into two groups: extractive summarization and abstractive summarization. Extractive summarization extract important sentences or phrases from the original documents and group them to produce a summary without changing the original text. Abstractive summarization consists of understanding the source text by using linguistic method to interpret and examine the text.

Abstractive methods need a deeper analysis of the text. It has the ability to generate new sentences, which improves the focus of a summary, thus reduces its redundancy and keeps a good compression rate.

Various methods can be used to implement these approaches for summarization. Graph theoretical approaches for summarization represents a document as an undirected graph, in which the nodes represent the sentences in the document. Two nodes in the graph are connected if the cosine similarity of the sentences corresponding to the nodes is above some particular threshold. The sentences corresponding to the nodes with the highest cardinality are considered important and are included in the summary.

Semantic graph based approaches extracts semantic triplets (Subject-Object-Predicate triplets) from each of the sentence in the document are used to generate a graph of the document. A sub-graph of this graph is selected using machine learning techniques and the sentences in the subgraph are used to form the summary.

Maximal Marginal Relevance is a method for summarization works better for longer documents. This strives to reduce redundancy while maintaining query relevance in re-ranking retrieved documents and in selecting appropriate passages for text summarization.

3. Literature Review

A. Maximal marginal relevance based Malayalam text summarization with successive thresholds

It proposes a Malayalam text summarization system which is based on MMR technique with successive threshold. Here the sentences

are selected based on the concept of maximal marginal relevance. The key idea is to use a unit step function at each step to decide the maximum marginal relevance and the number of sentences present in the summary would be equal to the number of paragraphs or the average number of sentences present in the text document.

Successive threshold approach

In this approach, the total number of paragraphs and sentences in the given text document is counted and if the total number of paragraphs in the input text document meets a threshold value, then take the value of $_n$ as number of paragraphs otherwise take $_n$ as average number of sentences in the input document.

The proposed system uses the following algorithm. The algorithm consists of two sections; the first section uses a unit step function that identifies the maximum marginal relevance that is the relevant sentences from the input document. Then the next section uses a successive threshold approach. By using this approach the total number of sentences in the final summary can be calculated. Process is explained below:

Input: Malayalam document

Output: Summarized document

1. Input a document to be summarized
2. Now the document is traversed and eliminates the words that are not useful (stop word removal).
3. Identify the most important word (by meaning) with the help of a Malayalam dictionary from the input document.
4. Starting with the starting position of the sentence until the document finishes
5. Using the unit step function and dictionary calculate the first level important sentence from the document by using the important word identified in step 3 The unit step function used in the algorithm is given as:
6. Then the second level sentence is identified using the first level sentence and the

$$u_k + 1 = \arg \max (Sim1(u_i, Q) - \max(sim2(u_i, u_j)))$$

Where

Q : User input document

u_i : Most important word/sentence

u_j : Remaining sentences in the document

U : Selected list of sentences

Dataset	Precision	Recall	F- Measure
Dataset1	0.535	0.565	0.543
Dataset2	0.5407	0.5805	0.5785
Dataset3	0.5917	0.6743	0.6537
Dataset4	0.6779	0.7896	0.7549
Dataset5	0.673	0.7826	0.7775
Dataset6	0.852	0.8910	0.8018

Fig 1.0

- dictionary by using the unit step function
7. The next level sentence is identified using the sentence identified in step 6 and the dictionary by using the unit step function
8. Repeat the step 7 until an appropriate number of sentences is in U. Which can be achieved by using successive threshold approach
9. Stop

This method is implemented on 6 different datasets of different sizes.

B. Minimum spanning tree based graph reduction approach for Malayalam text summarization

It proposes an extractive Malayalam summarizer which reduces redundancy in summarized content and meaning of sentences are considered for summary generation. A semantic graph is created for entire document and summary generated by reducing graph using minimal spanning tree algorithm. The summarizer system can be divided in to mainly three modules. (i)Pre-processor (ii)Semantic graph generator (iii)Semantic graph reducer.

(i). Pre-processor

The steps involved are mainly divided into (a) Sentence extraction (b) Removing special characters (c) stemming (d) stop word removal.

Algorithm: Pre-processing

Input: Entire text Output: Pre-processed content

1. Start
2. Extract each sentence from the input
3. Remove special characters from input
4. Perform Stemming
5. Perform Stop Word Removal

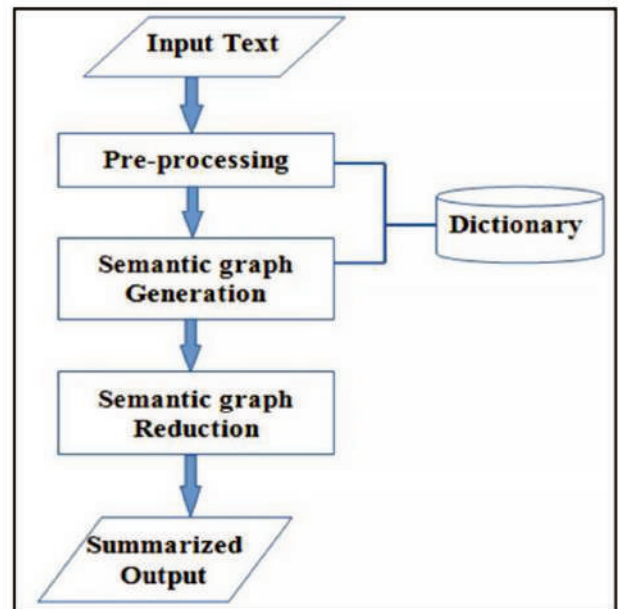


Fig 2.0

6. End

After pre-processing the input, a semantic graph is created based on the relationship among words. In the graph, each node represents sentence and the edge represent semantic relationship among them. Based on the relation each edge is annotated with a weight and rank of a node is the sum of weights of edges connected to a node. The weight is calculated based on number of common words, synonyms, antonyms other semantically related words shared among them. Spanning trees are created and the node with high ranking is placed in the output also these related nodes are removed from the graph. This process does iteratively till output becomes 25% of input.

(ii) Semantic graph generation

Semantic graph generator module is

responsible of constructing semantic graph G with the sentences S_1, S_2, \dots, S_k as vertices. Consider two arbitrary vertices S_i and S_j in the graph where $S_i \neq S_j$ is connected with an undirected edge if any of the below two cases are satisfied. Case1: Same word occur in both sentences S_i and S_j . Case 2: A word W_m in S_i is the synonym, antonym, hyponym or other Nym word W_n in S_j . The semantic graph is reduced by means of repetitive sub graph creation. It is achieved by introducing Minimum Spanning Tree (MST) concept.

(iii) Semantic Graph Reduction

This module is the core part of the system which actually do the summarization process. For this purpose a prefixed variable sum_factor is used. The variable sum_factor determines the size of summarized output according to the input size. For example: let $sum_factor = 0.1$ is the summarizer system will produce summarized output which is having 10% size of input text.

Algorithm: Semantic graph reduction

Input: Semantic graph Output: Summarized output

1. Start
2. For every node S_i , a. find $inv_rank_i = 1 / rank(S_i)$
3. Repeat
 - b. construct minimum spanning tree of the graph G
 - c. find a node S_i such that S_i has maximum inv_rank
 - d. using node S_i split graph in to G' and $G = G - G'$
 - e. select a node S_k in G' with minimum

inv_rank

4. Place S_k to output
5. Until (no of subgraphs \leq fraction of size of input for each subgraph G')
6. End

The variable sum_factor is going to use as a termination condition variable and the inverse of rank of each node S_i is calculated for each node and stored in the variable inv_rank_i . inv_rank_i is used to create minimum spanning tree (MST). The MST calculation assists to divide graph G into G' and $G = G - G'$, where G' contains the node S_k with minimum inv_rank value which implies that the sentence S_k is the most relevant in the context and it will be present in the output. Perform this action on G repeatedly till sum factor criteria is satisfied. This method works on the semantic layer ie, more than language concepts, meaning is considered here for finding similarity.

Disadvantage of this mechanism is that consideration is only for root words not similarity. Consider the case that top N and top $N-1$ ranked sentences shares common root words, which implies that the meaning of these sentences are same. Placing two sentences with similar meaning are not encouraged to place in the output because it leads to redundancy. Here, Reduction is done by using Prim's algorithm. One of the main advantages of the proposed system is that redundant information

Dataset	Precision	Recall	F- measure
Dataset1	0.720	0.818	0.621
Dataset2	0.571	0.667	0.615
Dataset3	0.600	0.750	0.667
Dataset4	0.400	0.500	0.445

Fig 3.0: Parameter Evaluation of Proposed Method

in the summary can be avoided by checking the synonym of different words concerned to a particular sentence.

C. Malayalam text summarization using graph based method

The graph theoretic approach extracts the semantically similar sentences. The similarity is determined by the different similarity scoring approaches like cosine similarity, longest common sub sequences, Levenshtein similarity etc. The relation of node is represented by the feature score of sentences.

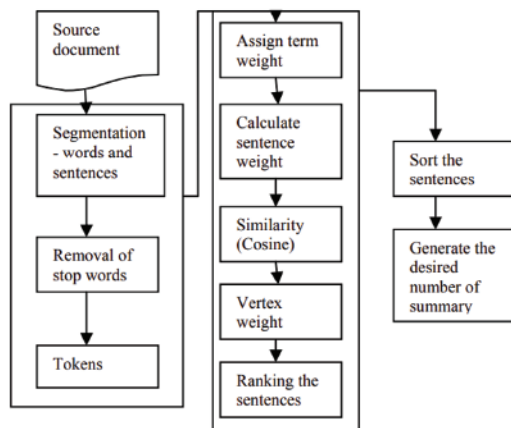


Fig 4.0: Architecture of Proposed System

Algorithm for Malayalam Text summarization

1. Input the .txt files.
2. The sentence and word tokenizer () split into sentences and words.
3. Removes the characters such as (,) . ! etc,
4. Content words can be compared to stop words list. If the word is included in the stop word list move to next word.
5. If it is not a stop word placed in word

dictionary and also keeps the sentence number.

6. Calculate the affinity weight of sentences
7. Calculate the sentence weight.
8. Calculate cosine similarity between the sentences
9. Calculate the vertex weight.
10. Rank the sentences on the basis of vertex weight and sentence weight.
11. Preset a threshold and extract the desired

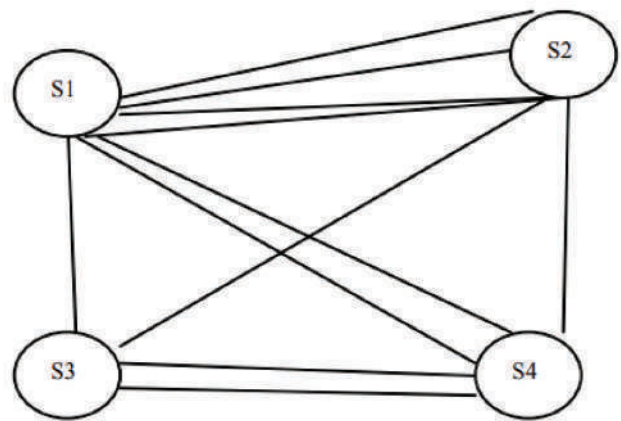


Fig 5.0: Graphical Representation of the Sentences

number of sentences.

Ranking of sentences using the graph theoretic method as explained below: Affinity weight of

$$S1 = (1/47 + 2/47 + 4/47 + 1/47 + 1/47 + 2/47 + 2/47 + 1/47 + 1/47 + 1/47) = 0.32$$

$$S2 = (1/47 + 2/47 + 2/47 + 1/47 + 1/47 + 1/47 + 1/47 + 1/47 + 2/47 + 4/47 + 1/47 + 1/47 + 1/47 + 1/47) = .4$$

$$S3 = (1/47 + 1/47 + 1/47 + 1/47 + 4/47 + 1/47 + 1/47 + 1/47 + 1/47 + 1/47 + 1/47 + 1/47 + 1/47 + 1/47 + 1/47 + 1/47 + 1/47 + 1/47 + 1/47 + 1/47) = .4$$

$$S4 = (2/47 + 2/47 + 2/47 + 4/47 + 1/47 + 1/47 + 1/47) = .26$$

Sentence weight of S1= $0.32/10=0.032$

S2= $0.4/10=0.04$

S3=0.04

S4= $0.26/10=0.026$

Similarity (s1,s2)= $14-6/14=0.6$

Similarity(s1,s3)= $16-8/16=0.5$

Similarity(s1,s4)= $10-8/10=0.2$

Similarity(s2,s3)= $16-13/16=0.1$

Similarity (s2,s4)= $14-13/14=0.07$

Similarity(s3,s4)= $16-12/16=0.3$

Rank of Sentences are:

S1= $0.032+.6=0.63$

S2= $0.04+.6=0.64$

S3= $0.04+.5=0.54$

S4= $0.02+.3=0.32$

Dataset	Precision	Recall	F-score
Dataset1	0.40	0.50	0.44
Dataset2	0.40	0.80	0.53
Dataset3	0.47	0.58	0.52
Dataset4	0.60	0.60	0.60
Dataset5	0.56	0.33	0.42
Dataset6	0.63	0.54	0.58

Fig 6.0: Graphical Representation of the Sentences

The result shows that generated summaries 51% of the sentences are semantically similar with human generated summaries

4. Discussion

There is no single evaluation scheme to evaluate all aspects of summary. So, a combination of evaluation methods are used to evaluate summary. Mainly two methods are used for

summary evaluation such as intrinsic and extrinsic evaluation. The intrinsic evaluation predicts the quality of summary based on content and co-selection measures. The co-selection measures are Precision, Recall and F-score. Precision score shows the fraction of the sentences chosen by the humans and selected by the system are correct. Recall score shows the fraction of the sentences chosen by humans is recognized by the machine. F-measure is computed by combining recall and precision.

Measure			
Method	MMR	MST	GRAPH BASED
Precision	0.852	0.720	0.63
Recall	0.891	0.818	0.54
F-measure	0.8018	0.621	0.58

Fig 7.0: Accuracy

From these three papers, Maximal marginal relevance-based Malayalam text summarization with successive thresholds provides higher accuracy. The accuracy is different according to the size and type of dataset taken.

5. Conclusion

Text summarization is growing as sub – branch of NLP as the demand for compressive, meaningful, abstract of topic due to large amount of information available on net. Thus, text summarization is the need of the hour and used by business analyst, marketing executive, development, researchers, government organizations etc. This paper discusses various methods for abstractive and extractive text summarization. As abstractive summarization requires more learning and reasoning, it is bit complex then extractive

approach but, abstractive summarization provides more meaningful and appropriate summary compared to extractive. The abstractive summarization requires heavy computational models for language generation. Such a situation the extractive text summarization produces the summary within the limited time. The combination of abstractive and extractive summarization provides a good result. Maximal Marginal Relevance (MMR) based methods are very efficient in Malayalam text summarization.

6. Future Scope

Results can be improved by incorporating a good data dictionary and a strong corpus which is currently lagging for Malayalam language. Also, the effectiveness of summary can be increased by adding some advanced features such as sentence-to-sentence cohesion, biased words etc. Through the study it is also observed that very less work is done using abstractive methods on Malayalam, there is a lot of scope for exploring such methods for more appropriate summarization.

7. Acknowledgement

I express my deepest gratitude to Dr. Jiju A Mathew, Head of the Department for providing me with guidance and facilities for preparing and presenting the paper. I am

thankful to my guide Mrs. Rameela Ravindran K for her valuable guidance and encouragement during the course of preparation of this paper.

8. References

1. Ajmal E B1 and Rosna P Haroon "Maximal Marginal Relevance based Malayalam text summarization using successive thresholds" International Journal on Cybernetics 5 & Informatics (IJCI) Vol. 5, No. 2, April 2016 DOI:10.5121/ijci.2016.5237 349
2. Rahul Raj M and Rosna P Haroon "Malayalam Text Summarization: Minimum Spanning Tree Based Graph Reduction Approach"
3. Kanitha D K, D. Muhammad Noorul Mubarak and S.A. Shanavas "Malayalam Text Summarization Using Graph Based Method" International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 9 (2), 2018, 40-44
4. Ajmal E.B and Rosna P Haroon "An Extractive Malayalam Document Summarization Based on Graph Theoretic Approach" 2015 Fifth International Conference on e-Learning
5. RENJITH S R and SONY P "An automatic text summarisation for Malayalam using sentence extraction" International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106, Volume-3, Issue-8, Aug.-2015
6. Kavya Kishore , Greeshma N Gopal and Neethu P H "Document summarisation in Malayalam with sentence framing"
7. Kanitha D K , D. Muhammad Noorul Mubarak 2 and S. A. Shanavas "Malayalam Text summarization Using Vector Space Model" International Journal of Engineering and Techniques - Volume 4 Issue 2, April-2018 ISSN: 2395-1303 <http://www.ijetjournal.org> Page 918

Subscription Information

Convergence - International Journal of ICT Academy of Kerala is a peer reviewed International Journal published annually by ICT Academy of Kerala, Technopark, Trivandrum, Kerala.

ANNUAL SUBSCRIPTION FEES	
Non-Members Institutions	Rs. 500/-
Members Institutions	Rs. 300/-
Faculty Members & Students	Rs. 200/-

Ordering Information

If you want to order multiple copies of the Journal, please contact:

The Chief Editor
Convergence, ICT Academy of Kerala
GF-1, Thejaswini, Technopark Campus
T'puram, Kerala - 695 581

Format & Guideline for Publishing the Article - ICTAK Journal

Cover page: The manuscript should be accompanied by a cover page containing the article title, the short title (not more than 5 words and which may be used in all correspondence), the names and affiliations of all the authors (specify order), along with their postal address, phone and fax numbers, and email address. Details of the authors' name and affiliation should not appear elsewhere in the manuscript. In the case of multiple authors, the cover page should indicate the designated corresponding author.

Second Page: The second page should contain the article title, the short title, the abstract (not more than 100 words), keywords (a maximum of 8 keywords), and an extended summary (not exceeding 300 words).

Body of the article: The recommended length of papers is 8000 - 10000 words, inclusive of tables and figures. Materials may be formatted in Times New Roman, font size 12 and double spaced. All tables and figures are to be serially numbered, sequentially following references to them in the text. All tables and figures are also to be presented in a separate WORD document and file names should clearly specify the paper to which the exhibits belong. All tables and figures should be in black and white only.

ICTAK Journal follows both British & American spelling wherever possible, explanatory theories/concepts and other background material of a historical or collateral nature, and case illustrations/anecdotal applications should be presented in text boxes to ensure they do not interfere with the flow of the main text.

References

Authors must acknowledge all the sources they have drawn upon, including direct quotations, as well as ideas, concepts, data, and exhibits. Only those references cited in the main text should be listed in the reference list. Source should be stated briefly in the text, following the author-date convention of by

the last name and the year of publication, in parentheses. Citations within the text would read, for e.g. 'According to Pawlak, (1991)...' or '... (Pawlak, 1991)'. These citations should be amplified in a list of references appearing at the end of the paper. The reference list should be in alphabetical and chronological order, and should include complete bibliographical details, as appropriate-the names(s) of the author(s), year of publication, title of the article/book, name of the journal, details of the publisher, volume and issue number, and individual page numbers, URL of online sources (online journals, magazines, or newspapers) with access date.

The prescribed style of citation is as follows:

Sample book references

Pawlak, Z. (1991). *Rough sets: Theoretical aspects of reasoning about data*. Norwell, MA: Kluwer Academic Publishers.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., & Tatham, R.L., (2006). *Multivariate data analysis* (6th ed.). New Jersey: Pearson Prentice Hall.

Sample reference to chapter in book

Ravallion, M. (2007). Transfers and safety nets in poor countries: Revisiting the trade-offs and policy options. In V. Abhijit, R.B. Banerjee, & D. Mukhejee (Eds.), *Understanding poverty* (pp. 203-230). Oxford University Press.

Copyright & Permissions

Authors must cede copyright of the article as finally published to ICTAK JOURNAL if it is accepted for publication, and certify that all copyright requirements in respect of material used directly or indirectly in the article have

been duly met. Copyright rests with ICTAK JOURNAL in respect of the material submitted for its use and dissemination in any form or medium, individually or in any collection or other configuration, print, audio-video, electronic or otherwise. ICTAK JOURNAL however grants permission to authors for using the submitted material (subsequent to publication in ICTAK JOURNAL) in any printed books or other publications or derivative works authored or co-authored by them. All other usage will be subject to prior written permission of ICTAK JOURNAL.

Further Details

Any correspondence relating to editorial matters and print subscriptions may be addressed to:

The Chief Editor
Convergence, ICT Academy of Kerala
GF-1, Thejaswini, Technopark Campus
T'puram, Kerala - 695 581
Email: editorial@ictkerala.org

Disclaimer: The editors reserve the right to accept or refuse an article for publication, and they are under no obligation to assign reasons for their decision.



INFORMATION & COMMUNICATION TECHNOLOGY ACADEMY OF KERALA ICTAK

Over the years, ICTAK has been pioneering its training model both for online and offline modes. We have created a unique hybrid model called Positive Engagement in Real-Time Learning (PERL) that has produced over 4.5 positive feedback ratings for the training conducted.

Microskill Programs

- Python Programming
- Java Programming
- C# Programming
- Programming with JavaScript
- Advanced Excel
- Social Media Marketing & SEO
- Business Intelligence with Excel & Tableau
- Machine Learning & Artificial Intelligence
- Data Analysis with R
- Front-end Application Development with Angular
- Front-end Application Development with React
- Robotic Process Automation with UiPath
- Ethical Hacking
- Software Engineering using Agile Scrum
- Problem Solving using Design Thinking



For more details
<https://ictkerala.org/courses>



ICTAK

ICT Academy of Kerala (ICTAK) is a Social Enterprise created in a Public-Private Partnership model (PPP) for imparting ICT skills to the youths of Kerala and improving their employability opportunities in the Industry. The Company is supported by Govt. of India, partnered by Govt. of Kerala, and the IT industry.

Long Term Training

- Certified Specialist in Data Science & Analytics
- Certified Specialist in Full Stack Development (MEAN)
- Certified Specialist in Robotic Process Automation
- Certified Cyber Security Analyst
- Certified XR Associate
- Certified Specialist in Digital Marketing
- Certified Specialist in Full Stack .NET Development
- Certified Specialist in Software Testing
- Certified Specialist in Full Stack Development (MERN)



Information & Communication Technology Academy of Kerala

(A Govt. of India supported, Govt. of Kerala partnered Social Enterprise)

GF-1 Thejaswini Building, Technopark Campus, Thiruvananthapuram, Kerala 695581

+91 75 940 51437 | 471 270 0811