Proceedings of the 2nd National Conference on

Computational Methods, Data Science and Applications

(NC-CMDSA 2021)

Editors
PROF. ABDUL WAHID
DR. PRADEEP KUMAR

MAULANA AZAD NATIONAL URDU UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
SCHOOL OF TECHNOLOGY

Proceedings of the 2nd National Conference on Computational Methods, Data Science and Applications (NC-CMDSA 2021) ISBN: 978-81-953418-4-9

Prof. Abdul Wahid

Dr. Pradeep Kumar

Editors

Proceedings of the 2nd National Conference on Computational Methods, Data Science and Applications

Applications

Table of Contents

Preface

Computational science deals with mathematical modeling, numerical analysis, computational aspects of numerical methods to solve complex problems. Data is the new oil, data science is a plethora of interdisciplinary scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data. Artificial Intelligence has become an integral part of our daily lives are affecting the way we live, take decisions, travel, shop etc. Applications of data science and computational methods can be found in simple situations such as ordering food to complex ones such as performing a surgery. This book is the outcome of the planned event which garnered attention from researchers, scholars and academicians of high repute. The papers selected and presented during the 2-day national conference on computational methods, data science and applications are compiled to form this book. We hope that the readers find the book interesting and that they are able to gain knowledge and identify problems for further research. As editors, we have taken all measures to ensure that the contents are errorfree; however, if any errors are found, please mail the page numbers, and errors to the below mail address.

Email: conf.csit@manuu.edu.in

Abdul Wahid
Pradeep Kumar

About the Editors

Abdul Wahid has received the PhD degree in computer science from Jamia Millia Islamia (A Central University), New Delhi, India. He is currently a professor at the Maulana Azad National Urdu University (A Central University), where he is the Founder Dean of the School of Computer Science and Information. He is an expert member in several academic and administrative committees. His research areas are software and web engineering, e-learning, machine translation. He is a professional member of IEEE. He is also a lifetime member of the Institution of Electronics and Telecommunication Engineers (IETE), India Society for Technical Education (ISTE). He is Member of Scientific and Technical Committee & Editorial Review Board, World Academy of Science, Engineering and Technology. He has Guided 6 students to a successful Ph.D. One scholar has submitted her thesis recently, and currently 6 are under his supervision. He has two SCIE Indexed papers in high impact journals and numerous Scopus-Indexed publications.

Dr. Pradeep Kumar is currently working as an Associate Professor in Department of Computer Science and Information Technology, School of Technology, Maulana Azad National Urdu University (Central University), Hyderabad, Telangana, India. He has more than 24 years of Teaching and Research experience. He has obtained his Doctoral Degree School of Information & Communication Technology, Guru Gobind Singh Indraprastha University (GGSIPU), New Delhi. His area of research is Software Reliability, Machine Learning, Soft Computing, and Data Mining. He has published more than 21 reputed journals, 5 International Conference proceedings, 3 Book Chapters and 3 Book in his credit. He has active membership in Association for Computing Machines (ACM) India, Computer Science Teachers Association (CSTA) USA, International Association of Engineers (IAENG), International Association Computer Science & Information Technology Singapore, and Universal Association of Computer and Electronics Engineers. He is also conducted two national conferences at Maulana Azad National Urdu University.

From the Chief-Editor's desk

It is a great pleasure to be a part of this event and to be able to take part in editing this wonderful e-book which is an outcome of the successful National Conference on Computational Methods, Data Science and Applications. The event was envisioned in the month of March and conducted in May 2021. I am happy to have a dedicated team that worked day and night to make this event a success. Even in such short duration and current situations, the event has been highly successful. The conference was organized to bring together like-minded individuals to meet, network and discuss the latest research trends in computational methods, data science and their real-time applications. Data is the new oil, and data science is the science of collecting, using algorithms to clean it and study it to find patterns that can be applied to solve many real-world problems. Computational methods are everywhere, from solving simple problems like bus route monitoring, cab recommendation to predicting the moisture level in the air, identifying the level of underground water and so on. It's use cases cannot be populated as an exhaustive list. During this event we were able to listen to a variety of guest speakers and authors who presented their work in this area to solve some of the problems identified. Papers were presented on methods that could possibly be able to solve the problems that appeared in the Covid situation, classical engineering problems and their smart solutions to improvements in mental health and education using computational methods and data science.

The research in the Department of Computer Science and Information Technology, School of Technology, Maulana Azad National Urdu University has been focused on some interesting problems such as Machine translation, personalized learning, bioinformatics, road dynamics and computer vision, web semantics, disease prediction, smart computing and networking and so on. Our team of dedicated research supervisors, research scholars and faculty aim to do research that will solve societal problems. With this note and endless hope of being part of much bigger events in future, I take leave of you while presenting you this book.

Abdul Wahid
Professor, CS&IT
Dean, School of Technology
Maulana Azad National Urdu University

From the co-editor's desk

It is an honor to be a part of a team that is a pro in organizing events and keeping things together so that events are successful and remembered for a long time. The national conference on Computational Methods, Data Science and applications was planned by anticipating the amount of research ideas this conference might generate. Who knows some of those ideas might be the baseline for future research projects and international patents?

After the event was launched in the pandemic situation and even in such a short notice we received an overwhelming response both from enthusiastic keynote speakers and authors who were excited to be a part of this conference. The dignitaries lauded the hard work of the team and congratulated on bringing together fine researchers and academicians together on a single platform. Prof. SM Rahmatullah, Hon'ble Vice Chancellor I/c, commented on how the act of thinking, contemplating, and reflecting about the creations of the Lord is the best prayer. The Chief Guest, Prof. Abdul Quiayum Ansari, elaborated on the evolution of teaching methods compounding it with research in information Technology. He also encouraged students to realize their potential, and instead of being a part of the herd, to try to be an outlier. Prof. Abdul Wahid, Dean, School of Technology, reiterated the details of programs in the department and the department is always keen and interested to take up challenging tasks that would be beneficial for the University. The keynote speaker Dr. Malik Yousef, took the audience through the intriguing journey of gene classification in the simplest manner possible. The conference received papers from all over the country. After a rigorous round of review, 27 papers were selected for presentation during the conference. The research papers ranged from the latest technologies to solve the new problems due to the pandemic, to classical problems in the networking domain. Some of the interesting papers discussed engineering problems and their solutions using computational methods and data science. Without much delay, I leave the readers to these interesting research papers.

Dr. Pradeep Kumar
Convenor,
Associate Professor, CS&IT
Maulana Azad National Urdu University

Acknowledgements

About MANUU, School of Technology, Department of Computer Science and Information Technology

Maulana Azad National Urdu University (MANUU) is a central university established in the year 1998 by an Act of Indian Parliament, with all India Jurisdiction, to promote and develop Urdu language, impart vocational and technical education in Urdu medium through conventional and distance modes and focus on women education. The university is offering has 7 School of Studies, 19 departments, 8 CTEs, 5 polytechnics, 2 satellite campuses, and 14 centres and directorates for education and research

The University is named after Maulana Abul Kalam Azad, the freedom fighter and outstanding scholar, who as the first Education Minister of Independent India envisioned and laid the foundation of technical, scientific and higher education in India. School of Technology was established as School of Computer Science and Information Technology (SCSIT) in the year 2014 with the objective to create a congenial environment for inspiring young minds and faculty. The school was renamed in June 2019. Currently, the Department of CS&IT is the only Department in the School at present in addition to 5 Polytechnics located at various parts of the country viz. Hyderabad (Telangana), Bengaluru (Karnataka), Darbhanga (Bihar), Kadapa (Andhra Pradesh), Cuttack (Odissa)

In order to maximize the interaction between industry & institute the Department started AICTE approved B.tech (computer science) programme from academic year 2015. Further to enhance academic ambience and research expertise, the department started regular M.Tech (computer science) in year 2015. Department is also engaged in Research Programs and at present 15 research scholars are pursuing their research in different domains of CS&IT with an overall strength of the department is almost 350 students on roll. The backbone of the department is its Dedicated Team of Faculty Members leaded from the front none other than by our respected Dean, School of Technology - Prof Abdul Wahid, Professor in Computer Science & Information Technology, MANUU.

The main objective of the School is to impart quality education in the field of Computer Science and other Engineering Technologies and also to provide a mix of courses with a truly interdisciplinary grouping of subjects that contributes to produce skilled manpower of the quality standards.

Research and Development is an integral part of the academic activities of School of Technology MANUU Hyderabad. Following are the key research objectives of the department of Computer Science and Information Technology

RESEARCH OBJECTIVES:

o   To create zeal amongst faculty and students towards research and innovation.

o   To create centres of Excellence in niche areas of research.

o   To work closely with the industrial needs that eventually will result in new or improved products, processes, systems or services that can increase productivity of the industry.

o   To foster collaborations for mutual benefits and to maximize Industry - Institute connectivity.

o   To establish collaboration with other universities, public and private sectors and identify R&D projects including consultancy services which could be undertaken at the institution.

o   To ensure smooth functioning and effective Management of R&D at the institution.

o   To organize and promote skill development trainings in emerging areas to enhance the employability skills of the students.

o   To promote self-employability skills by conducting various activities under Centre for Innovation, Incubation and Entrepreneurship (CIIE)

Conference Report

This two-day national conference in the virtual mode, was envisioned in the month of march by none other than our Dean Sb, Prof. Abdul Wahid. despite of all the hurdles we faced, this event is a huge success and I believe, everybody here will agree to it. Even with the ongoing pandemic, we received some wonderful papers and we could bring together such esteemed academicians and leaders on one single platform. The addresses by our dignitaries have not only been motivating but also they provoked new thoughts in each one us, preparing us for the times to come. We received around 40 good quality papers from authors residing in different parts of the country. Papers have come from Telangana, Andhra Pradesh, Karnataka, New Delhi, Jammu & Kashmir, and Uttar Pradesh. The Inaugural function was held on 24th May, 2021, 11:00 am, in the virtual mode, Curious listeners and encouraging delegates were a part of the inaugural session. The chief guest, Prof. Abdul Quaiyum Ansari, Prof. Dept. of Electrical Engineering was the central figure of the event. The session began with the recitation of the Quran, playing of the tarana, then the keynote address followed by the chief Guest's address. The inaugural was followed by paper presentations,. 13 papers were presented on day 1

The second day of the conference was purely a technical day, fully packed with loads of information. The first talk by Dr. Khalid Raza, Senior Assistant Professor, Department of Computer Science, Jamia Millia Islamia, delivered the first lecture of the day on "Computational and Big Data Challenges in Bioinformatics". He unravelled the mysteries in the budding and highly researched field of Bioinformatics in simple words. The audience were captivated in trance with the discussion. It was followed by a lecture on "Application of Computational Intelligence in Engineering Problems" by Dr. Mohammad Zakwan, Assistant Professor, Civil Engineering, MANUU, Polytechnic. His topic was a beautiful amalgamation of machine learning, fuzzy logic and engineering problems. His research can be used in future to solve some of the problems commonly faced by the society. The talks were followed by parallel paper presentations. The Valedictory function was presided by Prof. SM Rahmatullah, who congratulated the Department of Computer Science and Information Technology for organizing a power-packed conference and encouraged them for future events targeting international audience also. The Guest of Honour, Prof. Salman Abdul Moiz, Professor, School of Computer and Information Sciences, addressed the gathering and spoke to the audience focusing on the need of good research in current times. The conference report was presented by Conference convener, Dr. Pradeep Kumar; after intense reviewing, 40 papers on various

topics involving current research were presented in both sessions. The day ended with networking of the audience.

In conclusion, the conference portrayed a kaleidoscope of young researchers presenting their research and there has been new areas of inter-disciplinary research were discussed. The two-day conference was a success because of the strong coordination between the faculty, teaching and non-teaching staff, research scholars and students of the department. The Engineering section, estate section and guest house also assisted in organizing the conference. We wish to organize many such conferences in the future.

Co-Editor
Dr. Pradeep Kumar,
Conference Convener,
Associate Professor,
Department of CS&IT

List of reviewers

The following is our team of reviewers who we are thankful to, for reviewing the submissions and replying well in time with their valuable comments. The comments were very helpful to the authors during the revision and preparation of the final manuscript.

1. Dr. Syed Imtiyaz Hassan, Assoc. Professor and Head, Department of Computer Science and Information Technology, Maulana Azad National Urdu University
2. Mrs. T. Arundathi, Department of Computer Science and Information Technology, Maulana Azad National Urdu University
3. Dr. Khaleel Ahmad, Assistant Professor, Department of Computer Science and Information Technology, Maulana Azad National Urdu University
4. Mrs. Khaleda Afroaz, Department of Computer Science and Information Technology, Maulana Azad National Urdu University
5. Dr. Muqeem Ahmed, Department of Computer Science and Information Technology, Maulana Azad National Urdu University
6. Mr. Mohammad Islam, Department of Computer Science and Information Technology, Maulana Azad National Urdu University
7. Mr. Jameel Ahammed, Department of Computer Science and Information Technology, Maulana Azad National Urdu University

Contents

CMDSA-028	Detection of Phishing Attacks in Web Environment Using Unsupervised Machine Learning Sheo Kumar, Rajesh Tiwari, Neethu Choudhary CMREC
CMDSA-030	Road Crack Detection And Segmentation For Autonomous Driving Alisha Raza CS&IT, MANUU
CMDSA-031	SEEK – SEAT Rahmatunnisa Khatoon Muffakham Jah College of Engineering and Technology, Hyderabad
CMDSA-032	SCHOOL – BOT Afroz Samee Muffakham Jah College of Engineering and Technology, Hyderabad

An Empirical Illustration on Student Stress: An Outlook

Zohra Khatoon[1] and Mr. Mohammad Islam[2]

Department of Computer Science and Information Technology
Maulana Azad National Urdu University, Hyderabad, India
zohrazoya25@gmail.com[1], islam@manuu.edu.in[2]

Abstract: Usually we can tell that stress is the major problem in a student's academics. In this era of life stress performs a dominant role in a student's academics because of the various internal or external expectations that were made by their parents and teachers. It comprises of various stress just like attaining the higher grade marks, matching the socio-life of the college in comparison to their batchmates, also at the same time dealing with the financial difficulties of the family. So, it creates a lot of pressure for the students to even qualify a single exam or presentation. Hence it can proved that stress is the common part of a student's academics but unfortunately few of them didn't cope with this and try some suicidal activities. As per the data from GOI, a suicidal activity were attempted in every 55 minutes. According to the latest government report for the last 3 Years, 81 student's kills themselves in premier institutes like IITs, IIMs and other central universities. In 2016 alone 9,474 suicide reported among them One-Fourth of the cases reported were due to failure at examinations. Comprably, 2,403 in 2014, 2,646 in 2015 and 2,413 in 2016 suicide was committed for the same cause. This study evaluates the stress level of 50 participating postgraduate students in the university by using GSR and PPG Signals. In the proposed model, we used 4 different machine learning methods. For each method we applied the postgraduate student stress dataset to Compare and find the most accurate method to detect stress among postgraduate students.

Keywords: *Academic Stress, Physiological Signal, SVM, K-NN, Naïve Bayes, Decision Tree*

Introduction

Stress plays a serious role in emotional and physical wellbeing and is related to many diseases together with diabetes, depression and alternative chronic diseases [11]. Stress is an emotional or physical response for any such as conditions that makes anyone feel frustrated anger or nervous in daily livings. It can be experienced by the environment, body and thoughts. Stress becomes so terrible that you get used to it and it started to feel familiar and even normal to you and you are not able to consider how much it's affecting you even it makes a count. That's why it is more important to be aware of the common warning and symptoms of stress overload.

Academic stress means that education connected to stress. Educational stress is the main source of stress faced by youth nowadays. The more stressful is the expectation that parents had for their children, which grows on and becomes a large burden on the children that could not be carried by them anymore.

Academic stress is the main reason behind anguish. Academic means facing different exams and completing assignments or presentation within the shortest period of time with lots of liabilities. In other words, we can also say that academic stress is the interconnection between students to the environment, cognitive thoughtfulness, coping Strategies and physiological stimuli towards the stressor [45]. From all the above discussion we can say that education-related stress is the main reason for stress within students.

2. Literature Survey

The following table 1 shows summaries of stress among students. In this article paper, we dig on different research work of the researchers related to stress. Out of many research works we conclude 30 papers in tabular form in which year of publication & publisher, reference or author, methodologies or tools, findings or conclusion, and research gap or future work are shown:

Year of Publication & Publisher	Reference/ Author	Methodology / Tools	Finding / Conclusion	Research Gap /Future Work
2019 EL SEVIER	[1] Ravinder Ahuja et al.	**Student Dataset:** 206 Stress, Not stress **Performance parameter:** Sensitivity, Specificity, Accuracy, K-fold **Algorithm:** LR, NB, SVM, RF **Sensor:** HRV, GSR **Language:** Python	➢ The mental stress of the students is obtained during the internet usage period and before one week of the exam. ➢ Identify increasing stress levels & predict the stress before it can affect their life well-being. ➢ SVM 85.71 better than other Algo.	➢Data is less structured. ➢Research attribute to be explored more.
2018 Scopus	[2] K. Jayasankara Reddy et al.	**Participants Subject:** 336 Commerce:93 Management:104 Humanities:55 Basic Science:84 **Statical Analysis:** T-test, ANOVA	➢ The academic result is affecting student mental health & well-being. ➢ Student stress can be reduced by implementing yoga, meditation and bio-feedback.	➢We can carry this experiment on engineering and medical students for their reported higher stress levels.

Year/Source	Ref/Author	Methodology	Findings	Future Scope
2019 IEEE Journal	[3] S.M.Umar Saeed et al.	**Education Dataset:**33 (Participation has no past or current Brain Injury) **Sensor:** EEG **Parameter:** F-Measure, MAE, RMAE **Algorithm:** K-NN, NB, SVM, LR, MLP	➢ Classify a person's long-term Stress using EEG for labeling of stress & Control group PSS score & expert Evaluation. ➢ For Stress Classification Alpha symmetry is used for feature and potential biomarkers. ➢ SVM & LR Algorithm Accuracy is 85.20%.	➢ Small Dataset. ➢ More classifier Algorithms can be used. ➢ Various Age group could be included to improve accuracy.
2019 IEEE	[4] Ahmed A. Marouf et al.	**UG(CS)Dataset:** Bangladeshi student: 169 Low stress: 30 High stress: 18 Moderate stress:114 **ML Algorithm:** NB, DT, SLR, RF, LDA, BA, MLP, SMO, K-NN **Metrics:** Accuracy, Recall, Precision, F1 Score. **Tools:** WEKA.	➢ SMO gives an accuracy of 70%, which is higher than all Algorithms used.	➢ The accuracy of the model may be improved through the SVM Algorithm.
2019 IEEE Conference	[5] Amir Hasanbasic et al.	**Student Dataset:** 10 **ML Algorithm:** SVM, LDA **Tools:** MATLAB **Sensor:** EDA, ECG	➢ SVM Accuracy 91%. ➢ All the students who were examined were Suffering from high stress and anxiety.	➢ Small Dataset ➢ Fewer Algorithms. ➢ More Physiological sensors can be used to improve data accuracy.
2018 IEEE	[6] R. Karlo C.Pascual. et al.	**Engineering Students Dataset:** 30 subjects **Sensor:** HR, GSR, BT **ML Algorithm:** ANN Algorithm **Language:** Python **Tools:** Arduino **Software:** Raspberry Pi	➢ Detect the stress level of a student as low, moderate, high. ➢ Developed a device, which successfully detects the student stress level with an accuracy of 91.67%.	➢ More physiological signals to be used for data collection for different groups.
2018 Springer	[7] Aakash Verma et al.	**Sensor:** GSR, Pulse Sensor **Hardware:** Arduino Lilypad, OLED display	➢ Wearable device Senses and shows the patient's mind of state in color form. ➢ By using colors, the mood of the patients can be identified and useful to give them the proper medications.	➢ By using an ML algorithm, we may get better prediction or classification.
2018 IEEE Journal	[8] Mohamed Z Ramdan et al.	**University Student Dataset:** 20 men **Statical Analysis:** ANOVA **Sensor:** HRV, GSR	➢ Compared to 2D Display TV, Physical stress is highest in 3D Display TV	➢ We may also collect female datasets to get better physical stress data and we also compare the

Year	Author	Methodology	Key Findings	Future Scope
				gender differences.
2017	[9] KN.Fountoulakis et al.	**Dataset:** Schizophrenia **Patients:** 175 **Statical Analysis:** MANOVA, PCA, PCC	➤ A huge number of schizophrenic patients experienced maniac or depressive symptoms. ➤ Lower Depressive Rate in Female compared to males.	➤ By using various physiological sensors, we may get better data results.
2017 ELSEVIER	[10] Sriramprakash. S et al.	**SWELL-Kew Database**: 25 Subjects **Sensor:** ECG, GSR **ML Algorithm:** SVM, linear, SVM, RBF, K-NN	➤ A stress detector hardware device has been created that can determine the stress of working people. ➤ GSR, HR, and HRV are the immediate respondent of stress. ➤ SVM combining with RBF Kernel gets the highest accuracy of 92.75%.	➤ The dataset that is used consists of only 25 Subjects. ➤ Work is not the only reason for stress in working peoples it can be emotional, relational or maybe working environment.
2017 IEEE Conference	[11] Begum Egilmez et al.	**Student Dataset:** 09 **ML Algorithm**: Random Forest **Device:** Portable wrist work GSR, Android or iOS phone.	➤ Detection of stress by using a wrist-worn device that can identify the physiological response and accurately predict stress. ➤ Singing Level is the highest level compared to other Activity among College Student.	➤ A more accurate real-time stress detection model can be made by using various other algorithms. ➤ Accuracy can be enhanced by exploring more subjects or a large group.
2016 IEEE Transaction	[12] R. Subramanian Et al.	**Subjects:** 58 **Stimuli:** Video **Sensor:** GSR, ECG, EEG **Algorithm:** SVM, NB, RBF **Tool:** MATLAB	➤ ASCERTAIN database has been created that combines the personality traits and emotional states by using physiological responses. ➤ RBF and SVM give better personality trait recognition.	➤ This study is mainly person dependent.
2016 IEEE Conference	[13] IEEE (EMBS) International Student Conference, Design competition	**Parameter:** Mean, SD, Kurtosis, Skewness **Sensor:** GSR and PPG **Device:** Arduino Microcontroller & Android (BLE)	➤ Changes in biological stress are directly proportional to the heart rate.	➤ Dataset was not disclosed clearly. ➤ We do not have the accuracy of the related data. ➤ We can also use ML Algorithm.
2016 IJERT	[14] Anthonette D. Cantara et al.	**Computer Science Database:**119 Student:102 Fresh Faculty: 12 Senior Faculty: 5	➤ Computer users may also come under stress with the use of computers. ➤ Fuzzy Logic is used to create a stress detection	➤ More Physiological signals can be used. ➤ Various algorithms may be

Year/Source	Ref/Author	Methodology	Findings	Future Scope
		Sensor: Self-made ECG & GSR **Algorithm:** Fuzzy Logic **Tool:** MATLAB, Detecting data from Arduino board the C# programming language.	algorithm that gives an accuracy of 72 %.	tried to improve accuracy.
2016 Scopus	[15] Nitin Pandey et al.	**Dataset:** 30 subjects (10 each mild, moderate, severe acne) **Feature Extraction:** Mean, Standard Deviation, P-value **Statical Analysis:** T-test, ANOVA Test **Sensor:** GSR	➢ In female acne patients, physical and psychosocial stress is a major reason for acne severity. ➢ Timely stress management is necessary in these cases.	➢ We can use a machine learning tool.
2016 ACM	[16] Mohamed Abouelenien et al.	**UG&PG Student Dataset:** 50 Participants (Female: 35, Male: 15) **Sensor:** GSR, BVP, ST, BR **ML Algorithm:** Decision Tree **Tool:** MATLAB	➢ Thermal sensors outperformed physiological sensors on efficient stress detection. ➢ Data obtained from physiological sensors give better performance as compared to a single sensor's data.	➢ Data was collected in controlled lab-setting, it can be used in real-life situations to achieve higher detection rates. ➢ Data collection can be done by using other various ways such as, public presentations, quizzes & mid-term or final exams of students.
2015 Hindwai Journal of Biomedical Education	[17] Samira S. Bamuhair et al.	**Medical Student Dataset:** 378 subjects (male: 224) Premedical Year: 40% Preclinical Year: 26% Clinical Year: 33% **Statical Analysis:** T-test, ANOVA	➢ Female students discern lots of stress and stressor. Also In comparison to male students they have more coping strategies. ➢ If a student discerns high stress, then as a result coping strategy are also high. ➢ Medical Students received more stress than paramedical students.	➢ The study was bounded to one university campus. ➢ PSS scale that was used, may cause bias reporting that leads to exaggerated or underreported stress. ➢ Instead of PSS, Physiological Signals may use to collect the data for better results.
2014 International Conference	[18] A. Fernandes et. al.	**Sensor:** ECG, PPG, GSR **Device:** RemierM2000, ATMEGA 2560 Microcontroller	➢ 80% of the volunteers got physically stressed due to the exercise test while the remaining 20% were mentally stressed.	➢ We can use a machine learning tool.
2014 ELSEVIER	[19] Nandita Sharma et al.	**Dataset:** 13 Subjects (5 Males and 8 Females) **Age Group:** 16-25 Years **Sensors:** GSR and EEG **Algorithm:** SVM, GA-SVM, and ANN	➢ To represent stress signal, a curve is used that implies to a hyperbolic tangent function.	➢ Stress signals of an observer can be compared across multiple environments. ➢ Estimating the stress signal of a stressed individual

				with particular stress stimuli.
2014 International Conference	[20] L. Vanitha et al.	**Sensors:** GSR, BVP, ECG, PD, ST, EEG, FT, EMG and ET. **Classifiers:** SVM	➢ To determine stress is very important in human well-being. ➢ HRV is a reliable measure to detect stress. ➢ SVM classifiers get an accuracy of 92%.	➢ Dataset is missing.
2013 Springer	[21] Nandita Sharma et al.	**UG Student Dataset:** 35 (Male: 25, Female: 10 Age: 18year) **Sensors:** ECG, GSR, BP **ML Algorithm:** ANN, SVM, GA-ANN-SVM, GA-ANN, GA-SVM, 10-fold Cross-Validation. **Tool:** MATLAB	➢ GA hybrids produced better results to select suitable stress features and optimize parameter settings for classification models to capture stress patterns. ➢ Without GA the accuracy of ANN and SVM were 68% and 67% respectively but with GA hybrid the accuracy gets up to 89%.	➢ Reading in time segments can be include to resolve critical time segments to differentiate reading classes in terms of stress measurement.
2013 IEEE	[22] H. Kurniawan et al.	**Dataset:** 10 Subjects **Test:** SCWT, TSST, and TMCT **ML Algorithm:** K-Mean, DT, GMM, SVM **Sensors:** GSR (LEGO Mind storm NXT1), RCX wire connector Sensors. **Device:** VOICEBOX1 speech processing tools with RAPT.	➢ SVM best accuracy 92% ➢ Speech is a suitable indicator for determining stress	➢ Speech prediction is more person dependent. ➢ Relate both GSR and Speech signals we did not have more effective and accurate data.
2013 IEEE MALAYSIA	[23] Bong Siao Zheng et al.	**Dataset:** 25 Healthy subjects **Sensors:** ECG, GSR **ML Algorithm:** K-NN, Fuzzy-KNN **Statical Analysis:** One-way ANOVA.	➢ Through EMG Features we have the difference between non-emotional stress and emotional. ➢ Waveform Length feature gives us an accuracy of 70.85%.	➢ Time-frequency domain and Frequency Domain features can be extracted from EMG Signal. ➢ To improve the database effectively by adding emotional stimuli and PSO.
2012 Springer	[24] Nandita Sharma et al.	**UG CS Student Dataset:** 35 (male: 25, female: 10) Age: 18yrs **Physiological signals:** GSR, ECG, EOG, Sphygmomanometer **ML Algorithms:** GA, ANN **Tools:** MATLAB	➢ Stress in reading has been successfully classified using ANN. ➢ The GA and hybrid ANN provide a better-quality classification technique.	➢ Different ANN models and topologies may be applied to check the stress classification and investigate alternate optimization techniques.
2011 IEEE Conference	[25] Jorn Bakker et. al.	**Datasets:** 5 Subjects **Sensors:** GSR **ML Algorithm:** Fit Simple Regression Model.	➢ Finding the stress level of professional people who belongs to various domains. ➢ Detection of Stress is a very tough task & the	➢ Different Stimuli of data for stress detection, Categorization & Statics from the E-

			GSR Data Signals is highly dependent on various hidden Context.	mail, Social media, Calendar.
2011 IEEE Transaction	[26] Alberto de Santos Sierra et al.	**Student, Social work (UCM) Datasets:** 80 Female **Sensors:** GSR, HR **ML Algorithm:** Fuzzy logic	➤ A Detection system that as an accuracy of 99.5% in stress detection that uses 10s to find the stress template and 7s for stress detection. ➤ Accurate stress detection can be done only by using GSR and HR physiological signals.	➤ The detection system can be used for aliveness detection in biometric systems or can be used in ATM Machine to avoid forcefully dispensation of cash.
2011 Springer	[27] Nandita Sharma et al.	**Dataset:** 6040 Sample **ML Algorithm:** ANN **Physiological Signal:** GSR **Tool:** MATLAB	➤ In reading males and females, stress can be classified using ANN. ➤ In males and females stress is different $(p<0.01)$.	➤ GSR value may use with other classifiers like SVM for pattern recognition and investigating inconsistency in male datasets.
2008 ELSEVIER	[28] Mohammed Al-Fudail et al.	**Teacher Dataset:** 9 (Male: 3, Female: 6) **Physiological Sensor:** GSR	➤ Existence of technostress when teachers use technology in their classrooms. ➤ Technostress caused by the lack of fit increased work demands and lack of appropriate pedagogic preparation in teachers.	➤ Dataset is small which may be the consequences of the conclusion. So, we need to do larger-scale studies.
2007 Springer	[29] Armando Barreto et al.	**Dataset:** 32 subjects **Physiological Sensor:** BVP, GSR, ST, Pupil Diameter **ML Algorithm:** NB, SVM, DT **Tool:** MATLAB	➤ The SVM classifier shows the best performance on other classifiers. ➤ But if the Pupil Diameter signal were not considered, then the recognition rate falls at 58.85% in SVM.	➤ We can use the Pupil Diameter assessment instead of Webcam in Standard PC's to detect stress.
2019 IEEE Transaction	[30] Kizito Nkurikiyeyezu et al.	**Dataset:** 58 subjects **ML Algorithm:** Random Forest **Physiological Sensor:** ECG, EDA, EEG, EMG	➤ A personalized model has been presented by combining the samples of a large group to person-specific samples that act as a fingerprint and introduce as uniqueness to the new model.	➤ ML methods when used with the physiological signals, fail to conclusion and detect the stress of an unseen person. So, it is more person dependent and not appropriate for real world stress monitoring system.

From the above study we gather the knowledgeable idea about stress, its types and stress detection techniques using various parameters and physiological signals. While some of them are related to our experimental study on student stress analysis, they are either using PSS or ASS Scale for stress detection among students that have found not to be the most accurate method to detect students because participants may give the wrong entry for the questions. So here in this study we are using physiological signals like GSR and PPG with Machine Learning Algorithms to analyze stress among students.

2. Methodology

Machine Learning is an associate feature of Artificial Intelligence that gives systems the power to learn and improve from experience without being programmed explicitly. There are the following phases for analyzing stress using Machine Learning:

a) Problem Identification.
b) Data Acquisition
c) Data Pre-Processing.
d) Model Construction
e) Evaluation
f) Discovery Knowledge
g) Taking Action.

In this paper we use machine learning (ML) to analyze the stress and its level in university students and to predict the student's as highly stressed, moderate stressed and normal. That came helpful to stop the damage and suicidal activity in their life before happening.

A. Block Diagram

Figure 1 shows the sequence of approach towards the stress detection in student's academic cycle.
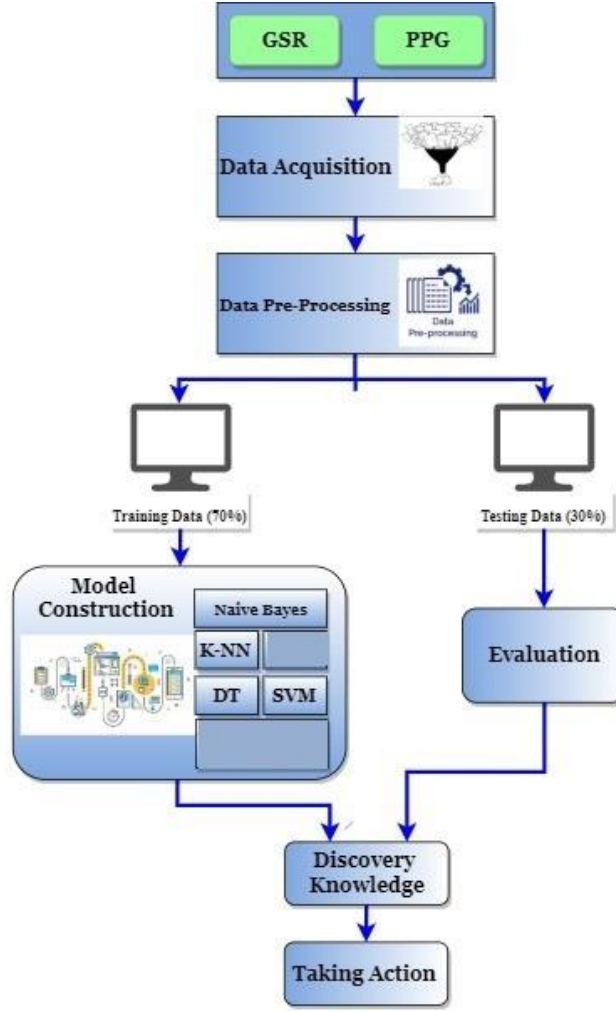
Figure 1: Block Diagram for Student Stress Analysis

B. Algorithm Study

There are too many prominent algorithms that can be used for the prediction and analysis of the stress among students.

a. Naïve Bayes

Naive Bayes, is a classifier algorithm and based on Baye's Theorem. It is a combination of multi-classifier algorithms where all the algorithms are independent of each other but all of them work for a common principle. It is a nonlinear classifier that provides Suitable outcome in real-world problems. Also, it works on the small training dataset to predict the parameters. Now, in a model where two events A and B are independent then, the result would be: [47]

$$P(y|x_1, \dots, x_n) = \frac{P(x_1|y)P(x_2|y) \dots P(x_n|y)P(y)}{P(x_1)P(x_2) \dots P(x_n)}$$

b. Support Vector Machines (SVM)

SVM is an algorithm in supervised learning that analyzes data for discrete and continuous analysis. SVM with labelled training data gives an optimal hyperspace that separates training datasets with a large margin to maximize the gap between the data of two classes. SVM creates a line between the data which separates those classes. The SVM is a most efficient classifier and it has been used widely for stress detection in GSR based studies. By using SVM we reduce the risk of data over-fitting and provide erformance.

c. K-Nearest Neighbors (KNN)

KNN is renowned as a widely used and simplest form of a classification algorithm. It is used to solve both classification and regression problems. It makes a good effect in real-life scenarios because of its non-parametric technique, it implies that it does not make any assumption regarding the distribution of data. Here a distance function is used to find a member of the training dataset, which is used to predict the classes. KNN adopts Euclidean distance to calculate the distance between an instance whose attributes values are $a_1, a_2 \ldots a_n$ (n is the number of attributes) and $b_1, b_2 \ldots b_n$ can be defined as[47]:

$$D_g = \sqrt{(a_k - b_k)^2}$$

d. Decision Tree

The decision tree is a very powerful and broadly utilized tool for classification and prediction. It is a DFD structure where each internal node shows a test performed on an attribute, each branch serves as the result of the test performed, and each leaf node has its class label. A Decision tree classifies its instances by sorting them down from the root to the leaf node, which briefs about the classification of instances [46]. An instance has been classified by testing the attributes, beginning from the root node of the tree and moving beneath to the tree branch relating through the value of the attributes. Similarly, this process is repeated for the subtree rooted at the new node.

2. Dataset Collection

This paper performs the empirical illustration of student stress towards their academic cycle. The experiment has been

Conducted with 50 subjects of postgraduate students among them 20 were male and 30 were females in the age boundary between 20-28 Years. All the participants are from MANUU University, Hyderabad in their postgraduate level of study. The experiment is based on few questions of their academics like Worry about exams, teacher's expectation, poor interest in subjects, lack of concentration during study hours, difficulty in remembering course materials, hesitation on asking doubts, difficulty in public speaking, incomplete and confusing study material, etc. During this interview we are using physiological signals like Galvanic Skin Response and Photoplethysmogram, which has to be attached with each subject on their left hand to measure the physiological signal designate by the electrical conductance of skin variability which is related with the student's response towards their academic stress.

3. Experimental Result

In classification the performance of the model is measured by using sensitivity, specificity, precision and F-Score.

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Speccifity = \frac{TN}{TN + FP}$$

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

$$F - Score = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

From the above equation we can calculate the metrics, where TP = True Positive, TN = True Negative, FP = False Positive and FN = False Negative[46].

The proposed method has been tested on 50 datasets of postgraduate students. Further the dataset has been divided by 30% for testing and 70% for training on applied machine learning algorithms. Also, the accuracy of the metrics has not fluctuated because we use the random state for each algorithm.

Table 2: Performance metrics of different classifiers.

Classifier	Precision	Recall	F1-score	Accuracy

SVM	0.75	0.75	0.75	81%
DT	0.55	0.69	0.61	68%
NB	0.95	0.75	0.81	75%
K-NN	0.62	0.56	0.58	56%

In Table 2, the result shows the performance metrics for the various classification algorithms. Except for K-NN all the metrics show an accuracy of above 65%. For each classification algorithms, all four metrics are calculated, and SVM gives the highest accuracy of 81% and K-NN got the lowest accuracy of 56%. Also, the calculated stress among participants are as below:

Table 3: Properties of Participants

Properties	Values
No. of Participants	50
No. of instance for Low Stress	22
No. of instance for Moderate Stress	15
No. of instance for High Stress	13

4. Conclusion & Future Scope

In this paper we attempt to find the student's stress and stress level in the academic life cycle of students. The system is trained and tested with 50 postgraduate student's datasets and got an accuracy of 81% approximately. It's the most important to mention this study only examines the latest literature papers available from 2007 to 2019 which is very much related to the stress detection. A total of 30 papers were studied to take the brief view on stress and its effecting factors on academic students. This research work may provide a better way for the development and basic understanding of stress detection on students and may lay the foundation for further studies within the field.

Physiological Signals, like GSR and PPG play an important role in the stress detection of students. If we consider the previous work of stress detection on students, the proposal of this paper will be able to remarkably improve the accuracy of stress detection and find the stress level of students as highly stressed, moderate stressed and low stressed. Furthermore, understanding the academic stress and stress level would help us to develop proper medication

techniques and counseling modules for university counseling experts to help students stress and their well-being.

In the future we can use large datasets with ensemble learning techniques to compare with various individual algorithms to get the best accuracy possible within various age groups of students also we can try to do the comparison of the stress level among male and female students.

References

[1]   Ravinder Ahuja, Alisha Banga, "Mental Stress Detection in University Students using Machine Learning Algorithm", *ELSEVIER, India*, Vol. 152, pp. 349-353, 2019.

[2]   K. Jayasankara Reddy, M.S. Karishma Rajan Menon and Anjana Thattil, "Academic Stress and its Sources among University Students", *Biomedical & Pharmacology Journal,* Vol.11, pp. 531-537, 2018.

[3]   Sanay Muhammad Umar Saeed, Syed Muhammad Anwar, Humaira Khalid, Muhammad Majid, Ulas Bagci, "Electroencephalography based classification of Long-term Stress using Psychological Labeling", *IEEE Journal of Biomedical and Health Informatics,* 2019.

[4]   Ahmed A.Marouf, Adnan F.Ashrafi, Tanveer Ahmed, Tarikuzzaman Emon, "A Machine Learning-based Approach for Mapping Personality Traits and Perceived Stress Scale of Undergraduate Students", *International Journal of Modern Education and Computer Science*, Hong Kong, Vol. 11, Issue. 8, 2019.

[5]   A.Hasanbasic, M.Spahic, D. Bosnjic, H. Hadzic, V. Mesic and O.Jahic, "Recognition of stress levels among students with wearable sensors", *18th International Symposium INFOTEH-JAHORINA (INFOTEH),* East Sarajevo, Bosnia, and Herzegovina, pp. 1-4, 2019.

[6]   R.K.C.Pascual, J.P.D.Serrano, J.M.A.Soltez, J.C.D.Castillo, J.L.Torres, and F.R.G. Cruz, "Artificial Neural Network Based Stress Level Detection System using Physiological Signals", 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management, Baguio City, Philippines, pp. 1-6, 2018.

[7]   Aakash Verma, Astha Dogra, Ketan Malik, and Meenakshi Talwar, "Emotion Recognition System for Patients with Behavioral Disorders", Springer *Link,* Singapore, 2018.

[8]   Mohamed Z.Ramadan and Mohammed H.Alhaag, "Evaluating the User Physical Stresses Associated with Watching 3D and 2D Displays over Extended Time Using Heart Rate Variability, Galvanic Skin Resistance and Performance Measure", *Hindawi, Journal of Sensors,* page.10, 2018

[9]   K.N.Fountoulakis, D.Popovic, M.Mosheva, M.Siamouli, K.Moutou, X.Gonda, "Mood Symptoms in Stabilized Patients with Schizophrenia: A Bipolar Type with Predominant Psychotic Features?" , *Psychiatria Danubina,* Vol.29, No. 2, pp. 148-154, 2017.

[10]   Sriramprakash.S, Prasanna Vadana D, O.V.Ramana Murthy, "Stress Detection in Working People", *ELSEVIER*, India, Vol. 115, Page. 359-366, 2017.

[11]   B.Egilmez, E.Poyraz, Wenting Zhou, G.Memik, P.Dinda, and N.Alshurafa, "UStress: Understanding college student subjective stress using wrist-based passive sensing", 2017 IEEE International Conference on Pervasive Computing and Communications Workshops, Kona, pp. 673-678, 2017.

[12]   Ramanathan Subramanian, J.Wache, M.K.Abadi, R.L.Vieriu, S. Winkler and Nicu Sebe, "ASCERTAIN: Emotion and Personality Recognition Using Commercial Sensors", IEEE Transactions on Affective Computing, vol. 9, no. 2, pp. 147-160, 2018.

[13]   "Monitoring of Arduino-based PPG and GSR Signals through an Android Device", IEEE Engineering in medicine and biology society (EMBS) International student Conference 2016 Design competition.

[14]   Anthonette D.Cantara & Angie M.Ceniza, "Stress Sensor Prototype: Determining the Stress Level in using a Computer through Validated Self-Made Heart Rate (HR) and Galvanic Skin Response (GSR) Sensors and Fuzzy Logic Algorithm", *International Journal of Engineering Research & Technology,* Philippines, Vol. 5, Issue. 3, page.28-37, 2016.

[15]   Nitin Pandey, Abdus Salam, Pooja Ojha, Sara Siddiqui & Zishan Zaidi, "Galvanic skin resistance in different phases of menstrual cycle in acne vulgaris", National Journal of Physiology, Pharmacy, and Pharmacology, 7(5), 455-457, 2017.

[16]   Mohamed Abouelenien, Mihai Burzo, and Rada Mihalcea, "Human Acute Stress Detection via Integration of Physiological Signals and Thermal Imaging", Proceedings of the 9th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA 16). ACM, New York, NY, USA, Article 32, 8 pages, 2016.

[17]   Samira S.Bamuhair, Ali I. Al Farhan, Alaa Althubaiti, Sajida Agha, Saeed ur Rahman and <u>Nadia O. Ibrahim,</u> "Sources of Stress and Coping Strategies among Undergraduate Medical Students Enrolled in a Problem-Based Learning Curriculum", *Journal of Biomedical Education,* Saudi Arabia, 2015.

[18]   A. Fernandes, R.Helawar, R.Lokesh, Tushar Tari and A.V.Shahapurkar, "Determination of stress using Blood Pressure and Galvanic Skin Response", 2014 International Conference on Communication and Network Technologies, pp. 165-168, 2014.

[19]   Nandita Sharma, Tom Gedeon, "Modeling a stress signal", *ELSEVIER,* Australia, Vol. 14, Pages 53-61, 2014.

[20]   L.Vanitha and G.R.Suresh, "Hierarchical SVM to detect mental stress in human beings using Heart Rate Variability", 2014 2nd International Conference on Devices Circuits and Systems (ICDCS), Coimbatore, pp.1-5, 2014.

[21]   Nandita Sharma, Tom Gedeon," Hybrid Genetic Algorithms for Stress Recognition in Reading", *Springer Link,* Australia, Vol. 7833, pp. 117-128, 2013.

[22]   H.Kurniawan, A.V.Maslov, and M.Pechenizkiy, "Stress detection from speech and Galvanic Skin Response signals", 26th IEEE International Symposium on Computer Based Medical System, Porto, pp. 209-214, 2013.

[23]   Bong Siao Zheng, M.Murugappan, S.Yaacob and S.Murugappan, "Human emotional stress analysis through time-domain electromyogram features", IEEE Symposium on Industrial Electronics & Applications, Kuching, pp. 172-177, 2013.

[24]   Nandita Sharma, Tom Gedeon, "Artificial Neural Network Classification Models for Stress in Reading", *Springer Link,* Australia, Vol.7666, pp.388-395, 2012.

[25]   John Bakker, M.Pechenizkiy and N.Sidorova, "What's Your Current Stress Level? Detection of Stress Patterns from GSR Sensor Data", 2011 IEEE 11th International Conference on Data Mining Workshops, Vancouver, BC, pp. 573-580, 2011.

[26]   A. de Santos Sierra, C. Sanchez Avila, J. Guerra Casanova, and G.Bailador del Pozo, "A Stress-Detection System Based on Physiological Signals and Fuzzy Logic", IEEE Transactions on Industrial Electronics, vol. 58, no. 10, pp. 4857-4865, Oct. 2011.doi: 10.1109/TIE.2010.2103538.

[27]   Nandita Sharma, Tom Gedeon, "Stress Classification for Gender Bias in Reading", *Springer Link*, Vol. 7064, pp. 348-355, 2011.

[28]   Mohammed Al-Fudail, Harvey Mellar, "Investigating teacher stress when using technology", *Computers & Education*, 2008.

[29]   Armando Barreto, Jing Zhai, Malek Adjouadi, "Non-intrusive Physiological   Monitoring for Automated Stress Detection in Human-Computer Interaction", *Springer Link*, Vol. 4796, pp. 29-38, 2007.

[30]   Kizito Nkurikiyeyezu, Anna Yokokubo, and Guillaume Lopez Wearable Information Lab

[31]   Kizito Nkurikiyeyezu, Anna Yokokubo, Guillaume Lopez, "The Influence of person-specific Biometrics in Improving Generic Stress Predictive Models", *IEEE TRANSACTIONS*, Aoyama Gakuin University, Japan, 2019.

[32]   Temma Ehrenfeld, "The Three Types of Stress", *Psychology Today,* 2018.

[33]   Professor Roger Azevedo, "Measure Skin Conductivity to Indicate Emotional arousal and Stress", *Imotions,* North Carolina State University.

[34]   M.F.Rizwan, R.Farhad, F. Mashuk, F. Islam and M. H. Imam, "Design of a Biosignal   Based Stress Detection System Using Machine Learning Techniques", *International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka*, Bangladesh, pp. 364-368, 2019

[35]   Lee Y.C, Jang E.H., Chung S.C, Sohn J.H, "ANS responses in negative emotions induced by audiovisual film clips", *Korean Journal of the Science of Emotion & stability,* Vol.10, Issue 3, Pages.471-480, 2007.

[36]   Mi-Hyun Choi, Su-Jeong Lee, Jae-Woong Yang, Ji-Hye Kim , Jin-Seung Choi, Hyung-Sik Kim , Jang-Yeon Park , Jae-Hoon Jun, Gye-Rae Tack, Hyun-Jun Kim, Soon-Cheol Chung, "An analysis of the correlation between young males personal aggression and their skin conductance levels during exposure to aggression images", *ELSEVIER,* Vol. 186, Issues 2-3, Pages 441-442, 2010.

[37]   Marieke van Dooren, J.J.G. (Gert-Jan) de Vries, Joris H.Janssen, "Emotional sweating across the body: Comparing 16 different skin conductance measurement locations", *ELSEVIER*, 2012.

[38]   Roisin Vahey, Rodrigo Becerra, "Galvanic Skin Response in Mood Disorders: A Critical Review", International *Journal of Psychology and Psychological Therapy,* Cowan University Australia, 2015.

[39]   P.H.Venables, D.A. Mitchell, "The effects of age, sex and time of testing on skin conductance activity", *Biological Psychology 43 (1996) 87-101,* 15 November 1995.

[40]   Giulia Cartocci1, Myriam Caratu, Enrica Modica, Anton Giulio Maglione, Dario Rossi, Patrizia Cherubino, Fabio Babiloni, "Electroencephalographic, Heart Rate, and Galvanic Skin Response, Assessment for an Advertising Perception Study: Application to Antismoking Public Service Announcements", *Journal of Visualized Experiments,* vol. 29, No. 2, pp 148-154, 2017

[41]   Christina Gregory, "Mood Disorders", *PSYCOM,* 8 January 2018.

[42]   M.V.Villarejo, B.G.Zapirain & A.M.Zorrilla, "A stress sensor based on Galvanic Skin Response (GSR) controlled by ZigBee", *MDPI*, Sensors, Spain, Vol. 12(5), pp. 6075-6101, 2012.

[43]   Maitree Baral, "81 Students commit suicide in Country's Premier Institutes in Last 3 Years: Government Data", *NDTV*, INDIA 2019.

[44]   Ananya Bhattacharya, "In India, one Student Commits Suicide every hour", *Quartz India,* INDIA, 2018.

[45]   Xiao, Juan, "Academic Stress, Test Anxiety, and Performance in a Chinese High School Sample: The Moderating Effects of Coping Strategies and Perceived Social Support", *Georgia State University*, 2013.

[46] Tom M. Mitchell, "Machine Learning", McGraw-*Hill Science / Engineering/Math,* New York, 1 March 1997.

[47] Peter Flach, "Machine Learning, the art and science of algorithms that Make sense of Data", *Cambridge University Press,* 2012.

Authors Profile

Miss. Zohra Khatoon is Completed a Master of Technology in Computer Science from the Department of computer science and Information Technology, Maulana Azad National Urdu University, Hyderabad, Telangana, INDIA in 2020. She has completed a Diploma in 2014 and a bachelor's degree in Computer Science Engineering in 2017. She has attended Workshops Conducted by MANUU in 2019. Her main Research Work focuses on Machine Learning, Data Mining, and IoT.

Mr. Mohammad Islam is Currently Working as an Assistant Professor in the Department of Computer Science & IT, at Maulana Azad National Urdu University, Hyderabad, Telangana, INDIA. He has 12 years & 6 months of teaching experience. He has completed his Bachelor & Master Degree in Computer Science Engineering. He is also pursuing a Ph.D. from MANUU, Hyderabad. His area of Research in Distributed System, Artificial Intelligence and Bioinformatics. He has active membership in SCI, CSTA, IACSIT, IAENG, ISTE and UACEE.

Network Flow Analysis using Machine Learning

Thumati Manju Priyanka, B.Mouna Shruthi, R Ravinder Reddy
Chaitanya Bharathi Institute of Technology (CBIT), Hyderabad
manjupriyankathumati@gmail.com, mounashruthi159@gmail.com,

ravinderreddy_cse@cbit.ac.in

Abstract: *During the past few years, network analysis and anomaly detection have become a critical process in cyber security.The tremendous growth in the networks, the analysis of the traffic is become a crucial problem. The analysis of the network traffic requires intelligent methods like machine learning approaches. Many organizations are using advanced modern methodologies to detect anomalies and preventing cyber-attacks. The aim of this work is to identify the distinctive behavior in the data congestion. To identify the pattern/characteristic in malicious flow in the network need a dataset for training of our model. The result of this analysis can help and tune the rules to detect the network anomaly.*

Keywords: *machine learning, linear regression, network flow, anomaly detection.*

1. Introduction

There has been an incredible rise in Cyberspace congestion in the recent past. A study finds the growth of 35% is observed in network traffic per annum. The enormous growth is leading to a constant bandwidth encounter between various web applications and instantaneously contributing to an increase in the number of refugee attacks. Thus, efficient web congestion management and analysis is required for the various security applications and itsoperations. It helps in finding the sudden traffic rises and shows any anomaly or malicious performance over the Internet by categorizing information from patterns obtained during the analysis. To identify the difference from malicious traffic to normal congestion, first we have to identify the difference between the flow and strange congestion. The differentiation helps to make the decision among the normal and anomaly of the traffic.

The data streams transferred in between a specified source and destination is called the flow, it is crucial for the identification of abnormal flow in the network in the analysis of anomaly detection. Generally, a flow includes complete data regarding the n number of packet and n

number of bytes belongs to the same stream, along with its time flow. The network flows are always will be helpful for system tracking and controlling.

Most of the researches use the machine learning algorithms for the Cyberspace congestion analysis. In order to identify the flow congestions, Stream analysis is one of the easy procedures which are used to identify patterns followed by the packets related to different kinds of flows. To find the TCP and UDP based congestion in the system need to define the proper data flows along with their source and destination addresses with port numbers. For ICMP based congestion, ICMP message type and code are used. The detected patterns in the traffic flows for size of the packet, count of packets, and time can help in classifying the traffic as normal or anomalous. The newly build models can be used as signature models in the system. As the new trend of data flows in the system are going to change the models more dynamically. The signature-based models are no more valid in the real time scenarios. The learnt parameters can be used for the detecting the anomaly I the real time.

Here, to determine the patterns, the dataset has prepared, which are obtained after flow detection of a datasets including one traffic and three normal traffic datasets. We detect streams for the network in contrast to backbone networks. Many of the related works discussed backbone network traffic for flow detection. We are considering all the datasets from traces for the flow detection other than packets sample flows. Here we are prepared the dataset for the real time and captured the packets with mentioned parameters.

1. Related Work

As the network grows, a large number of web servers are added to the networks, which necessitates the use of packet sampling in order to make the network more scalable. The network flows will be described as the various packets that pass through the controlling point. These flows are critical for identifying traffic jams and calculating other data. The statisticians investigated the precision and validity of sampling techniques for Network clogging [4] and high-speed settings[5, as well as the estimate of congestion flow characteristics from real-world data [6].

They demonstrate that even with a little quantity of data, a statistical picture of the network state may be displayed [4]. In large data sets, packet sampling in flow is used[12, 17]. Systematic sampling is used in Net Flow.

There has been a surge in interest in the use of flow-based approaches for anomaly and intrusion detection in recent years. Examples include [8,9,10], which uses principal component analysis to analyze traffic time series, and[8,] which presents a framework for network anemography.

Another example is [1], which seeks to detect worm propagation on an internal network basis in a high-range network. It identifies the challenge of identifying DoS attacks and frauds in the same setting. The authors focus on approximation header information in this instance since it can be sent via Net Flow (TCP flags). Furthermore, the methodology demonstrated is enthralling, because it demonstrates the difficulty of tracking changes over time.

The importance of timely study of flow data in discovering patterns is highlighted in [3]. The author developed a general-purpose platform for doing time-based side-by-side analysis of flow information in order to attack, with anemphasis on DoS assaults. Time series on flows, packets, and bytes are regarded relevant items [l] from the perspective of web-based monitoring because they provide for a dynamic and real-time overview of all online databased on the stream of data arriving from the exporter [12,15].

We learn about the use of traffic time analysis for detecting abnormalities and intrusions in this paper. We're particularlyinterested in determining if it's required to examine

1) flows ,2) packets, and 3) bytes time series, or whether oneor two of them will enough. In addition, we'd like to see if the result holds true in the presence of sampling. Our idea is unique in that it is based on a genuine case study that was conducted using high-speed data and 10 gbps networks. TheUniversity of Twente (UT) and SURFnet, the Dutch research network, both classified our measurements at the same time [7]. SURFnet uses a packet-to-packet ratio of 1:100.

When flow detection is highly helpful to know, a review of container datasets is highly valuable. The packet-to-packet ratio on SURFnet is 1:100. Most analytical strategies deal with false-positive rate since they are anonymous and less congested. Unrelated information about a smaller number ofdistinct containers taken from a particular stream can beused to detect a risky flow.

3. Methodology

The machine learning is proposing and solving the problems in the many domains, like financial, health care, image and computer vision and many more. The flow of the traffic analysis in the system is identified by learning the parameters in training. Here, we discuss the linear and multiple regression approach used for the prediction.

Here, we use Multiple Linear Regression to generate a model to predict the anomaly in the network. Firstly we will be training our model with some independent variables like port number, total number of forward packets, total backward packets, protocol, destination port, flow duration, length of forward packet, length of backward packets etc. Using this model trained with the above independent variables, we will be predicting the anomaly over a network.

In general the statistical formulation and calculation of Multiple Linear Regression cab be represented in the following equation.

$$y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon$$

1. **Y** is defined as dependent variable of predicted value which is scaled for the anomaly behavior.

2. **β_0** is defined as y intercept.

3. **β_1** it is the regression coefficient

4. n defines the independent variables how many that we are going to test in the network traffic.

5. **$\beta_n X_n$** =it is defined as regression coefficient of the independent variables.

6. **ε** is defined as the model error it defines how much estimation in our expectation y.

In order to get the optimal solution line for all of every independent variables, the Multiple Linear Regression technique calculates the most importantly on the three things:

1. The regression technique coefficient leads to the smallest part of the error.

2. It includes the overall statistics of the overall model.

3. The related p-value which depicts the relation in between dependent and independent variable.

It this technique it calculate the t-statistical and p-values with each and every regressions coefficients in this model. The model is used to analyses the best model based on the learnt parameters.

4. Dataset

As the network contains the data flow in between the source and destinations, to capture the dataset we capture the traffic on the web servers. Which enables us to get the data from the different sources? In the network traffic there are different parameters are there. Here we captured the specific features based on the feature selection techniques. Below tables illustrate some of the significant fields which are used for independent variables in generating the Linear Regression Model. The captured feature aong with the values are shown in the Table 1 & 2.

Table: 1 fields in network flow

Dst Port	Protocol	Timestamp	Flow Duration	Tot Fwd Pk	Tot Bwd Pk	TotLen Fw	TotLen Bw	Fwd Pkt Le	Fwd Pkt Le
0	0	08:31:01	112641719	3	0	0	0	0	0
0	0	08:33:50	112641466	3	0	0	0	0	0
0	0	08:36:39	112638623	3	0	0	0	0	0
22	6	08:40:13	6453966	15	10	1239	2273	744	0
22	6	08:40:23	8804066	14	11	1143	2209	744	0
22	6	08:40:31	6989341	16	12	1239	2273	744	0
0	0	08:39:28	112640480	3	0	0	0	0	0
0	0	08:42:17	112641244	3	0	0	0	0	0
80	6	08:47:14	476513	5	3	211	463	211	0
80	6	08:47:15	475048	5	3	220	472	220	0
80	6	08:47:15	474926	5	3	220	472	220	0
80	6	08:47:16	477471	5	3	209	461	209	0
80	6	08:47:16	512758	5	3	211	463	211	0
80	6	08:47:17	476711	5	3	206	458	206	0
80	6	08:47:17	476616	5	3	211	463	211	0
80	6	08:47:18	477161	5	3	211	463	211	0

Table: 2 fields in network flow

| Len Mean	Fwd Pkt Le	Bwd Pkt Le	Bwd Pkt Le	Bwd Pkt Le	Bwd Pkt Le	Flow Byts/	Flow Pkts/	Flow IAT N	Flow IAT S	Flow IAT N	Flow IAT
0	0	0	0	0	0	0	0.026633	56320860	139.3	56320958	56320761
0	0	0	0	0	0	0	0.026633	56320733	114.5513	56320814	56320652
0	0	0	0	0	0	0	0.026634	56319312	301.9346	56319525	56319098
82.6	196.7412	976	0	227.3	371.6779	544.1615	3.873587	268915.3	247443.8	673900	22
64285714	203.7455	976	0	200.8182	362.2499	380.7332	2.839597	366836.1	511356.6	1928102	21
77.4375	190.8312	976	0	189.4167	347.6426	502.4794	4.0061	258864.5	291724.1	951098	20
0	0	0	0	0	0	0	0.026633	56320240	203.6468	56320384	56320096
0	0	0	0	0	0	0	0.026633	56320622	62.2254	56320666	56320578
42.2	94.36207	463	0	154.3333	267.3132	1414.442	16.78863	68073.29	115865.8	237711	24
44	98.38699	472	0	157.3333	272.5093	1456.695	16.8404	67864	115746.9	237494	15
44	98.38699	472	0	157.3333	272.5093	1457.069	16.84473	67846.57	115645.7	237162	15
41.8	93.46764	461	0	153.6667	266.1585	1403.227	16.75494	68210.14	116178.2	238389	17
42.2	94.36207	463	0	154.3333	267.3132	1314.46	15.6019	73251.14	124959.5	256188	10
41.2	92.126	458	0	152.6667	264.4264	1392.877	16.78166	68101.57	115977.2	238034	16
42.2	94.36207	463	0	154.3333	267.3132	1414.136	16.785	68088	115553.1	237285	8
42.2	94.36207	463	0	154.3333	267.3132	1412.521	16.76583	68165.86	116324.5	238504	12
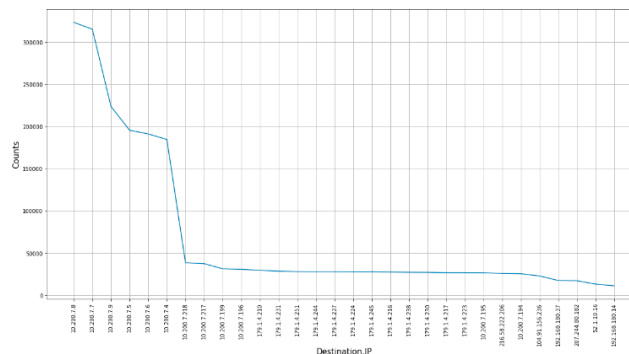
Figure 1: *The detected patterns of the dataset*

We plotted the graph in between the count of the packets and the destination IP as shown in the figure 1. Which depicts the relation of the two parameters.

5. Conclusion

With the increased surge in the network traffic data, patterns in the flow data have become more unpredictable. Hence network flow analysis has become indispensable in monitoring and analyzing the Netflow data. Here in this paper, we propose an effective solution for monitoring and analyzing the Netflow data on organization basis.

References

1. Alqudah, Nour, and Qussai Yaseen. "Machine learning for traffic analysis: a review." *Procedia Computer Science* 170 (2020): 911-916.

2. T. T. T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," in IEEE Communications Surveys & Tutorials, vol. 10, no. 4, pp. 56-76, Fourth Quarter 2008, doi: 10.1109/SURV.2008.080406.

3. Dubendorfer, T., Plattner, B.: Host behaviour based early detection of worm outbreaks in internet backbones. In: WETICE 2005: Proc. of the 14th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise, pp. 166–171. IEEE Computer Society, Washington (2005).

4. Gao, Y., Li, Z., Chen, Y.: A dos resilient flow-level intrusion detection approach for high-speed networks. In: ICDCS 2006: 26th IEEE International Conference on Distributed Computing Systems, pp. 39–39 (2006).

5. Munz, G., Carle, G.: Real-time analysis of flow data for network attack detection. In: IM 2007: 10th IFIP/IEEE International Symposium on Integrated Network Management, 2007, pp. 100–108 (2007).

6. He, G., Hou, J.C.: An in-depth, analytical study of sampling techniques for self-similar internet traffic. In: ICDCS 2005: Proc. of the 25th IEEE International Conference on Distributed Computing Systems, pp. 404–413. IEEE Computer Society, Los Alamitos (2005).

7. Izkue, E., Magaña, E.: Sampling time-dependent parameters in high-speed network monitoring. In: PM2HW2N 2006: Proc. of the ACM international workshop on Performance monitoring, measurement, and evaluation of heterogeneous wireless and wired networks, pp. 13–17. ACM, New York (2006).

8. Yang, L., Michailidis, G.: Sampled based estimation of network traffic flow characteristics. In: INFOCOM 2007. 26th IEEE International Conference on Computer Communications, pp. 1775–1783. IEEE, Los Alamitos (2007).

9. SURFnet (April 2008), http://www.surfnet.nl.

10. Zhang, Y., Ge, Z., Greenberg, A., Roughan, M.: Network anomography. In: Proceedings of the Internet Measurement Conference 2005 on Internet Measurement Conference, pp. 317–330. USENIX Association (2005).

A Review on Neuralink: A Fully Implanted Wireless Invasive Brain Machine Interface

Ch. Sujan, R Ravinder Reddy, P. Charithesh
Chaitanya Bharathi Institute of Technology (CBIT), Hyderabad
sujan2357@gmail.com, rravinderreddy_cse@cbit.ac.in, puppireddycharithesh16@gmail.com

Abstract: *Advancements in Artificial Intelligence, Neurosciences affects every aspect of society. Brain Machine Interfaces (BMIs) are rapidly growing with the increasing pace of advancements in Neurosciences, Neuro- technology, and Artificial Intelligence. BMIs have the potential to restore the sensory and motor functions in various clinical disorders. For an instance, BMI can help a paralyzedindividual in controlling the artificial limb or acursor with their neural activity. With the advancements in microfabrication, machine learning approaches there is rapid recentdevelopments in Invasive BMIs. We can designthe chips and electrodes at submicronresolution and predict the neural activity with enhanced adaptive decoding algorithms. Combination of neuroscience with modern engineering has risen unprecedented development in the BMI field.*

1. Introduction

Individuals with severe spinal cord injuries andparalysis require continuous care to perform daily routines, some have lost the ability to communicate. Discoveries in Neuroscience, advancements in AI and evolution of hardwareopened the bounds for restoring the motor [1] and speech abilities in people with above nervous disorders. Neuroscience proposed the use of Brain Machine Interfaces (BMIs) for disabled persons to interact with the external environment by translating the brain signals into commands to control the machines [2].

2. Related Work

Decades of research paved the path for buildinga Brain Machine Interface. Hans Berger is the first man to detect the Neuroelectrical Activity.To record the electrical activity in the brain he invented electroencephalography (EEG) in 1924. In 1970 USA's Defense Advanced Research Projects Agency started a program toexplore the brain signals and communications inside brain using EEG.

In 1976 UCLA's Laboratory for Brain Computer Interface outlined the systematic attempt to control a cursor with brain decisionsand reactions. UCLA's Professor Jacques J. Vidal coined the term BCI.

Philip R. Kennedy designed a neurotrophic wireless di electrode. In 1998 first invasive approach was used in human brain; these recordings are extremely high quality whencompared with EEG. In 1999 BCI was used to help quadriplegic individuals for movement ofhand.

In 2003 Cyberkinetics company demonstratedthe Brain Gate, where the tetraplegia patient controls and plays a computer game and controlthe prosthetics using brain commands. These milestones campaigned ever increasing pace inthis field.

3. Introducing the terminology

In this chapter we are introducing the related terminology and domain knowledge which describes the BMI field.

3.1 Neuralink

Neuralink is a Neuro-technology company founded by Elon Musk in 2016. Neuralink is focusing on developing a fully integrated wireless Invasive BMI with thousands of channels.

Figure 1: Logo of Neuralink

It is mainly designed for patients with paraplegia and tetraplegia (quadriplegia).Tetraplegia is a paralysis caused by an injury ofthe cervical spinal cord resulting in partial or total motor and sensory deficits in arms, legs, and torso. Paraplegia is the paralysis of lower body which includes both the legs and, in somecases, lower abdomen.

Neuralink, the name itself suggests "creating the link between the Brain and a Machine". Neuralink helps a person diagnosed with paraplegia or tetraplegia to control a computer or a smartphone directly with their brain signals.
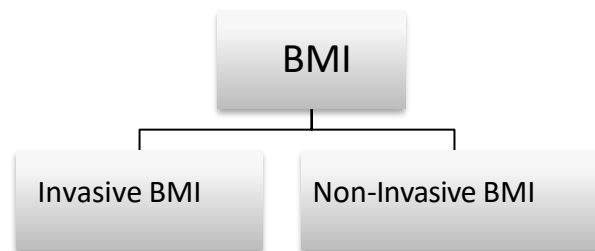
Figure 2: Non-Invasive Approach

3.2 BMI

Brain Machine Interface (BMI) is a system thatenables to communicate and operate a computer, a smartphone, or any other digital devices directly with the brain signals. Brain Machine Interface (BMI) sometimes called as Brain to Direct Neural Interface (DNI), Mind-Machine Interface (MMI), Neural Control Interface (NCI), Brain-Computer Interface (BCI), and Machine Interface (B2M). By decoding the brain activity and brain signals we can operate and control computers, prosthetics,or any other peripheral devices. Based on invasiveness, BMIs are broadly classified into Non-Invasive BMI and Invasive BMI.

Types of BMI

Non-Invasive BMIs are safe, do not carry risksand it can be easily implemented by placing theelectrodes and sensors on the scalp surface [5].Figure 2 shows the approach for Non-invasive BMI. MEG (magnetoencephalography), fNIRS (functional near-infrared spectroscopy), PET(positron emission tomography), fMRI (functional magnetic resonance imaging), and EEG (electroencephalography) are examples of Non-Invasive techniques [2][3][4].

Since the brain signals travel through the skull and scalp before reaching the sensors, they have low quality and attenuated signal recording in Non-Invasive techniques. There are approximately 86 billion neurons in thebrain. The recordings we get are produced by millions of neurons, so they do not give the task-related neural signals precisely.

On other hand, Invasive BMI requires a eurosurgery procedure which are performed by Neurosurgeon experts with the help of advanced Neurosurgical robots. Figure 3 showsthe Invasive Approach.
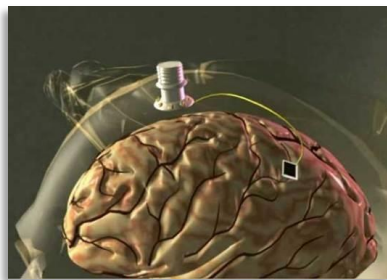

Figure 3: Invasive Approach

Invasive procedure involves insertingelectrodes into the brain tissue by opening the scalp and creating a hole in the skull. Brain is extremely delicate and sensitive, so Invasive procedure carries a lot of risk of brain tissue damage.

As the electrodes are placed very near to the ofneural signal generation source, recordings we get are of high quality and task-related neural activity. The increase in number of microelectrodes increases the recording qualityof signals and maximizes cortical coverage. Invasive BMI became the most interesting research area from the past 10 years because ofhigh quality signal recording from the brain.

3.3 Brain

The brain is one of the most complex organs inhuman body. Brain is responsible for almost allthe functions carried out in the body, and directly controls the ability to feel, walk, talk, think, and remember things. As shown in the figure 4 the three main divisions in the brain are the cerebrum, cerebellum, and brain stem.

The cerebrum, which is the largest part of brain,consists of 2 halves or cerebral hemispheres. The deep inner layer is called white matter andthe upper outer layer is called the grey matter (cortex).
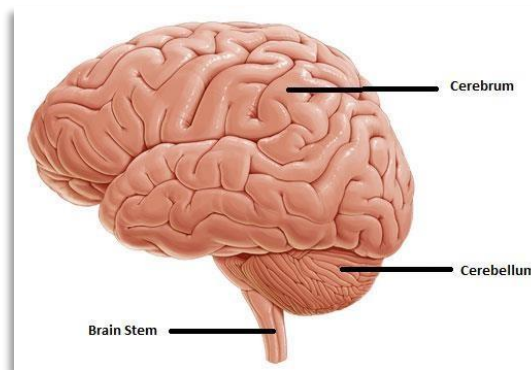
Figure 4: Three major divisions of Brain

The BMIs are still in the experimental phase globally. Neuralink focuses its experiments andstudies on the outermost layer of the brainwhich is called as cortex region.

The cortex region has divisions like Visual cortex – responsible for receiving the information from eyes, Motor cortex – responsible for executing and planning the motor movements, Somatosensory cortex – responsible for feel of touch and Auditory cortex – responsible for perceiving and interpreting the sound.

Figure 5A: Auditory cortex, Visual cortex,Somatosensory cortex on cortical surface

Figure 5B: *Motor cortex on cortical surface*

3.4 Brain Signals

The brain is made of billions of cells called Neurons. Neuron sends and receives the information from other neurons. Neurons can be of different types varying with the locations in brain, but they primarily have three parts: Dendrites, Axon, and Cell Body as shown in figure 6A. Initially a neuron's Dendrite receives the signal from another neuron, Cell body some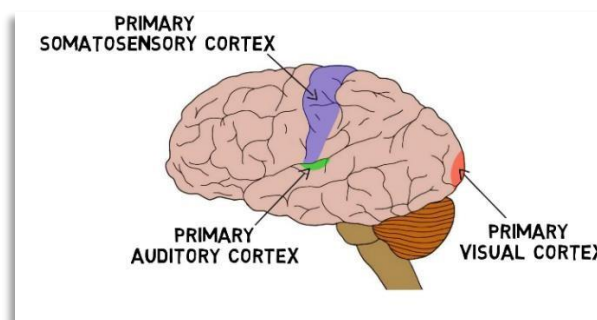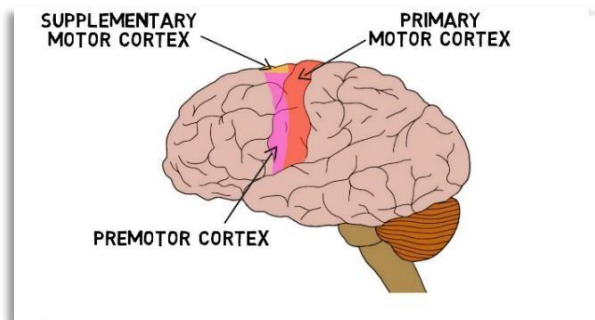times also called as soma computes the signal received, and finally Axon sends out the signal to other neuron. Neurons connect to each other through an axon and dendrite connection. These connections are called as synapses (figure 6B) and are responsible for sending and receiving the signals.

At these connections, neurons communicate with each other by releasing the chemical signals called Neurotransmitters (figure 6C). These neurotransmitters are released at the end of the axon because of an electrical signal called an Action Potential (figure 6D). The action potential is sometimes called as electrical spike. When a neuron receives the 'right' combination of synaptic input, it initiates an action potential. Action potentials travel down the axon produce an electric field. These electric fields can be detected by placing the electrodes near the neurons as shown in figure 6E. Recordings from multiple neurons helps in decoding the information and predict the intended actions that everything we think, touch, see, or feel.

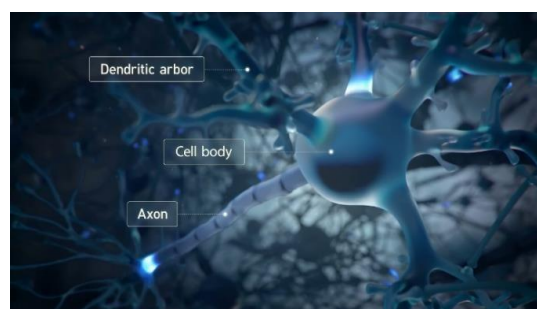***Figure 6A:*** *Primary components of a Neuron-Dendrites, Cell Body and Axon*
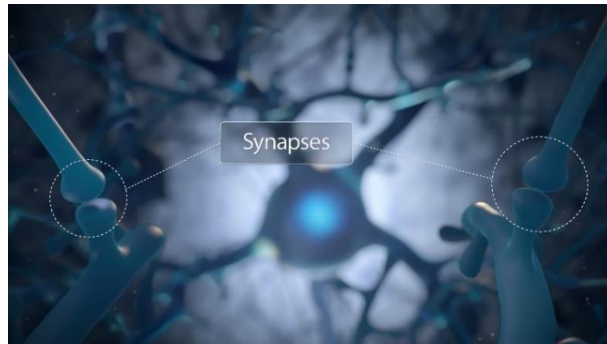
Figure 6B: *Axon-Dendrite connection between two neurons called the Synapse*

Figure 6C: *Chemical transmitters release atsynapse called Neurotransmitters.*

Figure 6D: *Release of Action Potential orSpike at synapse*

Figure 6E: *Electrode sensing the electric fieldfrom nearby neuron.*

3.5 Neuralink Vs Other Invasive BMIs
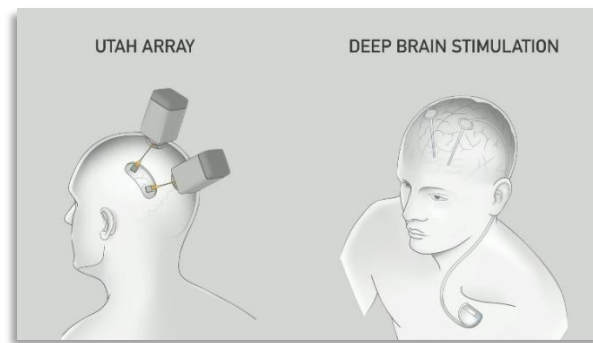
Figure 7: *Utay Array and DBS Approach*

There are only few FDA-approved devices for stimulating the brain and recording the Neural Activity. Utah Arrays and devices used for Deep Brain Stimulation (DBS) have been critical devices in neuroscience and clinical research which are FDA-cleared for neural recording.

Utah Array (figure 7) has just 100 channels or electrodes, whereas the stiff and long DBS probes have only four to eight electrodes. DBS is mainly designed to modulate the brain activity, but not to read the information from the brain nor to write into the brain [2].

DBS and Utah Arrays are completely wired. These implants are not completely contained within the body [3]. Big boxes of extracranial recording hardware and wires connected to the brain can carry the risk of infection [5]. The rigid spikes and electrodes inserted in the brain can disturb and damage the brain during the movement.
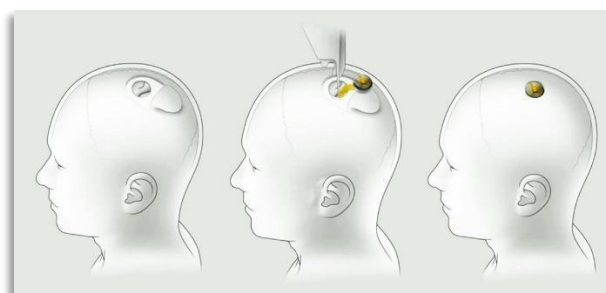
Figure 8: *Neuralink Approach*

Even the other devices that are being used in clinical trials do not have more than a few hundred electrodes. At Neuralink they created a device called 'LINK'. Unlike, the rigid spikes

31

used in DBS, the electrodes in Neuralink are very flexible and are placed individually by a neurosurgical robot to avoid blood vessels, vasculature [6].

The electrodes are placed only at interested locations of brain at pre-register insertion sites. The THREADS used in the Link are 800 timestinner than DBS probes (figure 9).

Figure 9: Comparison of electrode size ofDBS with Neuralink
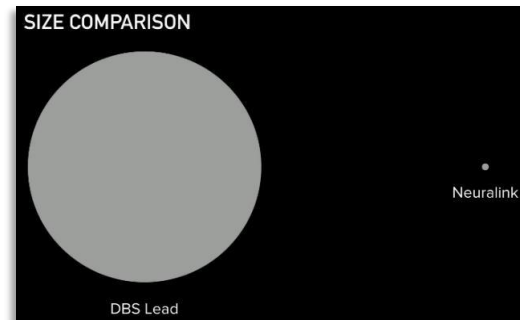
Neuralink built various configurations of devices with 3072, 1536, and 1024 electrodes (channels) [7]. The current clinically approveddevices have around 100s of electrodes perarray, but Neuralink designed an implant that has thousands of tiny and flexible electrodes, the channel count is almost 100 times more than the existing devices.The Link can perform real-time spike detectionon the data receiving from each channel, and this digitized processed data will all be sent wirelessly through Bluetooth. As Neuralink is wireless, there is no risk of infection unlike other invasive techniques.

4. Working of Neuralink

Neuralink built tiny array of electrodes which are inserted into brain tissue detects the electric fields generated by neurons. On-chip amplifiers amplifies these tiny micro signals, Analog-to- digital converters transform the amplifiedsignals into digital format. On-chip spike detection algorithms detect the spikes and sends the processed data to external mobile through wireless technology.

Figure 10: Components of Neuralink whichprocess, stimulate and transmit neural signals

4.1 Threads and Electrodes

Threads are ultra-thin and flexible polymers that contain the electrodes. Each array consists of 48 to 96 threads [7]. They are not simple strands of wires, instead, each carries 8 to 128 independent electrodes (figure 10). Polyimide is the primary dielectric and substrate used in the probes. Threads transfer the signals and information to the transmitter. Threads should be biocompatible with the surrounding tissue and should resist the corrosion from the chemical and salt fluids in the brain tissue.
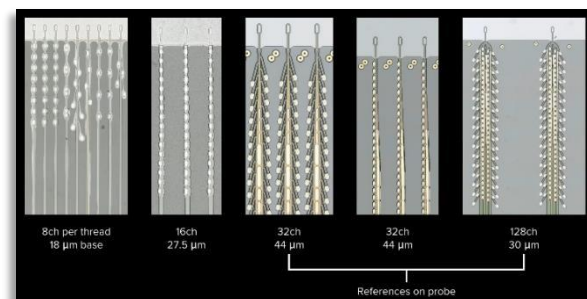
Figure 11: Different designs of threads likeTree and Linear Edge

When we insert a stiff probe, the brain tissue triggers an immune response that can limit the functionality and durability of the device [8]. Stiff probes can cause notable tissue lesions and get encapsulated by protein and glia deposits making the probes unusable forrecordings after few months or weeks [9]. To overcome this problem, Neuralink designed these threads to be on the same scale as neighbouring neurons in size, so threads are asthin and as flexible as possible.
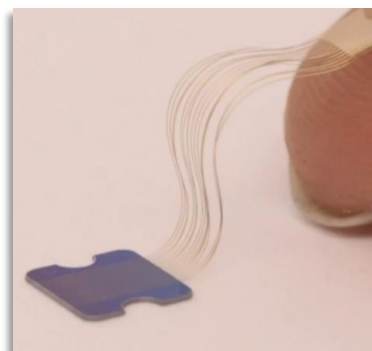
Figure 12: Neural Threads

Average human hair is human hair varies in diameter, ranging anywhere from 17 μm to 181 μm. Neuralink Thread thickness nominally ranges from 5 μm, it includes 2 layers of conductors and 3 layers of insulation like adding the corrosion resistant layers to the threads.

The length of neural thread is approximately about 20 mm. At sub-micron resolution, Techniques like Stepper lithography are used toform a metal film which makes it easy to handle before inserting into brain. Electrodes have a small geometric surface area which results in higher impedance and low charge carrying capacity, to overcome this limitation Neuralinkis using a polymer called Poly(3,4- EthyleneDiOxyThiophene) which is doped with PolyStyrene Sulfonate (PEDOT: PSS) andIridium Oxide (IrOx) [7]. PEDOT:PSS is highly electrically conductive and is highly resistant to oxidation.

4.2 Neurosurgery and Robot

Threads are very fine and thin; it cannot be manipulated by neurosurgeons directly using their hands. Threads are even very flexible andcannot go directly into the brain tissue on their own. Insertions should be safe and avoid vasculature with high precision and accuracy. Neurosurgeons need the help to insert these threads into brain tissue, so they created a Neurosurgical Robot [10] which can insert onethread (32 electrodes) in ten seconds.
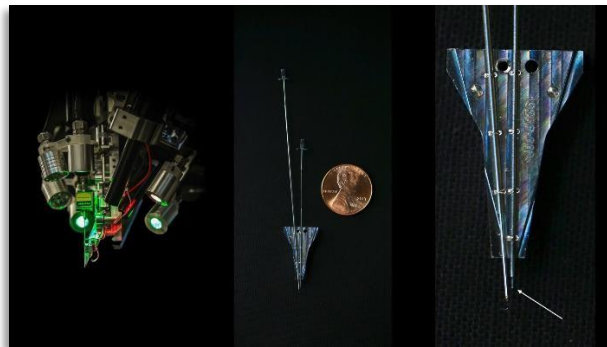
Figure 13: *The robotic electrode inserter ; Therobotic electrode inserter ;Needle pincher cartridge (NPC) compared with a penny for scale ;Enlarged view of the inserter-head.*

Optical stack, computer vision, and variousintegrated custom build software suits allow the pre-register of insertion sites and planning the path for insertions (figure 13). With all this planning we can avoid vasculature and blood vessels during the insertions and reduce tangling and strain on threads.

Figure 14: *Pre-selection of insertion sites*
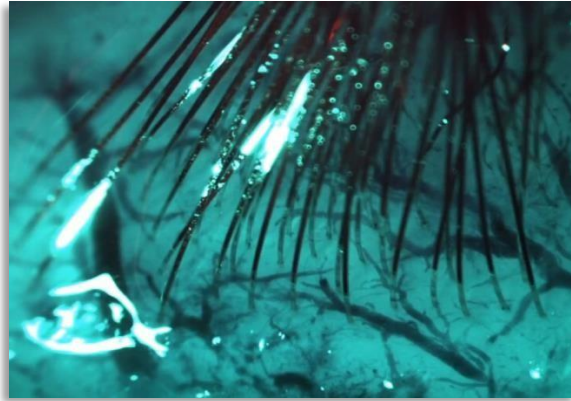
Figure 15: *cortical surface with implantedthreads and minimal bleeding*

The system consists of six independent light modules. Each light module can illuminate thelight with 405 nm, 525 nm, and 650 nm (whitelight) wavelengths independently [7]. The lightwith wavelength 405 nm can excite fluorescence from the polyimide substrate which helps the imaging system to guide the robot and distinguish the threads from greymatter (cortex region). Illumination of lightwith other wavelengths like 525 nm and 650 nm, Stereoscopic cameras, and custom build software-based monocular extended depth of field calculations allows the precise estimationof the regions over the cortical surface.

To insert the threads into brain tissue, first the neurosurgical robot should register the insertion sites. But breathing and heartbeat cancause the movements in brain, so robot uses the landmarks of the skull to make a common coordinate frame. To maximize the area of cortical coverage, the insertions are made withspacing about more than 300 µm between the threads. The length of each thread is nominallyaround 20 mm. To handle these thin and long threads before the insertion, parylene-c is deposited onto the threads to form a thin film. The threads

remain attached to the film until this neurosurgical robot pulls them off individually from the film for inserting into brain tissue. Over 19 surgeries the average success rate of insertion is of 87.1 ± 12.6 % (mean ± s.d.).

4.3 Chips and Electronics

The electronics used are custom-built Neuralink Application Specific Integrated Circuit (ASIC). The link converts the analog electrical signals into real-time digitized Neural Information.
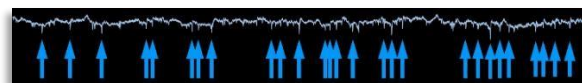
Figure 16: Analog Representation of Spikes
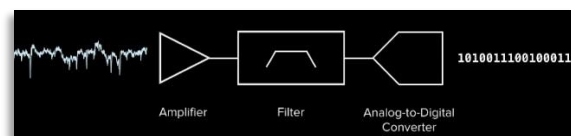
Figure 17: Conversion of Analog signals toDigital signals in LINK

The neural signals are very small (around microvolts) and are recorded in analog form. So, the LINK should have a high-quality recording stack that should amplify small neural signals that are in order of < 10 µVrms magnitude, digitize the amplified signals, and send the processed results wirelessly using Bluetooth Low Radio for real-time processing.The total computation should need minimalpower and on small-size chips which are on then an oscale.
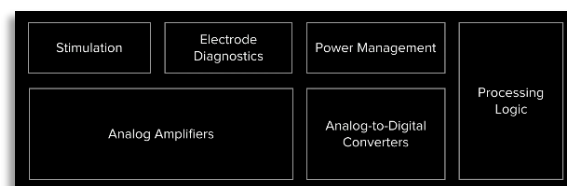
Figure 18: Basic Architecture of LINK

Design of Link may vary with version, but at the end, the components it needs to meet are Analog Amplifiers, Analog-to-Digital Converters, Processing Logic, Power management, Electrode Diagnostics, on-chipstimulation, and clock drivers. The link shouldhave very high-performance and high-quality signal amplifiers and digitizers. There are 256 amplifiers called Analog Pixels which are individually programmable, on-chip Analog-to-Digital Converters

(ADCs), and a peripheralcontrol circuitry for serializing the digitizedsampled outputs. The on-chip Analog-to- Digital Converters sample at 19.3 kHz with 10bit resolution. Each amplifier (analog pixel) can consume 5.2 µW and the whole ASICcircuit consumes around ~6 mW, whichincludes the clock drivers.

Using flip-chip integration, several ApplicationSpecific Integrated Circuits are integrated into a standard PCB (Printed Circuit Board). Each Link system consists of a Field Programmable Gate Array (FPGA); real-time accelerometer, temperature, and magnetometer sensor; and Bluetooth Low Radio for full-bandwidth data transfer.

Each chip has 16 wires (i.e., 256 electrodes) and there are four chips, that is where we get 1024 channels per neural link device [7]. Also, with the increase in number of electrodes, channel count increases, a lot of raw data that need to be sent wirelessly also increases. To send these huge raw and unprocessed digital signals to external devices wirelessly, the LINK consumes lot of power.

For efficient power consumption and usage, thedevices require on-chip spike detection which assures the characterization of neural spikesand the real-time identification of neural spikes. The custom-built chips on the Link meet these goals, which reduces per-channel chip size and consumption of power.

Link is a wireless device that is implanted and fully contained within the skull. The wireless charging of link is very much like the chargingof smart phones using Inductive Technology. A Compact inductive charger connects wirelesslyto the LINK to charge the battery from the outside overnight which gives the Link a full day battery life.

4.4 Neural Decoding and Electrophysiology

The real-time digitized broadband signals needto be processed and fed to algorithms. To identify the task-related neural activity and action potentials, Neuralink use custom built online detection algorithms. Conventional neurophysiology deals with training the algorithms with the spike data offline. But BMIevents need real-time spike detection to estimate the neural population dynamics [11].

From recordings of each 1024 electrodes, the Link amplifies and digitizes the voltage. By theactivity of neurons, the nearby electrodes can detect theses micro voltage traces called spikes(or action potentials).

The spikes are detected using custom designed algorithms. Later these spikes are aggregated into vectors of spike counts. One count forevery 25 ms over 1024 channels. For every 25 ms, the Link transmits these spike counts usingBluetooth low radio to a computer on which a custom decoding software was being running.

Artificial Neural Networks (ANN), Support Vector Machines (SVM), Linear filter with adaptive weights, Linear decoder [5], Kalman filter, Wiener filter, Hierarchical linear discriminant analysis classifier, Threshold based classifier, Filter bank canonical correlation and Continuous linear filter [12] arevarious strategies for decoding and predicting the neural signals.
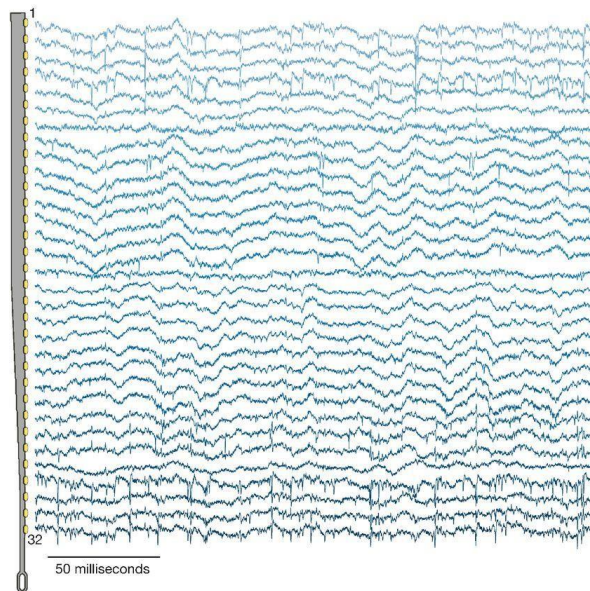
Figure 19: Unfiltered Neural signals receivedfrom single thread (32 electrodes)
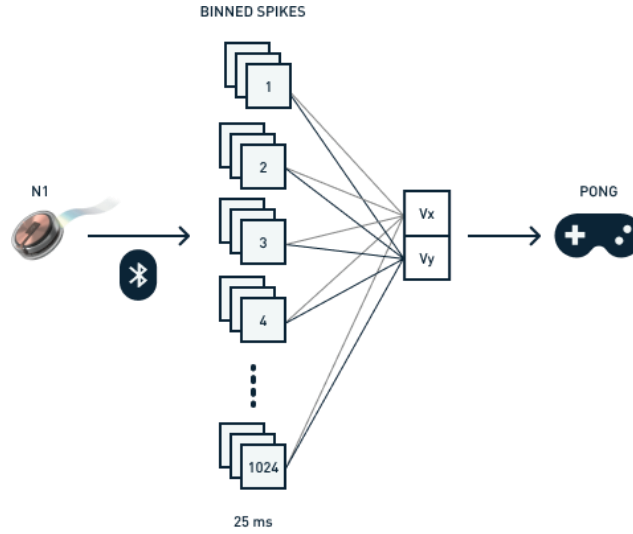
Motor neurons have different temporal neural properties. Therefore, the link needs spike counts from recent most 25 ms to the past 250 ms. So, to account this, the custom software re-arranges the spike counts at several timescales and then the weighted sum of these past, current, and recent spike counts is computed foreach dimension of control bypassing their firing rates through a decoding model [11]. Theoutput of the decoder is a set of velocity signalsfor each 25 ms bin, which can be used to directthe cursor movement on the computer screen.

The choice of 25ms bins is a very smart, as it isvery fast enough for all neural applications likeprosthetics and external devices. Many demonstrations and studies in this field uses 25ms or 50ms bin widths. The choice of Neuralink is using 25ms bin width while drawing history of 250ms on the computer for each velocity decode.

5. Limitations

Neuralink has not yet begun clinical trials in humans. They are still under trials on rodents and primates focusing on multiple cortical structures of the brain. Their demonstrations and research focus mainly on the cortical surface, not on the deep brain.

Neuralink demonstrated an implanted macaque monkey where the Links are placed bilaterally one in the left motor cortex and other on right cortex. The left cortex controls movements of the right side of the body and left cortex controls the left side of the body. Each implantconsists of 1024 channels making a total of 2048 channels. It shows the need of multiple implants to cover the whole region.

The chips have stimulation, which means electrodes not only read the information from the brain, but they can also write into the brain,but not demonstrated any work on that. As it

includes highly sophisticated technology, the initial cost of surgery and implant could be very high.

Privacy is just a myth in these days. We have been witnessing many data breaches in recent years. The data can be confidential andsensitive. In nearer future BMIs can develop inan unimagined way, if there is a breach withoutthe permission of user then it not only disclosesthe sensitive information but also every thoughtof the individual. The data breach from BMIs could be worse than leaked logs of emails and passwords. Security should be key aspect in BMI which may leads to reinvention ofCybersecurity into Neuralsecurity.

6. Future Advancements

Focusing on Deep Brain and applying for FDAapproval for clinical trials in Humans. Closed- loop therapies [13]; and demonstrations, research on stimulation can drive the future advancements in Neuralink.

Figure 21: Electrodes sensing the electricfields from nearby neurons

Electrodes sensing the electric fields from nearby neurons is simply reading the information from the brain (figure 19). Thedesign of chips shows that the Link is capable of stimulation. Which means not only reading the information from brain, but also one can write information into the brain. Each electrodecan deliver small currents. These currents change the electric field which makes the nearby Neuron to fire action potential. This action potential cause other neurons to fire thereby creating a chain of neural signals. So, by stimulating in right combination of spatiotemporal sequences at differentelectrodes can possibly create a pattern of activity. This activity can be of anything like feel, touch, or sensation. Stimulation canreduce the movement deficits in a person diagnosed with Parkinson's disease.

To treat spinal injured patients, Neuralink not only trying to make the patient interface with external digital devices, but also planning to make a secondary shunt below the spinal cord,so

that with triggers in brain signals reaches theshunt and shunt can transmit these signals down the body to make the patient capable of moving his legs [12]. They can potentially restore the physically mobility and motor movement. The neurosurgery at Neuralink is held with help of neurosurgical robot which takes the minute inputs from the neurosurgeons. Neuralink trying make this robot as fully automatic [7], where the complete procedure can be done by surgical robot alone.

7. Conclusion

Advanced methods of telecommunications are required to build a BMI. The delay in transmitting the signals should be very minimaland should meet the demands of transmitting enormous amounts of data. BMIs need to be invery small size and should be bio-compatible with high tissue permeability.

Technology at Neuralink is based on decades ofresearch in this field. The present clinically approved BMI systems do not have more than 100s of electrodes per array and uses the lengthy external connectors which come out from the skull. All the previous trails show therisk of infection when the devices are not fullycontained within the body and to connect with BMI the patient, need a caregiver or atechnician. Neuralink is building the BMI with1000s of channels which is 10 times higher channel count with the present BMI systems. The design of Neuralink is completely wirelesswhich ensures the zero chance of Infection.

References

[1] Fabien B "Targeted neurotechnology restores walking in humans with spinal cord injury" Nature 563 (2018)

[2] Edwin M. Maynard, Craig T. Nordhausen,and Richard A. Normann. "The Utah Intracortical Electrode Array: A recording structure for potential brain-computer interfaces", Electroencephalography and Clinical Neurophysiology 102.3 (1997)

[3] Taro Kaiju "High Spatiotemporal Resolution ECoG Recording of SomatosensoryEvoked Potentials with Flexible Micro- Electrode Arrays", Frontiers in Neural Circuits11 (2017)

[4] György Buzsáki, Costas A. Anastassiou, and Christof Koch "The origin of extracellularfields & currents-EEG, ECoG, LFP & spikes",Nature Reviews Neuroscience 13 (2012)

[5] Lebedev And Nicolelis "Brain-machine Interfaces: From Basic Science To Neuroprostheses And Neurorehabilitation"American Physiological Society (2019)

[6] Flavia Vitale "Fluidic Microactuation of Flexible Electrodes for Neural Recording", Nano Letters 18.1 (2018)

[7] Elon Musk & Neuralink "An integrated brain-machine interface platform with thousands of channels'' White paper (2019)

[8] Tarun Saxena "The impact of chronic blood–brain barrier breach on intracortical electrode function", Biomaterials 34 (2013)

[9] Aziliz Lecomte, Emeline Descamps, and Christian Bergaud. "A review on mechanical considerations for chronically-implanted neural probes" Journal of Neural Engineering 15 (2018)

[10] Timothy L Hanson "The "sewing machine" for minimally invasive neural recording" BioRxiv (2019)

[11] Dai J, Zhang H, Zhang S, Li Q, Liu X, Hao Y "Temporal-spatial model decoding for neuron ensemble spikes from rats' cortex" Science in China Series C, Life Sciences (2009)

[12] Marco Vilela & Leigh R. Hochberg "Applications of brain-computer interfaces to the control of robotic and prosthetic arms" Elsevier (2020)

[13] Weidong Chen, Shaomin Zhang, Yaoyao Hao, Kedi Xu, Nenggan Zheng, Xiaoxiang Zheng "Toward bidirectional closed-loop brainemachine interfaces BMIs a summary on invasive BMI research in China'' Elsevier(2020)

[14] S.K. Mudgal, "Brain computer interface advancement in neurosciences: Applications and issues" Elsevier (2020)

Ground Water Level Analysis & Prediction

Kabeerudin, R Ravinder Reddy, M Venkata Krishna Reddy

Chaitanya Bharathi Institute of Technology (CBIT), Hyderabad

kabeeruddin2297@gmail.com, rravinderreddy_cse@cbit.ac.in, krishnareddy_cse@cbit.ac.in

Abstract: *Management of ground water in India always suffered with serious problems like extensive irrigation in major canal commands and overexploitation of groundwater for all purposes. Exhaustion of water levels, salt water infringement, drying of aquifers, contamination of groundwater, water soaked and saltiness are the important results of over extraction, over usage of water and serious water systems. In India, the highest category of groundwater is consumed for the irrigation field. This field uses almost 88 per cent of groundwater which in turn drastically decreases levels of groundwater. The goal of the proposed approach in such cases, is to find and predict availability of the groundwater depending on parameters and inputs like history of availability of groundwater and data of water levels at the surface, weather forecasting's, usage of water, extraction of groundwater, other geographical data and target outputs which includes groundwater level scenario. The policies of the government are also mostly impacts on the ground water levels.*

Keywords: *Central Ground Water Board, Receiver Operating Characteristic, MLE-Maximum Likelihood Estimation.*

1. Introduction

Groundwater is the one available under Earth's surface in the layers of soil and in the cracks of rock layers. An aquifer is defined as the part of rock that yields a usable amount of water. A water table is called the depth at which the water availability in the soil or rock layers under the earth get saturated completely. Groundwater gets recharged from the rainfall on the surface. It may dry out in the summer due to humid heat. Groundwater is extracted for agricultural, household, and industrial usages by means of digging and operating wells. The study involved in the extraction, movement and distribution of groundwater is hydrogeology[1].

Groundwater is freely available, more natural and less affected by vulnerabilities than surface water. That's why it is generally used for providing drinking water. As an example, among all the states of the United States, California extracts large amounts of groundwater annually as a part of usable water storage. In the US, Underground reservoirs are used for storing more water than all surface reservoirs and tanks, including the Great reservoirs. In general, many municipalities in India rely on groundwater availability for day to day water supply. Ground water is always Earth's freshwater. Groundwater is defined as the water that exists underneath the layers of the earth within the spaces between molecules, particles and in holes, layers of soil, cracks and breaks in rocks. Usually groundwater always available not less than 80-100 meters of the Soil surface. Groundwater consists enormous constituents counting

microorganisms, life gasses, organic, inorganic and several naturally available materials. With growing population the usage is increasing, the levels of the ground water are falling. It is major concern in these days; need to improve the ground water levels. This work helps to improve the scenarios which can enhance the ground water levels. The machine learning is adopted to build a model which can predict the ground water levels season wise more accurately.

2. Related Work

Groundwater Management in India is facing a lot of problems due to huge extraction and extensive irrigation. Saturation of water tables, increase in saltwater, dry out of aquifers, contamination of groundwater, non availability of water etc. are major consequences. In different parts of the nation, the water levels are decreasing approximately 2 m/year. But in some areas, the water levels are increasing hugely as almost 1 m/year. Degrading the standards of groundwater due to several reasons is next immediate problem. Groundwater users of West Bengal become panic due to rise in acid content in shallow water levels. Because of all these problems, the freshwater availability for all purposes like irrigation, industrial, municipal and domestic uses is reducing. Solution is to be provided for all these groundwater problems otherwise water crisis will be severe in the coming future in India. Looking into this situation, the Indian Government has started various protective as well as useful measures in order to reduce the groundwater levels control management issues. But the above measures do not create any impact because of political and administrative reasons and lack of awareness. Among all the countries across the world, India is the largest user of groundwater. Groundwater satisfies more than 50% of the total requirement of fresh water in the country. It withdraws more groundwater than the US and China - the next two biggest countries to pull groundwater – combined [2]. The central water resources standing committee identified that groundwater forms the major share of agriculture and drinking water supply in India in the year 2015. Irrigation occupies the major portion of the groundwater extraction which is about 89 per cent, marking it as the largest sector user in the country. Household usage sector becomes second in use of groundwater with 9 per cent. Next is the industry sector which uses only two per cent of it. Totally, groundwater satisfies 80 percent of rural domestic water supply and 55 percent of urban water demand. This situation has drastically reduced groundwater levels in India by 62 per cent between 2008 and 2018, based on the report from the Central Ground Water Board (CGWB)[3]. The report stated that increase in population, fast urbanization; insufficient

rainfall and industrialization are the major reasons for immediate decline in groundwater levels in the country.

3. Our Contribution

Conventionally, levels at which groundwater available, is predicted using process-dependent models, such that which mostly rely on the in-depth knowledge gathered from system dynamics. These models demand a lot of extra spatial data related to the geological along with hydrological perspectives related to the aquifer. When compared, the opposite side is data-dependent models for groundwater level prediction with machine-learning methods proved effective. We are proposing a model which is based exclusively on the domain-related information and further joined into the framework through data dependent transformation by including new attributes. The aim of the proposed approach is to predict and sense the availability of groundwater levels depending on transient inputs of data that includes groundwater history, data of surface water level, weather forecasting data, extent levels of using land, rainfall data, groundwater extraction, some other relevant data along with outputs related to groundwater level.

In the proposed approach, a model is trained to predict groundwater levels continuously provided a set of relevant attributes. Data is collected by combining all the related attributes and target - groundwater level and taken up under supervised learning [4]. The following discussion focuses on the techniques used to obtain output:

A. Linear Regression Method

Linear Regression is an approach that the relationship linearly considering the scalar response as input and more than two explanatory variables in the field of statistics. The input variables are also called as dependent and the output variables are termed as independent variables. It is used for predictive analysis. It plots straight line representing the relationship between the variables.

B. Logistic Regression Method

Logistic regression always models a binary dependent variable statistically using basic form of logistic function even though that may lead to many complex extensions. Parameters of a model are estimated by logistic regression in regression analysis. S-shaped curve is generated which can have values between between 0 and 1.
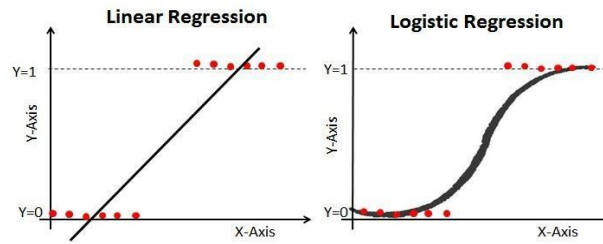
Figure1: Representation of regressions.

C. Confusion Matrix:

Confusion Matrix acts as a predictor of performance of a model on any classification problem. The number of adjusted and inaccurate forecasts is summarized with tally values and broken down by each class. This is projected always at the base of the confusion matrix. The confusion matrix shows the path of how our classification model is confused when it gives predictions on observations[5]. It helps us to measure the errors of our model while classifying the observations into different classes.

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Figure 2: Prediction

D. Random Forest:

Bootstrap Aggregation and various decision trees are used for performing both relapse and classification in this gathering method. Final output is decided by combining multiple decision trees instead of depending on individual decision trees.

4. Implementation

Implementation includes pre-processing of data, scaling the data, dividing the data into training & testing sets and building a model which can be trained by training data. It consists of an analysis of data and building a model for prediction from the given data

Data set has been downloaded from open govt data website. It consists of various information which results in the change in the concentration of groundwater level. It consists of various fields like Recharge from rainfall Monsoon season, Recharge from other sources, Domestic and industrial uses, etc.

The data set also consists of 4 categorical variables which are classified based on the information from various columns

```
In [3]: data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 29 entries, 0 to 28
Data columns (total 13 columns):
States                                                       29 non-null object
Recharge from rainfallMonsoon season                         29 non-null float64
Recharge from other sources                                  29 non-null float64
Recharge from rainfallNon-monsoon season                     29 non-null float64
Total_Rainfall                                               29 non-null float64
Natural discharge during non-monsoon season                  29 non-null float64
Net annual groundwater availability                          29 non-null float64
Irrigation                                                   29 non-null float64
Domestic and industrial uses                                 29 non-null float64
Total_Usage                                                  29 non-null float64
Projected demand for domestic and industrial uses upto 2025  29 non-null float64
Groundwater availability for future irrigation use           29 non-null float64
Situation                                                    29 non-null object
dtypes: float64(11), object(2)
memory usage: 3.0+ KB
```

After we provide the data set we will the following output screens of different types of
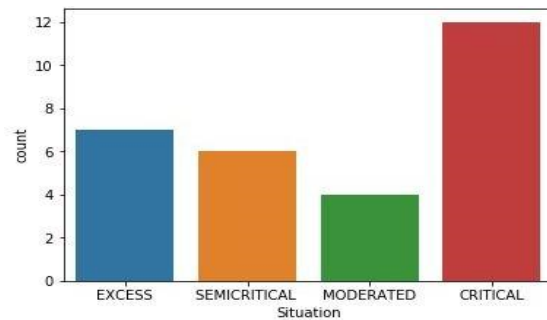
evaluation.
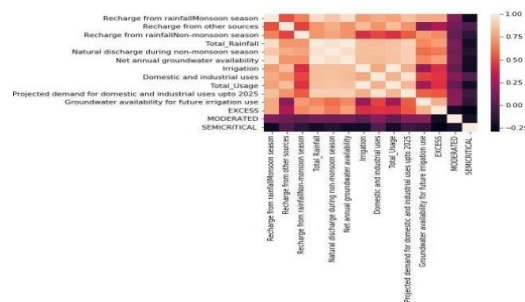
Figure 3: Count of Categorical variable
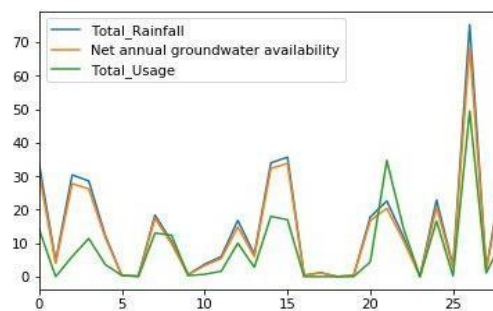
Figure 4: Correlation between the input variables

47
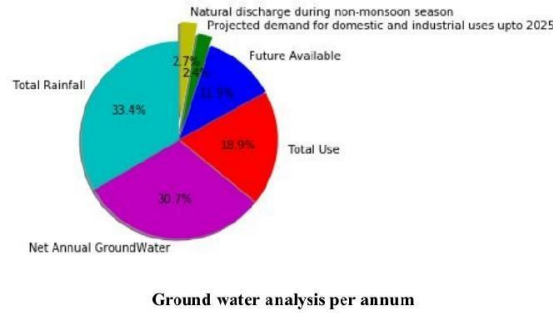
Figure 5: *Total Rainfall Vs Ground water availabilityVs Total Usage*
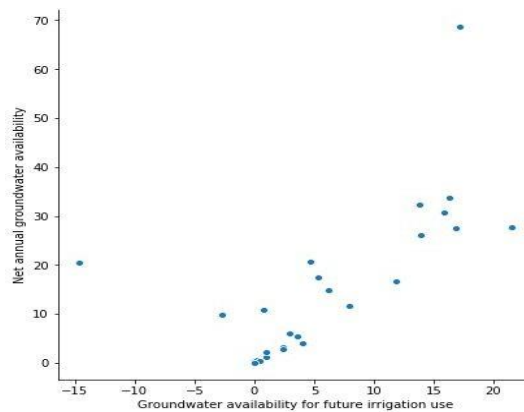
Figure 6: *Ground water analysis per annum & Future Availability Vs Net Amount.*

The above mentioned are the output displayed after we upload the dataset.

5. Conclusion

In this approach, we predicted the levels of groundwater. Four models that support data-driven methodologies are verified like linear, logistic regression, confusion matrix & random forests. Forecasting and Prediction of change in groundwater levels using the proposed approach proved to be an efficient model when compared in order to predict exact groundwater levels using traditional methods. Models used in the proposed approach and also tested are multifaceted with different estimator columns chosen from the data set provided. The model which we have built has an accuracy of 0.9375 and ROC curve score of nearly 0.88 which means the model that we have built is good and able to predict the situation with a probability of greater than 75%. Since the data set we have used is small and limited there are no performance tuning issues. But if the data set is large then things tend to change where we need to take a close watch at the performance tuning metrics and the outliers in the data

References

[1] S. Kumar, "Neural Network Water Level In Metropolitan C Runoff As A Par."Kumar," Neural Network Model For Prediction Of Ground Water Level In Metropolitan Considering Rainfall-Runoff As A Parameter. "International Journal Of Soft Computing And Engineering (Ijsce).

[2] Sujatha And Pradeep Kumar, "Prediction Of Groundwater Levels Using Different Artificial Neural Network Architectures And Algorithms."

[3] M.Kavitha And B.Pratap Naidu, "Application Of Soft Computing Techniques For Groundwater Level Forecasting." International Conference Proceeding Numbers Of 2012 On Computer Networks And Communication Systems. [Vol. 35. 2012].

[4] S. K. Gupta And R. D. Deshpande (2004). "Water For India In 2050: First-Order Assessment Of The Available Options". Current Science. [Vol.86, No.9,10 May 2014].

[5] Kaul. S. N., Mahajan.A. U., & Nandy, T. (1999). "Water And Wastewater Treatment, Recycle And Reuse". Journal Of The Indian Association For Environmental Management, 26, 74–90.

Social Media and Its Impact on Covid19 Outbreak in India

Jeelani Ahmed[1] and Muqeem Ahmed[2]

Department of Computer Science and Information Technology

Maulana Azad National Urdu University, Hyderabad

*jeelani.jk@gmail.com[1], muqeem.ahmed@gmail.com[2]

Abstract: *Today, almost everyone knows and uses the internet by means of various social media platforms. This social media is consisting of unverified messages, posts, stories, articles, and news. Due to the massive use of online news media and social media, there has been a sudden outburst of fake news on the internet. Earlier, spreading fake news was one of the difficult tasks but today, due to social media it has become very easy to spread fake news and make it viral. Fake news causes serious effects on the society we live in, it is now considering as one of the major threats to democracy of a country, journalism, and freedom of expression. Today, the world is facing one of its biggest pandemics since the Spanish flu of 1918. The Novel coronavirus outbreak has declared as a pandemic by the World Health Organization. Since the beginning of the covid19 outbreak in China, many misinformation/disinformation about it is being spread across the globe. The massive increase of misinformation around the coronavirus pandemic has never been known before. This paper critically examines the fake news since the beginning of the covid19 outbreak in India and presents some fruitful insights.*

Keywords: *Fake news, misinformation, coronavirus, covid19, pandemic, lockdown*

1. Introduction

Since the beginning of the internet, fake news exists in the world. Fake news is nothing but the propaganda or fabricated news consisting of false information to deliberately spread for the purpose of deceiving its readers [1].

The motive behind spreading fake news is to damage someone's reputation, mislead the readers, or to gain popularity. Fake news is spreading through both traditional and non-traditional media channels. Traditional media channels such as newspaper and television and non-traditional media channels such as Facebook, Twitter, etc. [2].

The social media platforms like Twitter and Facebook [3] are highly used media channels through which a massive amount of fake news is continuously shared [4]. Users of these social media platforms can share their opinions and views in an unstructured and unedited style. Today, fake news on these platforms is getting more views and impact on society as compared to traditional media outlets. According to a study of fake news velocity [4], tweets with misinformation and tweets having truth information have a ratio of 6:1 in terms of access to the readers.

Many times it has been seen that a massive rise in spreading fake news happened before any political election or any political or religious event but this time fake news is being largely shared during a pandemic [5]. The world is witnessing the worst situation of coronavirus pandemic and at the same time, our smartphones and internet are full of fake news that is being shared around the COVID19 [6].

The rest of the article is organized as follows: Section II comprises the methodology used and dataset information. Section III presents data analysis using which some crucial findings are drawn. Section IV has some techniques to spot fake news. Finally, section V concludes with a conclusion and future work.

2. Method

We have collected 395 samples of misinformation instances from a library maintained by Tattle Civic Technology. The collected misinformation headlines represent stories from the IFCN certified fact-checkers such as The Quint Webqoof, Indiatoday fact check, BOOMlive, AltNews, and Factly during the period of 28 Jan 2020 to 30 June 2020, that is the main period of news coverage of novel coronavirus. To understand this misinformation properly, we analytically annotated each news story is for the medium of propagation, media source, representative tags, and thematic-categories. Our analytics is limited to the number of fact-checking that has performed by the mentioned IFCN certified fact-checkers.

3. Findings

By systematically analyzing 395 misinformation samples during the period of novel coronavirus pandemic we have witnessed some useful insights which are discussed below
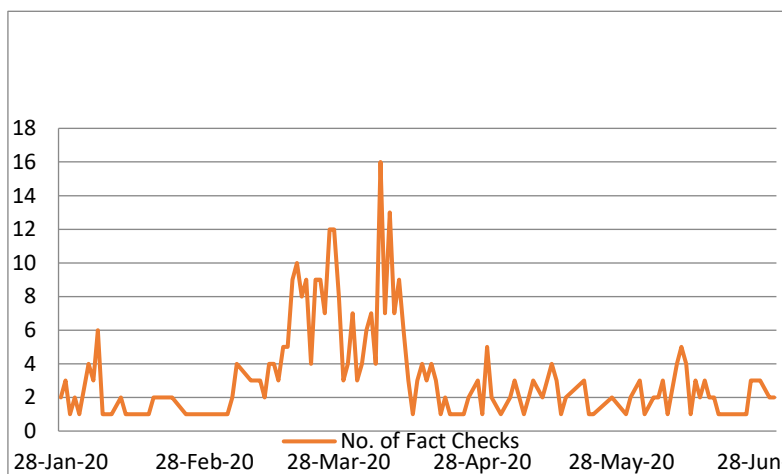
Figure 1. Number of fact checks during COVID19

The rise in fake news about the novel coronavirus has suddenly increased in March after the announcement of Janta Curfew by Prime Minister Narendra Modi to maintain social distancing. Later it will rise when the lockdown has implemented to break the chain of transmission and abstain from coronavirus. The number of fact-checks and the weekly cumulative growth has shown in the below figures.
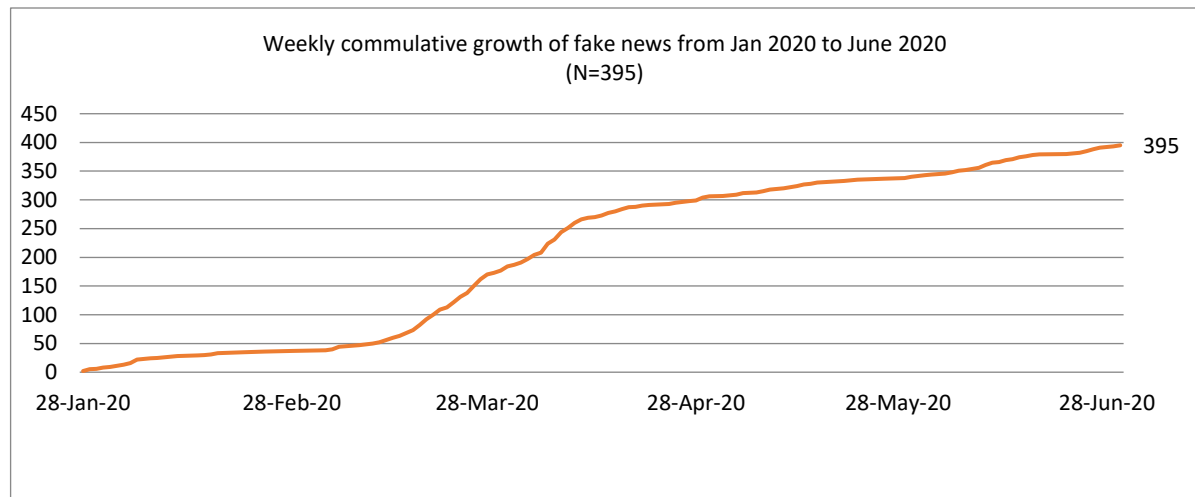
Figure 2. Weekly cumulative growth of misinformation from 28 Jan 2020 to 30 June 202

As we can see there has been a massive growth in the misinformation between March and April following a religious event held in New Delhi. This made people blame a particular community for intentionally spreading novel coronavirus.

We have classified misinformation into nine categories to understand the pattern of spreading misinformation easily. Table 1 below discusses the categories and their respective definitions and number of misinformation cases.

As we can see in above table 1 there are two categories i.e., Communal and Government around which misinformation has been rising consistently. The Communal misinformation is due to the rising fake news coverage around a particular community for allegedly spreading coronavirus. The second category of misinformation is due to the police brutality in the lockdown and fake news about the government advisories to contain the novel coronavirus. The fake remedies, unproven preventive measures, and false news related to the covid19 vaccine lead to an increase in Cure, Prevention, and Treatment Category. The pie chart in figure 3 depicts this summary of fake news topic trends.
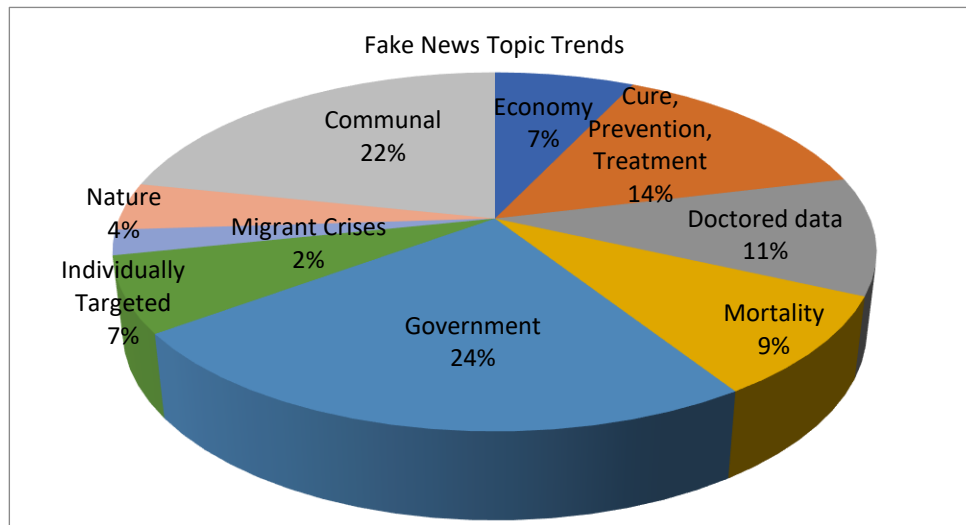
Figure 3. *Pie chart depicts fake news topic trends during novel coronavirus*

Table 1. *Categories of misinformation amidst novel coronavirus*

Sl.No	Category		Cases	Description
1	Communal		86	Misinformation related to religious events/entity and news that thought to be spreading religious hatred.
2	Cure, Prevention, Treatment		56	Misinformation which has suggestion for unproven remedies, cure, vaccine and false preventive measures.
3	Doctored data		42	Data which has manufactured falsely and intentionally to spread fake information.
4	Economy		28	Misinformation references to fake business data, panic buying and other things which impact economy.
5	Mortality		36	Posts related to illness, deaths of the people due to the coronavirus.
6	Government	95		Posts related to government's notifications, announcements and orders for public, political parties and police.
7	Individually Targeted	27		Misinformation which has incorrectly manufactured to target a person/entity.
8	Migrant Crises	09		Misinformation with references to the Migrant issues.
9	Nature	17		Misinformation with references to the environment.

4. Content type

We have found that misinformation is being spread in almost every type of content whether it is Text, Image, Audio, or Video. The content types and its percentage share of 395 fake information during novel coronavirus are shown in the below pie chart.

It has seen that all types of content are widely misused for spreading fake news, for instance, video content is used to show the disturbing graphics of misinformation whereas tweet largely uses text content to mislead its readers.

5. Month wise Propagation of Misinformation

We also find that the mode of propagation of misinformation over each month since 28 Jan 2020 to 30 June 2020 during coronavirus pandemic as shown below
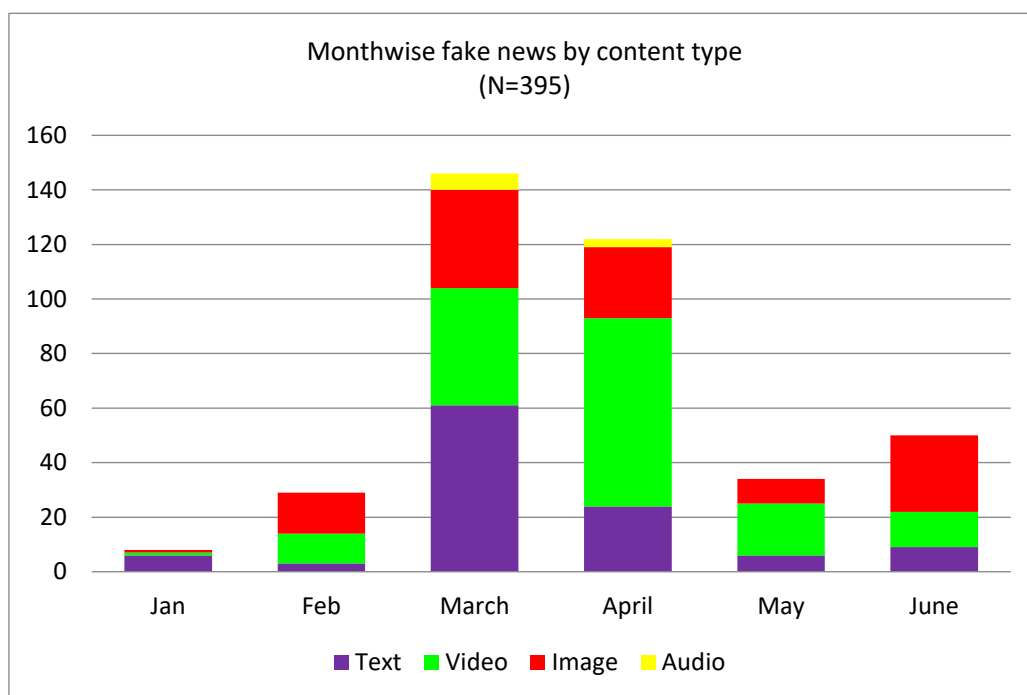
Figure 5. Month wise fake news with mode of propagation

It has found that in March, April almost every type of content has used to spread fake news following a religious event in New Delhi that was organized by a particular community. Video and Image contents have mostly used to spread misinformation on almost every social media platform.

We also annotate misinformation categories with the type of propagation. Each fake news category shared almost every content have shown in below figure 6.
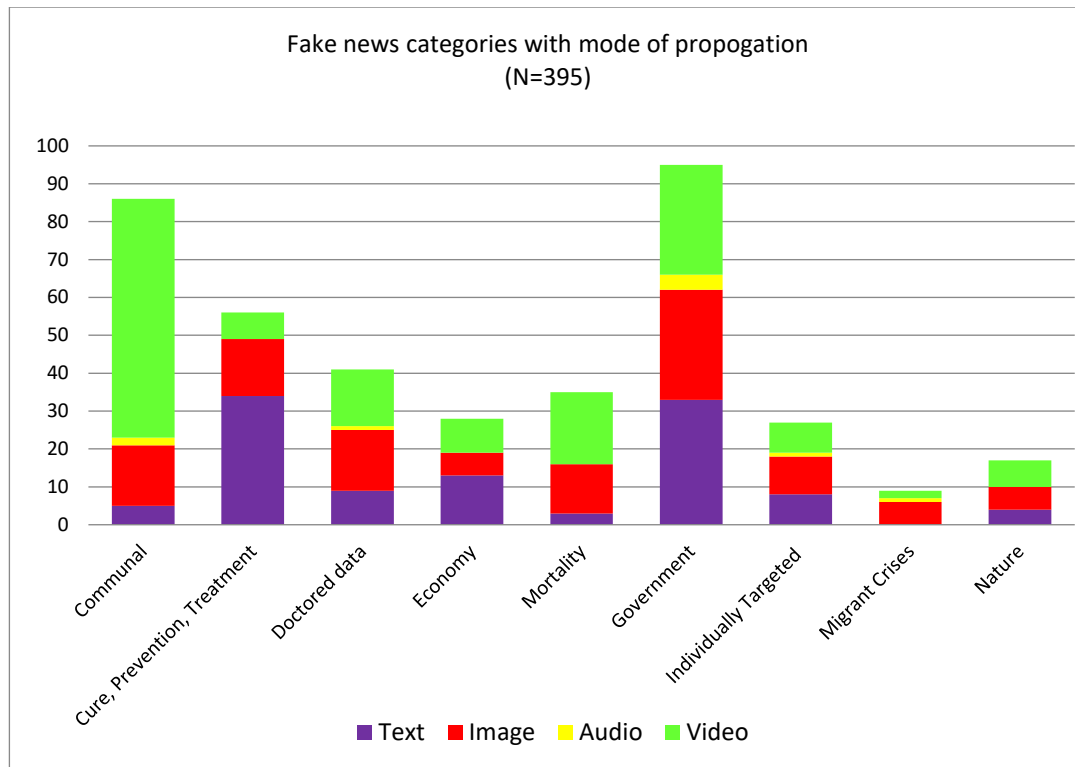
Figure 6. *Fake News Categories with mode of propagation*

The Government and communal categories of misinformation mostly hold the video content type and few categories hold the audio content type in very less amount to propagate fake news. Images and Texts have largely used to propagate misinformation that lies in the Cure, Economy, and Doctored Data categories.

6. Misinformation by Mainstream Media

.
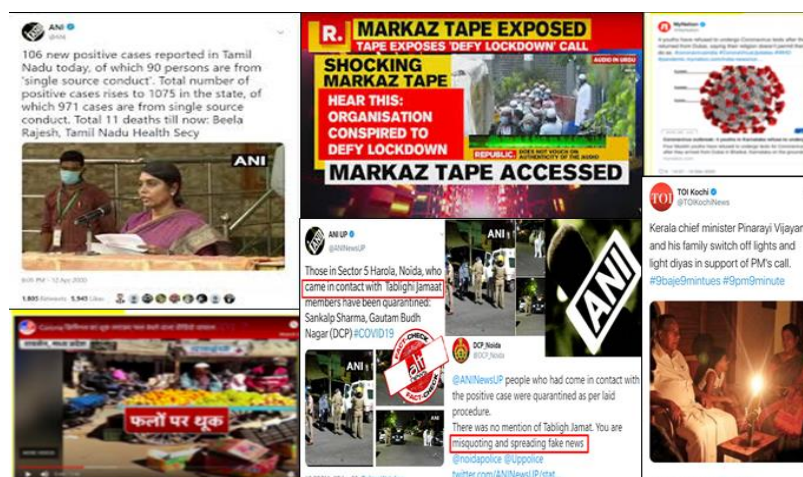
Figure 7. *Screenshot of Misinformation spread by various Mainstream Media*

We have found that the source of fake news is not limited to social media itself but reaches the mainstream media of India too. The Indian mainstream media ranging from regional media, digital media to the high TRP national news media which includes News18, TV9, ABP News, TOI have involved in broadcasting misinformation. Figure 7 is an example of such misinformation reported by mainstream media

Some media channels fall multiple times into broadcasting fake news which has illustrated in below figure 8. The mainstream media ranges from Print Media, TV Media, and Digital Media and as well as regional media of different languages.
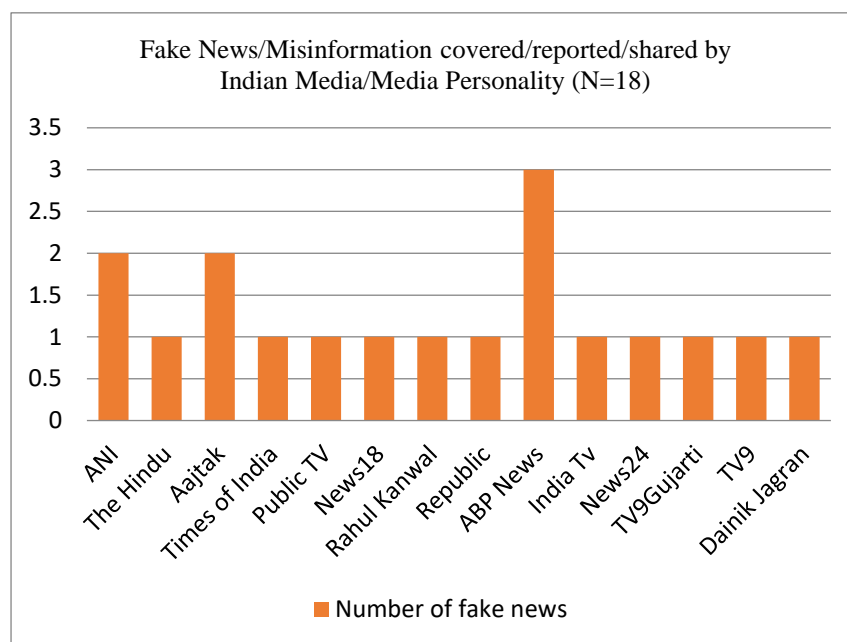
Figure 8. Number of fake news spread by various media channels/media personality.

7. Locations of/from Misinformation in India

Sharing of misinformation is not limited to some cities only; instead, it is spread all over India. From our data of misinformation, we have found that most of the fake news is about/from metro cities. We have collected some samples of fake news and compiled it into Indian state names from where they belong to or related to is shown in figure 9.
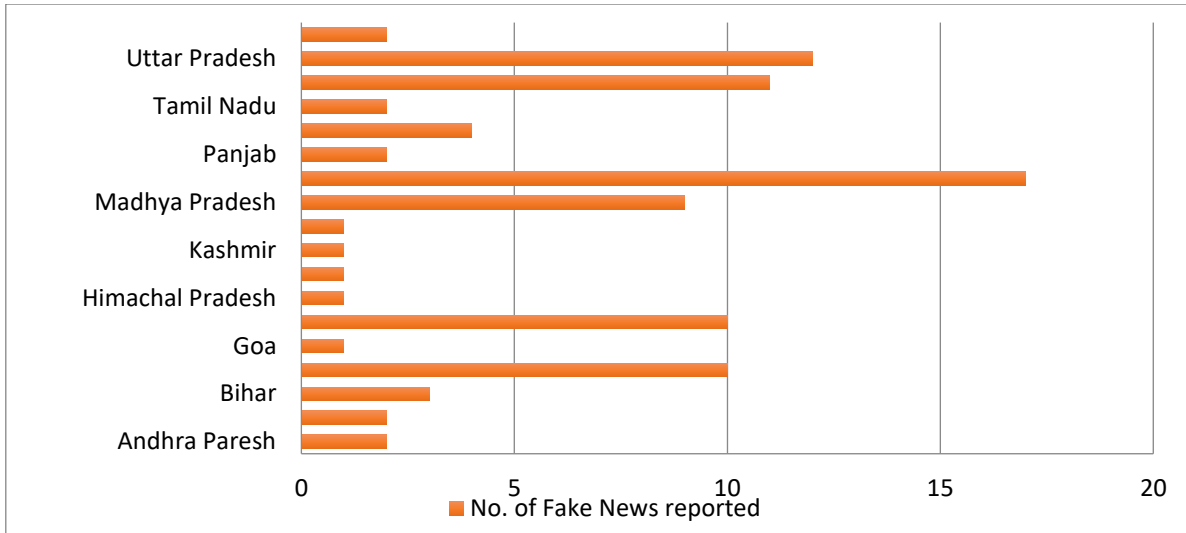
Figure 9. Number of fake news reported from/about each Indian state

We find out that fake news is spread all over the country mostly from or about the Maharashtra, Uttar Pradesh, Telangana, Delhi, and Gujarat and least from the Uttarakhand, Manipur, Goa, Kashmir, and Kerala.

8. Misinformation Categories
8.1 Communal Purpose

Misinformation related to religious events/entities and targets a particular community has kept in this category. As the following word cloud shows in figure 10, communal related misinformation has most of the tags related to Muslim/Islam objects. These tags show how a religious community was held responsible for allegedly spreading novel coronavirus.

8.2 Cure, Prevention and Treatment

Misinformation that has suggestion for unproven remedies, cure, vaccine, and preventive measures are falls into this category. As shown in the word cloud figure 10, we see most of the tags associated with the treatment of COVID19 like home-remedies, Ayush, cure, prevent, etc.

The number of misinformation concerning the Cure, Prevention, and Treatment of coronavirus was at a peak at the beginning of the pandemic but it eventually came down after clarification from many health organizations, doctors, medical experts, and government, etc.

57

8.3 Doctored Data

Data that has manufactured falsely and intentionally to spread fake information comes into this category. As shown in figure 10, all the tags associated with this category are manufactured, morphed, doctored, fake, graphic, etc.

Doctored Data is not new into the fake news; the main purpose behind this type of data is to intentionally spread the morphed or manufactured data to mislead its readers. In the midst of the COVID19 pandemic, we find many such doctored videos, audios, and images which mislead social media users intentionally.

8.4 Economy

Misinformation that references fake business data, panic buying, and other things that impact business and economy have put under the Economy Category. As shown in figure 10, various tags associated with the economy are essential-items, lockdown, money, markets, rupees, milk, etc.

As bogus messages are spreads with false intention and to create panic buying in the markets. Fake images and videos were shared with misleading claims during the nationwide lockdown in India amidst coronavirus pandemic.
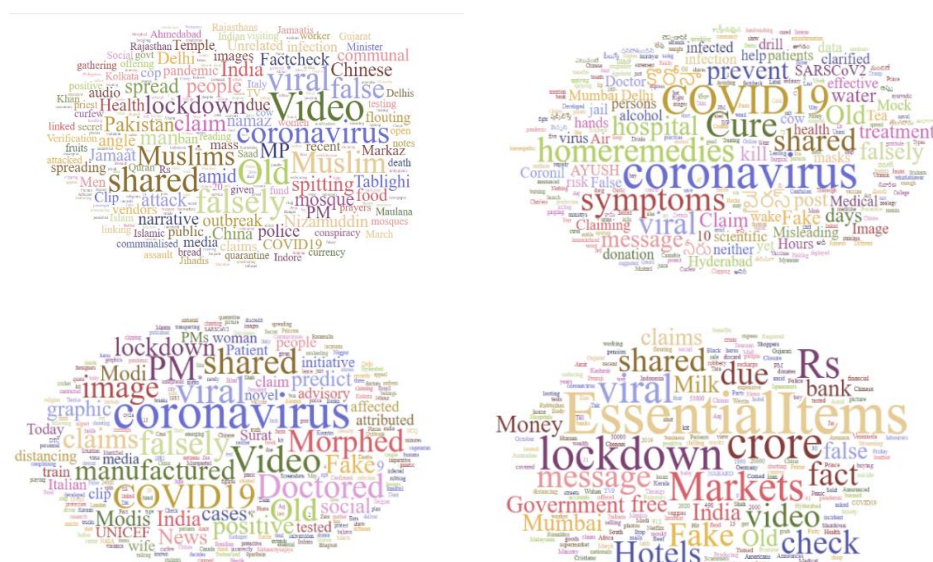
Figure 10: *Word cloud of all the tags associated with communal, cure and prevention, Doctored Data and Economy misinformation respectively.*

8.5 Mortality

Mortality misinformation category deals with the false posts related to the illness, deaths due to the coronavirus. These messages contain misinformation related to suicides, deaths, and illnesses of the people during coronavirus pandemic.

When mortality kind of misinformation spreads that contain graphics of deaths, images of people suffering from symptoms, it leads to creating panic and distress among newsreaders and social media users. These images and videos of dead bodies did not have any relations with the actual statistics of coronavirus deaths.

8.6 Government

Posts related to government's notifications, announcements, and orders for the public, political parties, and police has classified into this category. The tags which are related to this category are lockdown, police, ministry, army, curfew, etc. are shown in the below figure 11.

We have seen that misinformation spreads regarding the government advisories, announcement, and various notifications about the lockdown that was announced by the government amidst coronavirus pandemic.

8.7 Individually Targeted

Misinformation that has incorrectly manufactured to target a person/entity has classified into this category. We have seen that during the pandemic, many famous personalities and political leaders have falsely attributed to fake news and get targeted individually. All the tags refer to this category are Sachin Tendulkar, Amitabh Bachchan, Kanika Kapoor, Rahul Gandhi, etc. are shown in the word cloud in the below figure 11.

8.8 Migrant Workers

Misinformation with references to the migrant issues has classified into this category. The word cloud for this category is shown in the below figure that has all the tags such as migrant, lockdown, hungry, beaten, train, etc.

On 24 March 2020, the nationwide lockdown was announced by the government due to the COVID19 outbreak since then everyone got stuck at their places especially the migrant workers. These workers started facing many issues due to the suddenly announced lockdown and afterword's much fake news spreads on social media as well as form the mainstream media. One such example is ABP Majha News telecasted the fake news of train boarding migrant workers from Mumbai Bandra railway station, whereas there was no such

announcement by the railway ministry. We have found that there are several images on social media in which police show beating the migrant workers for not following the lockdown guidelines.

8.9 Nature

Misinformation with references to Nature and Environment has classified into this category. Various tags associated with this category are mutton, bats, meat, animals, market, chicken, human, etc. shown in figure 11. Since the beginning of the coronavirus outbreak, misinformation about its origin was at a peak. As nobody had the complete and authentic scientific information and many unproven facts were spread as the cure and treatment for the COVID19. For example, misinformation about chicken-eating will cause coronavirus infection, whereas there was no such proven fact about eating chicken was true.
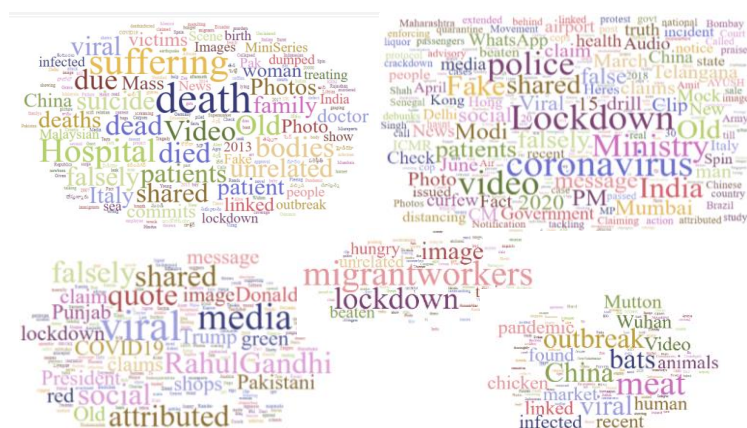
Figure 11: *Word cloud of all the tags associated with Mortality, Government, Individually Targeted, Migrant Workers and Nature misinformation respectively.*

9. Conclusion

It is obvious that with the rapid global spread of coronavirus infection, the world is also witnessing the rapid growth of misinformation around coronavirus to which the World Health Organization has labeled as a 'massive infodemic'. The fake news which contains myths, bogus remedies, conspiracy theories, and fake guidelines about COVID19 can cost people's lives [5]. We have carried out a detailed and critical study of such fake news that spreads between Jan 2020 to June 2020 and discovers some useful insights that would use to tackle fake news. We have mentioned some methods to spot the misinformation. Proposing a scientific tool to detect such fake news on the internet would be the future of this work.

References

[1] Carson, J., & Cogley, M. (2021, January 7). Fake news: What exactly is it – and how can you spot it? The Telegraph. https://www.telegraph.co.uk/technology/0/fake-news-exactly-donald-trump-rise/

[2] Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and Resolution of Rumours in Social Media. ACM Computing Surveys, 51(2), 1–36. https://doi.org/10.1145/3161603

[3] Jayaraj, J. (2020). Covid-19: When The Misinformation Is Spreading Faster Than The Virus. Journal of Thoracic Disease and Cardiothoracic Surgery, 1(1), 01. https://doi.org/10.31579/2693-2156/005

[4] Langin, K. (2018). Fake news spreads faster than true news on Twitter—thanks to people, not bots. Science. Published. https://doi.org/10.1126/science.aat5350

[5] Depoux, A., Martin, S., Karafillakis, E., Preet, R., Wilder-Smith, A., & Larson, H. (2020). The pandemic of social media panic travels faster than the COVID-19 outbreak. Journal of Travel Medicine, 27(3). https://doi.org/10.1093/jtm/taaa031

[6] Vicario, M. D., Quattrociocchi, W., Scala, A., & Zollo, F. (2019). Polarization and Fake News. ACM Transactions on the Web, 13(2), 1–22. https://doi.org/10.1145/3316809

Vaccination Studies through Integration of Zigbee Communication Protocols in Fixed Wing Air Vehicle for Coronavirus Disease (Covid-19) Pandemic Regions

Syed Mohammed Shoaib[1], Shaik Abdul Khader Jilani [2]

[1]Engineering Physics Lab, MANUU Polytechnic Kadapa Campus, School of Technology, Maulana Azad National Urdu University, Gachibowli, Hyderabad-500032, Telangana State., India
[2]Department of Electronics & Communication Engineering, Madanapalli Institute of Technology and Sciences Madanapalli-517325, Chittoor District, Andhra Pradesh., India
Email: smdshoaib@gmail.com

Abstract: *Present dangerous pandemic of Corona virus disease (COVID-19) caused by severe acute respiratory syndrome Corona virus 2 (SARS-COV-2). This outbreak was first identified in WUHAN China on 1st December 2019. As of latest reports more than 15.8 Cr COVID-19 cases have been reported in 185 countries and territories as a result more than 32.9 Lakhs deaths happened and this figure is increasing rapidly day by day. The World Health Organization declared the outbreak to be a public health emergency of international concern on 30th Jan 2020 and recognized as it is a pandemic on 11th March 2020. This Virus spread between people during close contact, often via small droplets produced by sneezing, coughing or talking and via touching. Fever, Cough and Shortness of breath are common symptoms of this pandemic COVID-19 disease. Vaccination supply using unmanned Micro Air Vehicles (MAVs) on long-term effect of COVID-19 pandemic disease is a present challenging research problem before us. The vaccination and data collection in and surroundings the COVID-19 pandemic regions using MAVs helps to better understand the present COVID SECOND WAVE propagation in India. Airborne bacteria, fungus like black fungi, white fungi, pollen and viruses like COVID-19 causes different type of allergies and lung diseases. However aerobiological research may help us to better understanding the relationship between aerial organisms, human health, ecosystems, including important food resources. Autopilot-equipped MAVs can exactly sample along pre-defined coded flight plans and sharply regulated altitudes with the help of GPS (Global Position System). They can provide even considerable benefit when they are networked together in coordinated COVID-19 sampling missions: such research calculations can produce further information about the aerial transport process rapidly and efficiently. Therefore, the task of coordinating multiple independent vehicles to address scientific and other missions has like to increase the attention from researchers.*

Keywords *Vaccination, Aerobiology Instrumentation, Unmanned Micro Air Vehicle (MAVs), Global Position System (GPS) and COVID-19.*

1. Introduction

In this present investigation, flight vehicle path planning, real-time sample analysis, control and coordination policy plan are examined for unmanned autonomous aerial vehicles with the help of Zigbee protocols. Based on geometric concepts, A time optimal path planning algorithms, which was easy enough to be resolved in real time was obtained. The present research work focuses on the validation-controlled trials of aero dispersion models of COVID-19 vaccination to identify and estimate the spreading of viruses and lung diseases. The use of Zigbee protocols allows us to control MAV and monitor various sensors in real-time with distributed control. Low power, long distant and authentic data communications features of Zigbee are fully utilized in the present pandemic COVID-19 research work. The present novel

method would be a significant improvement upon the existing system due to the presence of an Zigbee based control and monitoring system. The objective of the present investigation is to implement advanced communication Zigbee protocol, which provide most reliable and secure data transfer for guided unmanned aerial platforms along with GPS technology which can locate the MAV with considerable accuracy and collect the vaccination data from all the three pre-determined geographical pandemic COVID-19 regions (**Red Orange** and **Green** zones) at accurate altitudes. In this present research work path planning and control plan of action will be focused that are to be used in field experiments and vaccination supplying in and around the pandemic regions.

The air i.e., commonly known as the atmosphere of the Earth basically made of the combination of two major gases Nitrogen (78%) and Oxygen (21%). Plant and animal pathogens, insects, seeds, viruses and pollen or other living and non-living organisms that utilize the atmospheric air to alter habitat may incorporate aero biodata. Not only different gases, air also contains minute particles, like dust which contain microscopically living and non-living microorganisms. These living and non-living aerial microorganisms have a significant effect on human beings, but very little is so far found out about the processes and mechanisms that affect their movement. When we are dealing with CORONA Virus-2 (SARS COV-2) many fascinating questions may rise that how aerial transport mechanisms transfer viruses from one place to another place and one person to another person. To stop the spread of CORONA VIRUS-2, their role and interaction with other living organism is also a big challenge to the researchers. Aerobiological and Vaccination Survey in and around the COVID-19 pandemic regions may help us to better understanding the relationship between aerial organisms, human health, ecosystems including important food resources.

2. Experiment

We all know that many types of allergies, and lung diseases are happening due to airborne bacteria, fungus, viruses and pollen. These living organisms are responsible for pulmonary, tuberculosis, Diphtheria or Influenza and many more pandemic diseases like COVID-19. A special risk management system is required to early find out about this type of viruses, fungus, pollens and protect all human beings. To identify and predict the trajectories of latest SARS COV-2 a mathematical models of diverse complexity may be used in order to exactly find out and estimate the risk of an infection at COVID-19 pandemic regions. From this investigation it is observed that long-distance aerial transport plays an important role in the advancement of pandemic disease analysis. This research investigation also aims to inspire other research teams

to continue venturing into the MAVs with an increased concentration in the applicability aspects using latest communications and global position systems on long-term behaviour and environmentally friendly rehabilitation technologies of COVID-19 cases.

3. Vaccination Supplier

In the present investigation we would like conduct complete aerial survey on COVID-19 and provide Vaccination to COVID patients on urgent basis using MAVs in and surrounding the COVID-19 pandemic regions. The survey would like to carry out in different selected pandemic regions in A.P and Telangana. The complete aerial survey data collected surroundings the COVID-19 regions using MAVs ease to greater understand the atmospheric shift of microorganisms. Unmanned Micro Air Vehicle (MAVs) along with fitted camera, Aero Core 2 for Overo MAV control board and different type of sensors used in the present module is shown below

Fig. 1. Unmanned Micro Air Vehicle (MAVs) for the aerial survey on COVID Vaccination
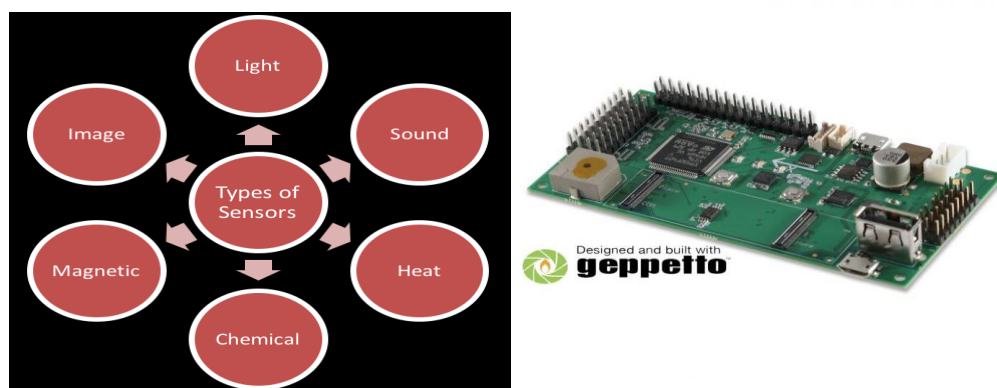
Fig. 2. Aero Core 2 for Overo MAV control board and different type of sensors

However this module is successfully tested in our laboratory and trying for the patent along with Major Research Project proposal for the implementation at ground level with the approval of permission of Govt. of India. Various workshops, seminars and webinars would also be conducted across Universities in order to ensure the dissemination of all knowledge gained during the investigation among University students.
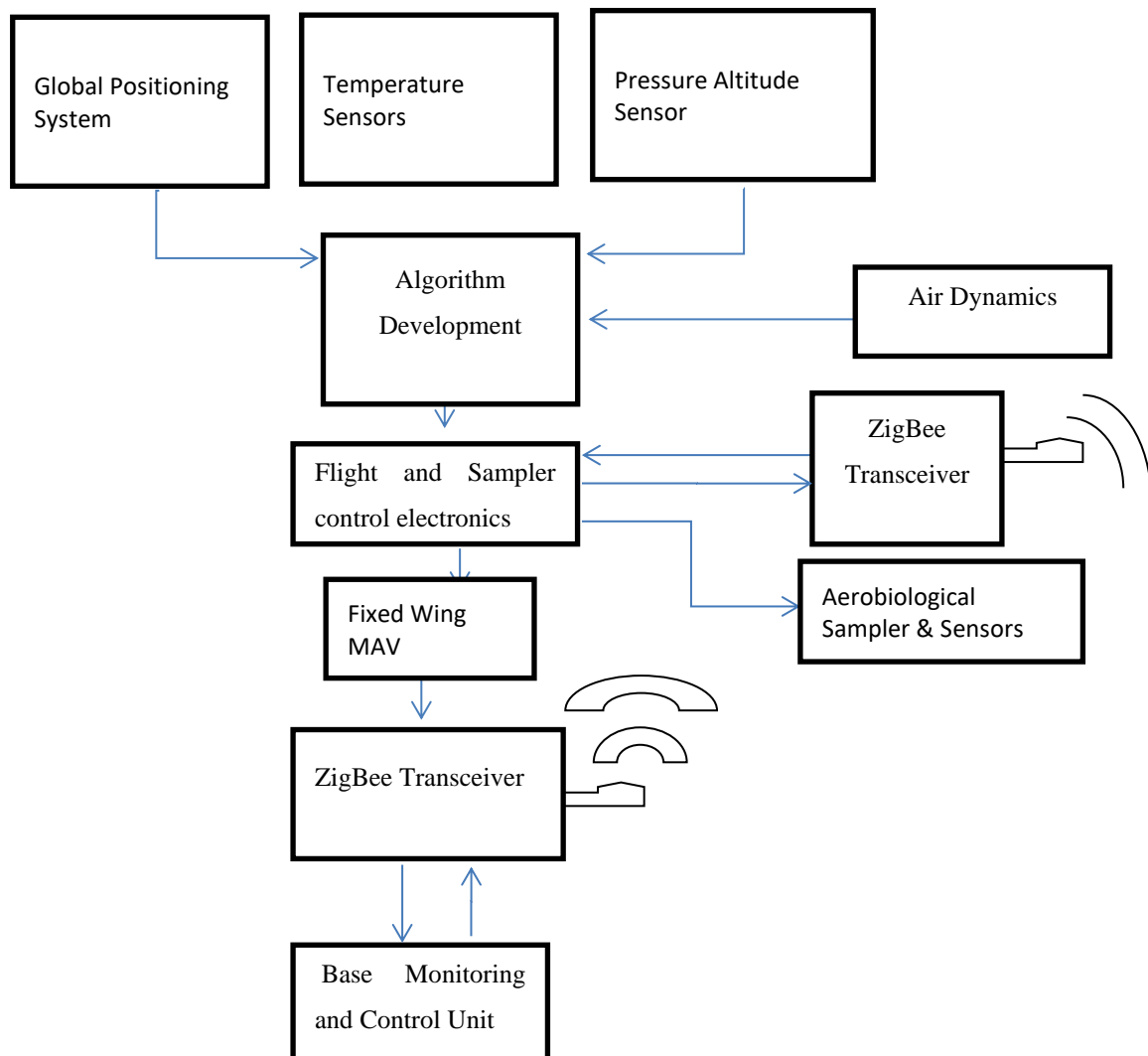
Fig.3 *Flowchart depicting the various stages of the investigation*
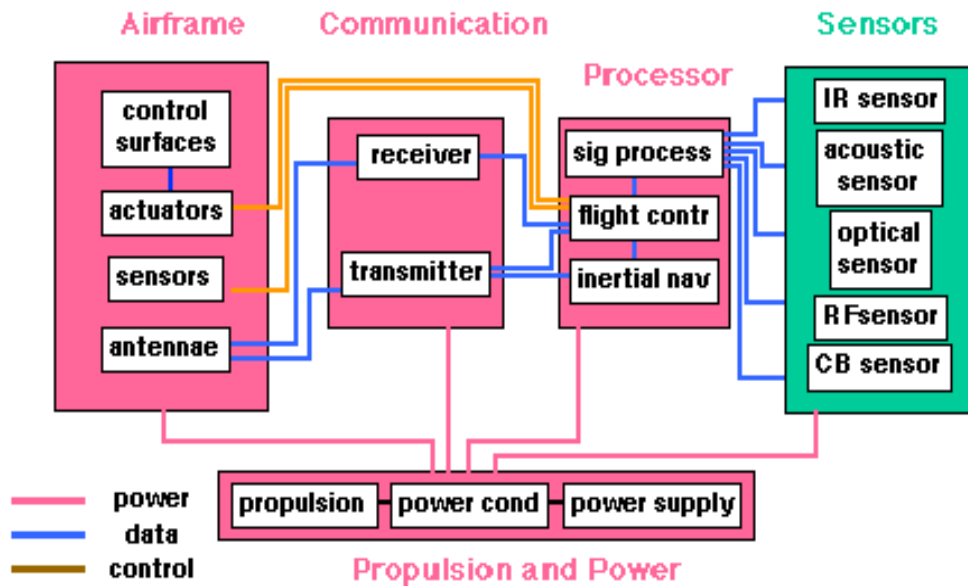
Fig.4 *Unmanned Micro Air Vehicle (MAV) system Modules*

Result

The study indicates that in the atmosphere there is huge variation in micro-organisms with respect to their abundance and distribution. Lot of unknown things has to be finding out in these micro-organisms present in the air. For this we have to thoroughly study two patterns i.e., environmental diversity gradient and the biogeography pandemic regions. Here we study the main characteristics of land i.e., water, air environments, and how these characteristics may donate to same and various biogeography patterns over these fields. With this study we can understand these patterns of lithosphere, atmosphere, ionosphere coupling concept.

A few different papers comprising of a few different approaches have been published on the use of the algorithms that are the topic of this research. Background information from a variety of sources has been studied to augment our knowledge and understanding of the technologies. In this present investigation we aim to draw from the substantial applied research available in this area to develop a novel application for MAV for the purposes of aerobiology studies in CORONA (COVID-19) pandemic regions.

4. Discussion

Our country is very active in aerobiology studies for the last six decades and contributing effectively to the world. The pioneer Aerobiologist from North East India was H.K. Baruah, Ph.D. (Cantab., 1942) and is now regarded as the Father of Aerobiology in the region. He started his aerobiological research works in the region after visiting Imperial College of Science

and Technology, London where he worked with P.H. Gregory, F.R.S. under Colombo plan fellowship (1955-56).

"THE INDIAN AEROBIOLOGICAL SOCIETY" (IAS)" is established in 1961 and is located at the Division of Palynology and Environmental Biology, Bose Institute, Calcutta. The society is affiliated with the International Association for Aerobiology. It is responsible for publishing research papers in Indian Journal of Aerobiology which is noted Indian Journal for Aerobiology. This society conducts various workshops and conferences in the field of Aerobiology. This motivation helps to conduct the complete aerial survey on COVID regions and supply of Vaccination accordingly. Govt. of Telangana also accords the special permission from Govt. of India to use this type of MAVs in this pandemic situation.

Recently we entered in to development of MAVs, Currently various organizations and research institutes like, SBMJCE NAL, IITs, IISc, MSRSAS etc. are undertaking research in the field of MAVs. A research work by a group from IIT, Kanpur focuses on Feature Based Object Tracking Using PTZ Camera. But our present work is very innovative and we wish to extend the scope and applicability of this by fitting GPS on MAV and enabling the MAV to locate exactly on required location and altitude for the study of CORONA (COVID-19) regions. However many International Research workers utilized this MAV for different studies.

5. Conclusions

An MAV system with the present Zigbee control can patented upon maturation of the results obtained through different funding agencies. Once a patent is secured, the technology could be used to identify and estimate CORONA VIRUS (COVID-19) regions, pollens distribution on crop fields, Ariel bacteria / virus estimation over pandemic atmosphere, epidemic virus estimation in humans at all etc. The potential uses with respect to epidemic virus distribution studies and supply of vaccination make it ideally suited for various environmentally socially beneficial activities. Various technical experts are being consulted for the development of the algorithms and the MAV. Technical consultancy to various Central Universities, State Universities, AIIMS, various colleges and industries like town planning, DRDO, Armed Forces, Large Dams, Private Security Agencies etc. would be provided after the investigation results reaches a certain degree of completion and robustness. The system once developed can be commercialized due to applicability in various health aspects, environmental, agricultural related activities. Good results economic gain could be expected to

begin shortly after the successful demonstration of the systems capabilities through proper channel.

6. Acknowledgment

References

1. Vincent C. C. Cheng, Susanna K. P. Lau, Patrick C. Y. Woo, and Kwok Yung Yuen, (Oct 2007) Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection, Clinical Microbiology Reviews, Vol.20, No.4, p.660-694.
2. D. E. Aylor, M. T. Boehm, and E. J. Shields (2006) Quantifying aerial concentrations of maize pollen in the atmospheric surface layer using remote-piloted airplanes and Lagrangian stochastic modeling. Agricultural and Forest Meteorology, 45:1003–1015.
3. R. Kennedy, A.J. Wakeham, K.G. Byrne, U.M. Meyer and F.M. Dewey (July 2000) A New Method to monitor Airborne Inoculum of the Fungal Plant Pathogens *Mycosphaerella brassicicola* and *Botrytis cinerea* Applied and Environmental Microbiology, Vol.66, No.7, p.2996-3000.
4. J. N. Bakambu and V. Polotski (2007) Autonomous system for navigation and surveying in underground mines: Field reports. J. Field Robot., 24(10):829–847.
5. D. G. Schmale III, B. R. Dingus, and C. Reinholtz (2008) Development and application of an autonomous unmanned aerial vehicle for precise aerobiological sampling above agricultural fields. Journal of Field Robotics, 25(3):133 – 147.
6. S. A. Isard and S. H. Gage. (2001) Flow of Life in the Atmosphere. Michigan State University Press, East Lansing, MI.
7. R. M Baxter, (1977) Environmental effects of Dams and Impoundments, Annual Review of Ecology and Systematics, Vol .8, 255-283,
8. Marcus W. Beck, Andrea H. Claassen & Peter J. Hundt, (2012) Environmental and livelihood impacts of dams: common lessons across development gradients that challenge sustainability, IJRBM, iFirst, 1-20.
9. Novel Coronavirus—China. World Health Organization (WHO). (https://www.who.int/csr/don/12-January2020-novel-coronavirus-china/en/ Retrieved 15 May 2020)
10. WHO Director-General's opening remarks at the media briefing on COVID-19 11th March 2020". (https://www.who.int/dg/speeches/detail/who-director-general-sopening-remarks-at-the-media-briefing-on-covid-19

The Role of Artificial Intelligence in Educational Technology

Mohd Yousuf[1] and Abdul Wahid[2]

[1,2]Department of Computer Science and Information Technology

Maulana Azad National Urdu University, Hyderabad, India

yousufmanuu@gmail.com[1], wahidabdul76@yahoo.com[2]

Abstract: *Today's education plays a vital role in an individual and a society's development. This research has been discovering and developing the use of Artificial Intelligence (AI) in the educational field and coming up with potential applications that have not been known before. Artificial Intelligence explores the new possibilities of innovation in educational technology. Today, artificial intelligence becomes the core element of the modern education system and a basic tool to get competitive advantages of the market. In this paper, we highlight the applications of artificial intelligence in education, existing tools and applications, the latest market and research trends, opportunities, pitfalls, and current limitations of AI in education. In particular, the work reviews the various applications of AI in education including classroom monitoring and recommendation systems, intelligent tutoring, sentiment analysis, students' retention, and drop out the prediction and student grading and evaluations. For this detailed review study, we analyzed various works of related domains, sub-domains like big data in education, educational data mining (EDM), and learning analytics. In this work, educational applications are analyzed from different perspectives. One side gives a comprehensive depiction of the platforms and tools developed as the product of research work. On the other hand, it recognizes the limitations, potential challenges, and opportunities for future enhancement, and this provides a baseline for future research in the e-learning domain.*

Keywords: *Artificial Intelligence, E-Learning, Intelligent Tutoring Systems*

1. Introduction

Human life is revolutionizing by the usage of Artificial Intelligence. AI is changing the field of education to unleash insights about understanding how to get crucial information on how to personalize the learning experience of students, how students learn, the way of obtaining more information for the decision-making process, and the tools that offer the way for students to access the domain. The complexity of the educational system and the difficulty of the traditional way of teaching can be address by Artificial Intelligence. A gigantic quantity of data is getting created by the e-learning that will allow AI to solve difficult challenges in education and provides advanced, smarter educational technology solutions.

Various government agencies, education institutions, funding agencies, and industries have been considering the role of AI in education as one of the hot topics of market research from

the last decade. Based on AI Market, the US Education Sector will grow by more than 47% from 2008 to 2022 [1].

By using data-driven methodologies, educational research problems can address using different approaches by different research communities. This is since AI communities solve research issues focusing on algorithms whereas the data mining research community focuses on big data.

Country	Government Initiative
Australia	Australia has become the first country in the world whose university named the Deakin University of Australia used Watson humanoid bot. Since then, the Australian government has started using Telepresence robots to teach children in remote areas.
Belgium	In 2019, the Flanders government signed a contract with the British platform century tech, which works on artificial intelligence. Using this contract, they planned to use this intelligent system in 700 universities and municipal schools around the region.
China	China started using educational robots to teach children educational programs and spoken language. For this purpose, more than 600 kindergartens started uses these robots since 2009. In 2018, Keeko robot-teachers were assigned in more than 200 schools.
France	Since 2008, Educational humanoid-type robots, manufactured in the county has used in the laboratories and universities of the countries.
Finland	In 2016, a robot named Elias pilot project launched in primary schools. It is a language and math teacher who understands and speaks 23 languages.
Germany	In 2015, Germany began to experiment with robots for teaching children and used it as a virtual assistant for university students.
India	Since 2017, families have started using social robots to interact with children. In 2018, to teach students from remote region government purchased Alexa robot.
Israel	In 2012, Israel made the first official robot for science and technology lessons to teach grades, 5-6 students. The bots were also used for kindergartens in starting 2016.

Japan	First, in the world, one of Tokyo's schools has experimented with an android-type robot that was able to express six emotions.
Singapore	In 2016, for preparing STEM school subjects, 160 preschools in the country began utilizing on-screen robots.
South Korea	In 2010, 29 robots used in 21 schools in Daegu as part of a large-scale robotics teaching project. In 2015, to study English 5,000 kindergartens used this robot.
Switzerland	In 2015, a robot used to improve children's handwriting skills of ages 6-8.
UAE	In 2014, 30 the most modern robots with integrated intelligence were launched in Merryland International School in Musafa.
United Kingdom	In the United Kingdom, to assist disabled children, Milo robots are being used in almost all schools since 2012. They teach thousands of students in more than 400 schools.
United States	The largest ECOT virtual school is established in Ohio in 2011 to use Artificial Intelligence in modern education.

Table 4.*Educational robots around the world*

Higher education is profoundly dependent on the development of new intelligent machines that are not only capable of processing huge data but should also be self-learning and improvement. In the educational market, many countries see Artificial intelligence as an international competition for the rapid development of advanced and smart education platforms and tools. Today, many countries provide the basic knowledge of artificial intelligence to its younger generation along with improving the quality of education.

Experts have both views regarding the usage of AI in educational purposes. The educationists and economists support the usage of AI, however, some of them contradict too. For example [2], AI algorithms help to enhance the level of education by data gathered from every interaction between students and teachers in virtual and physical classes. As Guilherme's research discovered there is a need for enough research on the integration of machine to person relationship with respect to education.

Scientists believe that AI has the potential benefits for the education that favor both teachers and students. According to a study [3], it is noted that future learning will be based on applications of virtual reality. Education will embed with the day to day latest technology and specialists will learn how and where this advanced technology may be useful. Authors of study [4]suggest that due to new technology and the arrival of intelligent computing systems, the role of teachers may change in the new education system. AI can replace real learning with virtual learning and can help students in many ways to enhance their learning. In most cases, the teacher may turn into a facilitator, whose work will be facilitating the new technological education paradigm.

The traditional old research problems of education are now under the scanner with several challenges are being investigated using AI and machine learning, and there is a hope of advanced and smarter solutions for education in the future. Similar to other fields, AI has applications in education as well; it has revolutionized education by enhancing course content, teaching methods, and other materials. In classrooms also, AI has changes things and enhanced their learning experiences. In education, AI could apply in many forms, such as improve course content, support teachers, grade students, and evaluate student performances, predict student performance.

At the international level, many countries begin to invest in artificial intelligence to teach its students with modern education. Table 1 shows such initiatives by governments around the world.
According to the above table, Artificial Intelligence is dominating the educational market with the latest and advanced research. Artificial Intelligence has both positive and negative aspects of different stakeholders of the education system shown in Table 2.

	Pros	Cons
Students	It enables students to monitor their learning progress. Students can opt for distance learning efficiently and effectively. AI helps in developing new and better learning materials by integrating with new technologies. It offers 24/7 digitals assistant to students. It provides continuous and more flexible training and learning sessions.	Low moral values support between teachers and students. Lack of motivation and encouragement for students to study.
Teachers and lectures	AI provides support for enhancing the training courses. AI offers content on smart devices, create tasks, and can manage a student's audience. Teachers and lecturers can easily evaluate the learning outcomes of students. AI offers the facility for providing feedback and assignment.	Teachers can replace by AI itself. It is not distracted or tiring but it quickly responds.

		Computer devices may experiences errors.
Universities	AI offers profile validation and verification, online programs, online admission, support of online exams. Protection of students' private data AI will turn universities into lifelong learning centers.	Robots are unable to maintain discipline in the class. Student's creative work is hard to evaluate.
Parents	Parents can check real-time feedback. AI provides quick information about progress. Low-income families can also access the education.	Constant communication with robots results in Dehumanization

Table 2.Pros and Cons of AI in Education

2. AI Applications in Education

2.1 Personalized Learning

Education is the most essential need for transformation the world witnessed till now, whether it is globalization, internationalization, and the educational management itself. The traditional learning which includes teachers, students in a static classroom with a unidirectional way of teaching, reading static text materials, and writing exam assessments to assess the learning skills of all students are being eroded. Contemporary learning directions congregate to interactive, tailored learning, student-focused models that assist a single or group of students with improved comprehension, closer interaction, better engagement, and broader scope coverage of learning outcomes.

The flipped classroom models are well suited for adaptive and personalized learning concepts. For instance, authors in [5] show recent educational technology advances and also instructional models design for facilitating a tailored learning experience. Several advantages of this method are identified by the authors that have a positive impact on the students that directly affect the final grades and dropout rates. In the flipped classroom model, technology plays a very crucial role in operations, feedback analysis, and design of the classroom models including analytics that offers insights about commitment from stakeholders and time management.

Personalized learning experiences have a recommender system for the support of recommending courses and study materials. For example, to personalized recommendation services, various web usage mining techniques have been used, and a personalized Web-based learning system was proposed in[6]. The web page classification method used for this approach, and as per detailed domain knowledge, attribute- oriented induction has been used. According to[7] in many online systems, as well as the e-learning process, recommender systems have been a beneficial tool for recommending items. However, when a learner uses any recommendation system with e-learning, less research has been done to measure its results. Rather much research has done on the performance measurement of the recommendation system while predicting the recommendation items instead of the learner's outcomes. This work focuses on the comparison of the learning outcomes when a learner uses many kinds of e-learning recommendation systems. After comparison, the authors proposed a new e-learning recommendation system, which uses content-based filtering techniques and previous excellent learners' ratings and recommends the study materials and resources and the framework enhanced the student's performance. The work proved that learners who engaged with this e-learning recommendation system had gained improved outcomes.Authors in [8] applied a recommender system to offer corrective actions for result-oriented shortcomings of higher education students like administrative level recommendations, for example, changing location in a study plan or changing a course's pre-requisite and including additional practice examples.

Personalized learning is an interesting method that provides unique e-learning content to individual students as per their requirement and assists to develop quality content for every student's needs. This method is very useful in supporting students during their study and avoids student's failure in the exams. Today, many education institutions are taking the initiative to implement this approach. Some examples of personalized education platforms are illustrated in Table 3.

Platform Name	Stakeholders	Purpose
Cerego	Companies and Universities	It is an algorithm-based platform that makes an optimum schedule for proficiently directed learning and assesses individual memory retention.
Highlight	University Students	It is a cloud-based tool using which student's progress may track.

Immersive reader	Students with dysgraphia and dyslexia	It is a learning tool that is accessible to everyone who needs an easy reading but this tool is specially designed to support students with dysgraphia and dyslexia.
Knewton	College and University Students	It is an Analytics-based tool that recommends personalized content to the learners based on their learning pattern.
Watson Education Classroom	College and University Students	It is a cloud service solution that improves student outcomes by helping teachers improve teaching classes by personalized content for each student.

Table 3. Personalized learning platforms

2.2 Intelligent Tutoring System (ITS)

The increasing gap between the shortages in qualified specialist teachers globally and the growing number of learners can overcome using the Intelligent Tutoring Systems (ITS). For supporting and enhancing traditional school curricula in thousands of schools, the US and other countries are currently using ITS systems [9]. Knowledge-based domain information is used to design such ITS systems. When combined these recent technologies with the educational domain transforms towards more tailored, complex and interactive learning approaches like learning by teaching or by games, thus proving that with the knowledge-based approach, limitations of ITS systems can make more apparent.

The Dynamic problems that require constant learning can be suitably solved using AI and data science technologies. It is not a new thing as these technologies extract new, knowledge, and unseen insights from high-dimension, non-structured data in a much more effective way than mining knowledge or expertise from human teachers. AI is also very operative in predicting student skills, mental states, results, and cognitive needs and consequently recommending the right course of action. For instance, ITS systems with AI enhancements are applicable in modeling efficacy, student emotions, and ability to perform scientific inquiry, and then generate recommendations automatically [9].

Open Learner Models (OLM) which is a branch of ITS systems whose purpose is to open up AI learner's model in terms of learning and teaching, human cognition. While ITS systems research focuses on how AI can be used in education efficiently whereas OLM research focuses on the essential components needed to make AI models explainable and interpretable

with respect to learning. Interpretable AI offer a framework for the implementation of knowledge-based and AI systems in education and beyond.

2.3 Classroom Monitoring and Visual Analysis

In modern offline and online education, students have several technologies- ambitious advantages at their fingertips but sometimes suffer from demands which lead to dropouts and classroom under-utilization. Measuring or increasing the efficiency of room utilization by using modern AI technology as instruments is a new topic predicting room utilization as an age-old problem. To monitor classroom attendance, authors [10] used on-campus sensor instruments while respecting student privacy. For identifying the best sensor technology, many measurement methods are evaluating in a lab experiment in terms of convenience, accuracy, and cost.

AI applications supported Technology Enhanced Learning (TEL) in many sub-fields. One such field is to understand the difficult task of understanding the various dimensions of TEL in schools. The reason behind such difficulty is the limitation of monitoring classrooms for an extended period to analyze students' learning experience and teachers' teaching methods.

Authors in [10], used sensors to collect observation data for more than two months of TEL classrooms. This collected data is observed, analyzed, and visualized over time and offered as insights teachers and academic administrators for reaction and corrective action enhance student learning. In a TEL environment, EDM and AI methods are deployed to handle the complex learning aspects at a higher level of precision.

AI and EDM play a vital role in the transformation of the traditional classroom into modern education. One such application comprises an assessment of the student level of engagement in the classroom by analyzing facial expressions of students. Authors in [11] proposed a system that determines student participation and their level of enjoyment in the class using AI and facial recognition technology that identify student emotions by accessing data feeds from video cameras of the classroom. Today, due to the generation of huge online data, much of the research is focused on digital and online learning environments and MOOCs. Apart from this, the physical classroom has gained the attention of many researchers as well. Authors in [12] studied different technologies developed to collect and analyze educational data and reviewed some case studies. Both learning environment digital and physical classroom settings are studied and presented many features learning environment. For quantifying teaching and learning processes, different aspects of the learning process have been assessed

and analyzed. The authors have also presented the pipelines that control data and information from both digital spaces and physical spaces.

2.4 Student's Retention and Dropout

Both online and offline learning platform facing student dropout problems and student retention and dropout is a universal factor which affects all over the world. In the US, every year, there is a vast growth in the dropout rate of Baccalaureate programs that are reaching nearly 30%, and it is hard to address because of insufficient quantitative analysis of causes and remedies. Authors in [13] have analyzed a dataset of 32500 students at a public institute for modeling the student dropout rates. They concluded that most of the dropouts happened at the early stage of the courses. Here, AI can play a vital role in predicting and preventing the causes of dropouts. If we talk about MOOCs environments, the dropout rate is considerably higher compared to offline programs.

AI has also been proved very effective in students' retention and dropout prediction similar to other application domains. Using AI, dropout rates, and dropout probability is calculated by identifying the demographic, ambient, and distinct aspects related to learning activities so that effective intervention and prevention remedies can be designed by the education administrators to control the dropout rates. The accuracy of numerous AI algorithms concerning the prediction of student retention rates at university levels has been analyzed by authors in [13]. They found that with random sampling Random Forest (RF) algorithm is an optimum method. The proposed system also handles huge extents of dynamically shifting data including user evaluation metrics and data structures.

Authors of [14] conducted qualitative research with a sample of 75 professors, tutors, and researchers. They used a combination of sequential, systematic, thematic, and explanatory approaches on findings from thematic analysis. They discovered that many factors contribute to the rising dropout rate regardless of education institution efforts. These factors include mental health, family pressure, workload, attendance, and engagement. Apart from these students' retention and dropout is depends on some critical factors such as academic integration, institutional social-environmental contribution, and the sense of community. Thus, for predicting the dropout rate, there is a need to go beyond basic modeling and analyze the impact of demographic, psychological, family factors, and socio-economic factors to conduct an effective analysis of causes of dropout.

To analyze large datasets from MOOC systems, a study in [15]applied visualization analytics methods and techniques. This data correlated with predicted dropout rates with the learning activities of MOOC subscribers visually. The visualization analytics performed to allow AI experts to design better predictive models and content designers to project more engaging and suitable content.

2.5 Student's Grading and Evaluation

For the prediction of student behavior and in-class student performance, accurate models are being developed in the market using several AI techniques. For example, two wrapper methods were proposed by researchers in [16] for semi-supervised learning algorithms designed to predict student performance in their final examination. While training the models with unlabeled data, higher classification accuracy is achieved for semi-supervised methods[17].

In the Modern education system that is supported by AI technologies predicting the performance of students is very important that helps education administration to take necessary measures. These measures will prevent the student dropout rate at the end of the semester and helps in identifying the weak students who need extraordinary support. Student learning difficulty is also one of the aspects of predicting student performance apart from grading and marks. The learning difficulties faced by students in a digital course has been predicted and analyzed by authors in [18]. For this purpose, data logs are analyzed that are collected from a Technology-Enhanced Learning environment (TEL) system. The above task has been performed by many Artificial Algorithms including Naïve Bayes Models, Decision Trees, Artificial Neural Networks (ANN), Support Vector Machines (SVM) as well as Logistic Regression. While solving varying difficult digital design exercises, student click behavior is monitored and collects meaning information such as average time, the total number of activities, total related activity for each exercise, the average number of Keystrokes, and average idle time. In news sessions, AI models were trained using old sessions to predict student performance.

2.6 Recommendation Systems in Education

The Learning Management Systems (LMS) can be easily incorporated with AI services due to the advances in the design and accessibility of AI packages and tools. The LMS collects and stores student assessment results and submits reports and elementary analytics to academic managers either for enhancing the quality of the programs or for daily operations

processes. The ITS, has proved and achieved better results when compared with the studies from printed materials and traditional classroom instruction.

To improve learning quality, various AI and data mining algorithms can be applied to recommend remedial actions in academic learning. For motivating the learners to participate in an e-learning process for enhancing their learning vigorously, a customized web content recommendation system has been proposed in [19]. Usage mining and web content techniques have been used in this system. Meaningful web content has been produced by using Web mining. To distinguish e- Learners' navigational examples, web utilization mining is utilized. E-Learners navigational patterns are useful to perceive the shortcomings and premiums, and again visited web content mining and helped to anticipate student's presentation. At that point, the proposed framework could give customized, successful, and productive web content. A personalized recommendation has been provided by web content mining and Learner preferences, which can be clustered by collaborative and content filtering techniques.

Training data like course level, course domain, lab option, and section size are enough to assist experts in selecting remedial action that is recommended for subsequent assessments. For each rubric line from the master pool, appropriate actions are selected by using a multi-label classification algorithm. In the application of remedial actions, AI provides obvious strengths and is manifested with fairness, consistency, efficiency.

2.7 Student's Performance Prediction

The powerful educational platforms can predict a student's future performance in a course that facilitates educational interventions and remedial actions promptly. The Educational Data Mining (EDM) provides some research areas and most salient applications that support the development of AI models for the prediction of student performance and uncovering hidden insights and patterns. Nowadays, in the field of academic performance analysis and prediction, many studies are being conducted. One such study has been conducted by authors in [7], who predicted the academic performance of university students by empirical investigation and comparison of several classifiers, data sources, and ensembles of classification techniques. They combine information from several data sources against a single-sourced trained model. Further, they compare and analyze the performance of ensemble techniques. To achieve this, many algorithms that include SVM, ANNs, and DT were used and compared individually.

For predicting student performance, an ensemble-based semi-supervised approach is also proposed by authors in [13]. Early prediction results show sufficient accuracy. To notify

students about their probable outcomes early on in the academic semester, authors designed an AI model targeting students in introductory programming modules. The authors achieved better results in terms of accuracy and F-Measure with Decision Tree (J48).

2.8 Sentiment Analysis in Education

To better understand student opinions and make adjustments to the content or presentation of the learning material, sentiment analysis of attempts to enhance the learning process in an e-learning environment by analyzing students' feedback. Sentiment analysis is a very useful technology and already shows its results in other fields such as natural language processing, social media, healthcare, and education, etc.[20]. It is sometimes known as Opinion Mining. It is a difficult job that involves various stages, such as collection, storage, and analysis of collected data using a combination of machine-learning and knowledge-based techniques.

There is enough amount of research done on the sentiment analysis of student's feedback involving text or their social media posts about materials, teaching methodology, and curriculum. For example, students' learning diaries were analyzed by authors [21] to predict students' emotions, sentiments, and opinions about their learning experience. According to[22], for the effective development of e-learning systems, knowledge, and evaluation of user opinions is a crucial prerequisite.

To classify emotional aspects of students, researchers[23]propose a Temporal Emotion-Aspect Model (TEAM) that tracks emotions over time with two main outputs: a) emotions evolution over time b) aspect probabilistic distributions that are emotion-specific. Authors discuss that for a better understanding of learning requirements, the temporal nature of student feedback in MOOCs environments stipulates that students' emotions and learning activities can be tracked.

2 Market Organization and Tools

There are rich AI applications in education. The leading market companies are Microsoft, Google, Quantum Adaptive Learning, Cognizant, Nuance Communications, Amazon, IBM, Pearson, Blackboard, and Third Space Learning. To help learners and teachers in diverse fields, these organizations developed numerous tools and platforms.

This section discussed such platforms and tools and their key features. Based on the application of the education domain these platforms and tools are classified. Fig. 1 discussed the taxonomy of AI-employed educational tools.
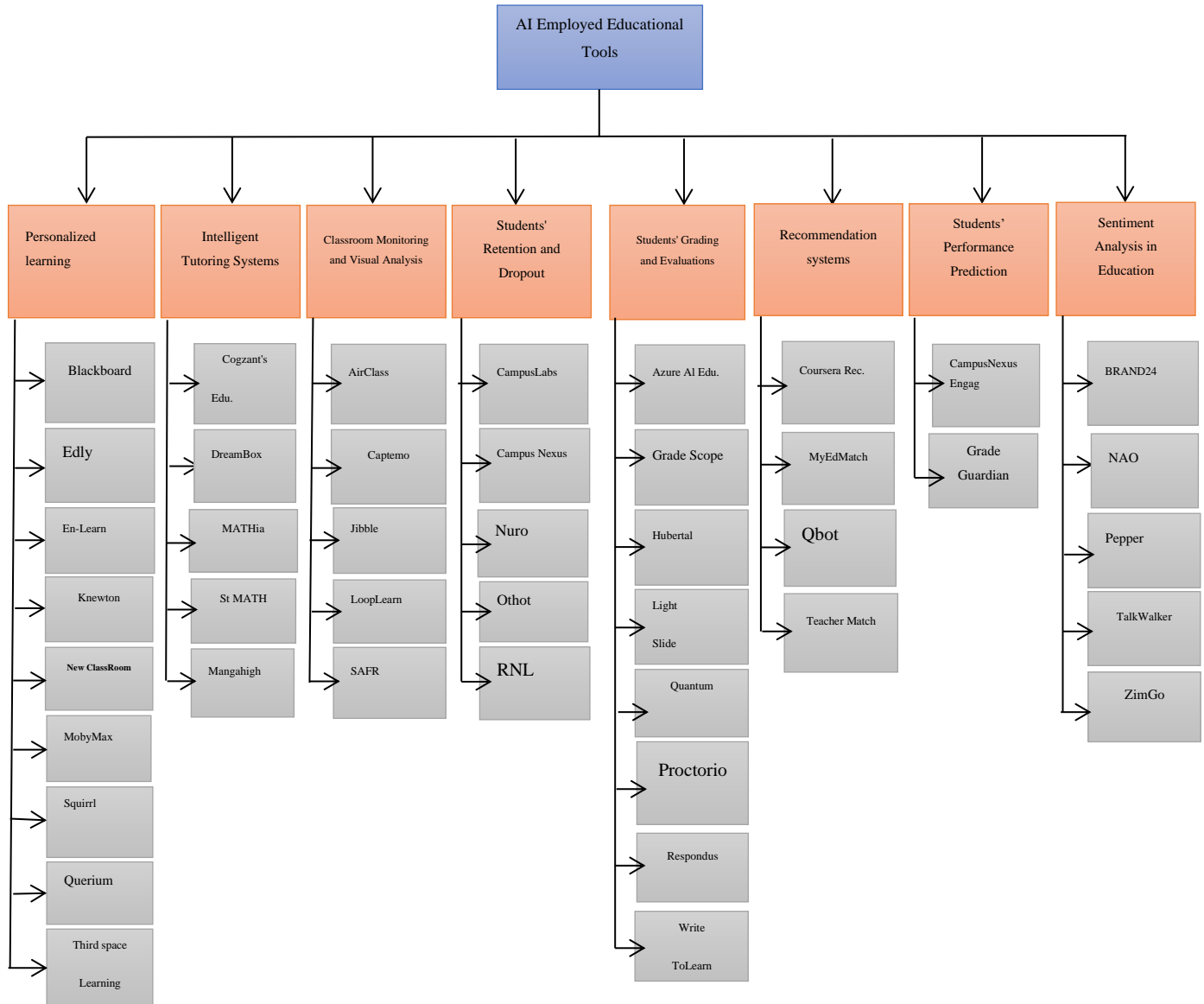
Fig. 1. *AI-Based Educational Tools*

3.1 Personalized Learning

Personalized learning platforms and tools use AI algorithms to support instruction based on the student's level of understanding and previous knowledge. Some of such personalized learning tools are:

Edly: This tool equally assists both teachers and students by offering them actionable insights via learning analytics. By employing AI algorithms in diverse features of the learning analytics, this tool makes the learning process smoother and better. On the one hand, it enables teachers to keep track of students' progress during the course. On the other hand, it assists students by designing the course contents as per the learner's capacity.

EnLearn: In creating personalized content, this tool offers a complete adaptive learning ecosystem by including all participants such as teachers and curriculum, students. Further, for targeted students, it extends its content instead of just trundling the content.

IBM Watson Content Analytic: To improve learning outcomes, millions of college students and professors use Watson's cognitive capabilities. It assists several students to succeed by providing personalized content for students based on mastery.

Knewton's Alta: Alta is an adaptive learning tool that offers personalized content to students in various domains of mathematics, engineering, science, and technology. This tool creates personalized content using AI from openly available high-quality content, resulting in better quality inexpensive and globally accessible personalized learning materials.

NewClassrooms: This tool schedule personalized math learning experiences using learning analytics.

MobyMax: For all K-8 subjects, this tool uses AI to determine and fill the learning breaks with flexible and distinguished learning materials. Students can learn at their speed with lesson plans and automatically generated practice sheets.

Squirrel: An AI Learning tool that comes with a supervised adaptive learning experience. The Squirrel tool is powered by AI-driven adaptive engine and custom-built courseware.

Querium: This tool offers step-by-step, bite-sized lessons, and personalized tutoring assistance. The platform uses AI to assists learners with STEM abilities so that they can learn advanced education.

Third Space Learning: This tool assists learners in mathematics with premium math resources and personalized content. This platform also provides weekly reports of students' progress for the purpose of targeting individual students. The platforms assess all the registered students and design personalized materials for each student before proceeding with the training.

3.2 Intelligent Tutoring Systems

Cognizant's tools for Education: The tool uses Virtual Reality (VR) and AI to offer the learning experience. This tool also improved students' experience across all corners of students' journeys, including reporting, assessments, learning, onboarding, admissions, and e-credentialing.

DreamBox: To enhance students' analytical skills and decision-making capabilities, this tool uses AI techniques. DreamBox developed to assist learners in solving and systemically understanding mathematics and promote student's growth and deeper conceptual understanding of the subject.

Carnegie Learning's MATHia: MATHia is an AI-Powered tool and closest to a human coach. This tool can mirror a human tutor by utilizing AI and cognitive science techniques. To facilitate a better math learning experience for every student, MATHiaU is intentionally designed by a team of cognitive and learning scientists.

MANGAHIGH: It is a game-based tool for learners that involves solving social competitions, interactive games, and puzzles to excel at mathematics.

ST Math: It is a game based Pre K-8 tool, which is used to solve mathematical problems by leveraging the brain's innate spatial-temporal reasoning ability. The platform offers students a unique approach to learning by resolving challenging puzzles and non-routine problems.

3.3 Classroom Monitoring and Visual Analysis

The classroom monitoring and visual analysis tools are used for monitoring students in the class, attendance management, and visual analysis of the virtual class. Some of such tools are:

AirClass: It is an emotion and facial expression recognition system that works by automatically analyzing student's responses to a lecture. The tool detects whether student's eyes are opened or closed during the lecture using facial expression recognition algorithms. Further, it analyzes the learner's commitment and interests in learning through facial emotion recognition and analysis.

Captemo: Emotion Recognition: It is an AI-based tool that recognizes student emotions and expressions in a classroom for various reasons. This tool can also be used by the teachers to know the feedback about the effectiveness of their teaching. Using this tool, student behavior in a classroom can also be monitored by the educator.

Jiblle Attendance Platform: It is an AI-based attendance tool that uses information from many data sources such as GPS data and student's photos. To ensure a reliable and cheat-free attendance system, this tool uses facial recognition algorithms with GPS data. The platform also generates reports and retrieves time-sheets with actionable insights on student's attendance.

LOOPLEARN: Automated Attendance Keeping: It is an AI-based automatic attendance tool that relieves the school administration from the tiresome job of marking students manually by automating the school attendance by employing the AI and Computer Vision (CV) techniques. Using this tool, we can generate insightful reports on students' attendance.

Secure Accurate Facial Recognition (SAFR): SAFR is a general-purpose AI and AI-based facial recognition platform for various applications like analytics, convenience, and security. It is also used for security purposes in schools for providing security to school kids by allowing authorized people to pick students from schools. For this process, the door of the school is opened only if the tool identifies the kid's parent or guardians photos that are stored in the database.

3.4 Student's Retention and Dropout

Educational interventions and remedial actions are the services provided by the Student's retention and dropout tools. Some of such tools are:

Campuslabs: Campuslabs using AI algorithms to offer actionable insights from campus-wide data for an early alert on student retention. The tool can recognize the student at risk and various causes for student's retention and supports them by identifying the best resources and most effective programs for the students.

CampusNexus Succeed: This tool employed AI algorithms that focus on institution-wide engagement and specialized student retention functionalities. Due to the presence of advanced predictive analytics and AI algorithms makes it the first choice of educators to monitor and identify students at the risk changes the policies and re-allocates the resources accordingly.

Nuro Retention: To cope up with different scenarios of a student life-cycle with real-time data analytics capabilities, Nuro retention covers different aspects of a student life-cycle. The tool utilizes data from many datasets such as SRP, SYSA, MYSA, and CSI, which are integrated with AI algorithms for students' retention prediction.

Othot: The tool collects data from various sources and supports advanced data analytics to provide various participants in education throughout student life-cycle. The platform contains AI algorithms in the predictive analytics module that assist all participants in decision making about enrollment and students' retention.

RNL Student Retention Predictor: Using this tool administration can predict students' retention and students at the risk of dropout. This tool works by allocating from 0 to 1 for each new

student. Using this scoring technique administration can keep an eye on students and assist them to associate resources whenever they needed. It also recognizes the possible reasons and variables impacting student retention directly on an institution.

3.5 Student's Grading and Evaluation

Students' assignments and tests can be analyzed, assess, and score by automatic grading tools. Today, for marking student's assignments and tests, various tools are available at large scale. Some of them are:

Microsoft's Azure Cloud AI Tools for Education [84]: Media Analytics, API, Cognitive Services, and Office Graph API can be integrated with AI to offer deep analytical insights into student performances, and then using Microsoft PowerBI it can be visually displayed.

Gradescope: This tool evaluates, grades exams, course work comprises projects, quizzes, technical reports, and assignments using AI algorithms. Both fixed templates and variable-length assignments can be evaluated using this tool.

Hubert.ai: This tool provides a cognitive computing assistant using which test-taking feature can be enhanced further. Using this assistant surveys can be converted into dialogue tests through back-to-back questions to students. A deep learning-based text analytics model analyses every response of the student that automatically analyzes and categorizes feedback. Essential learner abilities and skills like background reasoning, ethical reflection, creativity, and imagination can also be evaluated using this tool.

Turnitin's Lightside: This tool equally assists both teachers and students. For the teacher, it helps to evaluate and mark the student's work. For students, it assists students in evolving their writing skills by offering them suggestions and feedbacks. For checking and evaluating their writing skills, it provides a spell checker. Quantum Adaptive Learning and Assessment: This tool interprets student work by using artificial intelligence and explains each step along with feedback on whether students' answers are right or wrong.

Proctorio: Proctorio facilitates educators in many features of examinations, content analytics, and content protection, identification, and verification of candidates' IDs that make this platform a fully automated exam proctoring tool.

Respondus: It is an online evaluation and testing tool specially developed to assist in making online exam questions and inhibiting cheating. For this purpose, using its exam tutoring tool a huge collection of online exam questions is offered.

WriteToLearn: It is a web-based tool that works by evaluating students' writing skills and check the understanding of the text as well as evaluate the student's skills of grammar and spelling. The aim of this tool is to develop students' ability to summarize what they learn.

3.6 Recommendation Systems in Education

Recommendation systems work by recommending the personalized course, stud material, and learning path for the students. Some of such recommendation systems are:

Coursera-Recommendation Tools for Courses: It is an AI-based recommendation system that recommends courses to students based on student interests. While recommending process, this tool considers different parameters such as current courses registered by students and student's knowledge.

MyEdMatch: It is a platform that connects schools and teachers with common interests, goals, and beliefs. This tool makes the recruitment process faster and smoother by equally supporting administrators and teachers.

TeacherMatch: It is an AI-based tool that analysis the students in the pool maintained by the platform. The valuations are based on the four aspects, namely, attitudinal factors, cognitive abilities, teaching skills, and teachers' qualifications. For this purpose, to predict the teachers' impact on students' achievement, this platform uses an inventory like Educator's Professional Inventory (EPI). This platform uses AI techniques to identify useful insights in experiences, abilities, skills, knowledge, and other features essential for teachers' effectiveness in a certain context.

Qbot: It is a joint venture of Microsoft, Antares Solutions, and the University of New South Wales, Sydney. Qbot is an AI-based chatbot that assists students in many ways. Using this tool students can keep themselves busy in an interactive conversation and answers their questions. A learning platform is built using AI, which can be accessed by students throughout the day. Qbot also recommends students various leaning materials and online sources to benefit them.

3.7 Student's Performance Prediction

These tools help in analyzing and predicting a student's performance in examinations. Some of such platforms are:

CampusNexus Engage: To improve engagement and drive students and institutional success, this tool makes use of advanced AI and business analytics. CampusNexus engage offers many services that comprise predictive analysis to predict students' performance.

GradeGuardian: This tool supports various participants including students, advisors, educators, and policymakers by providing a complete solution for modern education challenges. This platform predicts schools' and students' performance in both the short and long term by utilizing advanced AI algorithms.

3.8 Sentiment Analysis in Education

Sentiment analysis tools can assist to enhance the learning process in an e-learning environment and a classroom. This tool works on better understanding student's opinions and emotions by analyzing the facial expressions and modifies the content or delivery of the respective learning material[24]. Some of the tools are:

BRAND24-AI-Driven Sentiment Analysis: It is a general-purpose sentiment analysis tool, which can be used in various application domains including education. Students' negative feelings about the learning process can be detected using this platform the teaching process can be adjusted as per these feelings to change the negative attitude of students about learning.

SoftBank Robotics's NAO: NAO works in the same way as Pepper does and intends to assist businesses in dealing with their customers/visitors'. NAO is extensively used in education and research also it has adopted by other businesses and health care centers. NAO is armed with numerous microphones, touch sensors, and speech and object recognition capabilities.

SoftBank Robotics's Pepper: This tool aims to facilitate interaction with human and personalized recommendations. Pepper engages humans in an interactive conversation by possessing face and human emotions/expressions recognition capabilities. For any business domain including education, Pepper could also be optimized.

Talkwalker: This sentiment analysis tool is used in education and it generally monitors social media networks and extracts meaningful information. To promote institutions in a better way, this tool can help educators in terms of social analytics.

ZimGo: This tool can be used for various purposes in education. This tool can be used to extract and analyze people's feelings, attitudes, and responses by utilizing AI and Natural

Language Processing (NLP) tools. Based on the data obtained from social networks of poor students, this tool has been modified to continually monitor students for signs of stress and depression.

4. Discussions: Insight, Pitfalls, Future Research and Open Issues

4.1 Insights

Today, the AI impact on education is continually rising. Education is becoming dependent on information technologies and the capability of these technologies to assimilate various data from various sources. Implementing education with AI is becoming an essential requirement instead of a luxury. It is very difficult for someone to ignore what AI can provide concerning education to solve problems of current and future education. For instance, AI can be used to solve safety and security issues in schools. By the integration of AI, safe cyberspace, classroom, school, and educational environment can be possible.

4.2 Limitations of AI in Education/Learning

Over the last few years, by delivering outstanding performances in various application fields AI has revolutionized the education sector. The overall quality of the education system is getting improved by data analytics and mining techniques. For example, personalized learning programs and AI-based interactive tutors are assisting students a lot. Likewise, the administrative tasks in the education sector are significantly reduced by AI. However, the requirements of modern education cannot be fulfilled by AI alone. There are many features of education where AI alone cannot add much. In education, AI has some limitations and pitfalls as well that are mainly categorized based on social and technological aspects. There can be technological pitfalls of AI that may be due to training data or because of conceptual/ algorithmic limitations. Some of such pitfalls are:

Extraction of interpretable and actionable insights is difficult: To enhance students' learning, the extraction of actionable and interpretable insights from the educational data using AI is difficult. For example, authors in [25], present many case studies where it shows AI-based predictions are not enough to understand and improve the learning process. Instead, the authors proposed an explanatory learning model by employing AI and Human-Computer Interaction (HCI) techniques. Using these techniques, the authors' show how technology could be made more useful for learners.

Course content generation Failure: At a timestamp, for a student, AI techniques recommend particular chapter/course content. According to authors[26], learner's performances can be impacted when the same material is presented in a different sequence. The generation of course content on an urgent basis is a very challenging task.

Lack of clarity and flexibility: There is a lack of clarity and flexibility when students are getting trained with virtual teachers (Robots). Further, a teacher is one who motivates his students during teaching but the robot lacks such capabilities. While the creation of personalized content, learning analytics, and content creation contributed to a greater extent but there is a lack of clarity and flexibility when this content is delivered by virtual teachers (robots)[27].

Lack of training data: The AI techniques work well when there is an adequate amount of training data, which has a significant impact on their prediction capabilities. In a sensitive and high-stakes environment, it is very difficult to obtain a sufficient amount of training samples, such as the education sector, where institutions don't want to take any risk with students [28].

High risk due to biased data: For accurate analysis and prediction, there is a need for accurate and reliable data. To make accurate analyses and predictions, AI algorithms need accurate and comprehensive data. According to authors of [29][30], AI in education is highly vulnerable to the risk of biases, due to which there is a higher chance of inaccurate conclusions and false predictions.

Security concerns: With the increasing popularity of AI, dependency on AI is also increasing which is leading to various serious privacy concerns [4]. The education institutions not only focused on quality but also on data privacy. According to the authors of [15], school administrations need to be ready for AI from a policy point of view and need to handle student data carefully.

Apart from the limitations of AI in education, some pitfalls are not directly linked with AI algorithms but have a negative impact on society. Such pitfalls are:

AI may put kids at risk: Deploying AI in education has both pros and cons, due to AI students may get addicted to technology like tablets and phones, which is dangerous to their personality and health. Due to access use of technology in the learning process also causes students to get isolated and limits their social interactions. AI may also limit students'

problem-solving capabilities and creativity by increasing student's dependency on the machine to solve every problem.

Increase the gap between rich and poor: Installing AI tools in education will enhance the dependency on the luxurious technology that may keep poor students away from quality education.

Increases the power cost: AI deployment in education leads to more power consumption that will increase the schools' budgets.

Increase joblessness: AI deployment in education causes joblessness, similar to other domains, in education when AI is deployed at a large scale, this will decrease the workload considerably, leads to joblessness in the market. In the education sector, AI could be implemented at various levels such as security, teaching, and administrative tasks, which may significantly reduce human labor.

Isolation and Individualization: Instead of collective learning and teaching, AI in education may lead to isolation and individualization.

4.3 Future Research Directions

AI has great future research directions concerning education. Future research directions on AI in education can be categorized into four areas:

4.3.1 Teaching Methods:

To recognize the finest teaching pedagogy that matches each learner interests and skills, much research is still needed. With the help of a customized teaching pedagogy, students can grasp the new concepts and course material very efficiently [31].

Personalized Tutoring: For diverse levels of learners, it is very difficult to provide personalized tutoring and real-time feedback beyond math topics. The MathiaU is a personalized tutoring tool which offers a better math learning experience to every developmental Math student [32].

Technology Integration in Classroom: Trial-and-error is the model using which most technology products are brought into classrooms. The role of integrating technology is given to teachers while the task of learning all subjects using technology is assigned to learners. Due to many reasons, designing and developing a model for effective technology integration is challenging and critical. Knowledge areas, learners' skills, and interests and technology

products are very diverse. Many researchers focus on and trying to find the best way of integrating technology in classrooms [33]. Technology integration has some possible clear negative impacts that cannot be overlooked [6][34]. This makes technology integration a challenging task concerning the development of an effective AI model.

4.3.2 Supporting Educator Effectiveness

Minimizing Biased Evaluation: Minimizing personal biases is one of the main challenges that each educator face when it comes to evaluation and grading. When it comes to relationships and judgment, human behaviors are hard to predict. In this situation, AI can help in protecting against the internal biases by providing insights into student's performances based on data. It is very challenging and needs careful considerations while designing AI techniques that can assist in decreasing biased evaluation while keeping teacher's attitudes is mind.

Identifying Students at Risk: AI can play a crucial role in detecting and predicting student's dropout rates.

Scheduling Efficiency: Optimal scheduling of learning lessons and activities are connected to optimal learning. Due to various contributing aspects like understanding how people learn a topic, availability of qualified teachers, level of the learner, age, and availability of resources, the design of optimal teaching schedules become very difficult. Effective scheduling optimized, and adaptive teaching policies are the topic of research concerning AI modeling. Online job scheduling using AI has been proposed by the authors in [35].

4.3.3 Improving Education Systems

Predicting student's future: Due to the various student's backgrounds, needs, environmental aspects, biological differences, skills, etc. developing AI tools to predict the best career paths and specialization areas is a challenging task. To intelligently predict a student's future, a comprehensive AI software application is needed. Authors in [36] discussed a study that predicted employment at graduation using AI tools.

Mistakes Implications: There can be critical implications when we commit some mistakes in education. When AI is used in making decisions, it is crucial to recognize the potential risks and consequences at various levels. If the AI algorithm recommends the incorrect study material or unsuitable clip to learners, it can lead to serious economic or social issues.

Therefore, before implementing any AI technique in education, researchers need to integrate a risk factor to quantify error implications and potential mistakes.

Quick Generation of Course Contents: In course content recommendations, AI techniques have been proved very effective. Though, it is fascinating to research how AI can be implemented in course content generation for an individual learner [37].

4.4 AI Issues and Concerns

Identifying Ethical and Privacy Issues: While AI provides solutions in various domains comprises education; various ethical concerns may come into the light and cause restrictions. Developing AI algorithms for education with keeping ethical concerns in mind is a challenging task. Also, it is serious to prevent using AI is leading to critical biases when it comes to identifying patterns by analyzing data. When machine analysis our private data and detect a pattern, this itself is a privacy concern. For instance, accessing student's online search data and detecting behavior patterns can negatively lead to long term impacts. Therefore, AI researchers must find ways to train algorithms and analytics. Various ethical issues of using AI were addressed in the study discussed by authors in [38].

Security Implications: While designing an AI algorithm, security is very critical and very prominent. Before applying AI techniques on educational data there is a need to distinguishing between sensitive and insensitive data. Hence, such intelligent AI techniques need to be developed that can deal with data in classified and careful ways.

5. Conclusion

This paper reviewed AI applications in education from various points of view emphasizing the market prospects and the future scope for AI in education. We have also presented the latest research trends in AI, already existing tools and applications developed using AI in education, present limitations, and pitfalls of AI in education. Specifically, we presented a detailed summary of the literature in many application domains such as classroom monitoring, intelligent tutoring systems, sentiment analysis, students' dropout, and students' grading and evaluation. We also presented recent key market players, platforms, and tools in the numerous applications of AI in education.

Today, the discussion about modern technologies such as AI in education, the value, and harm of AI in education is very crucial. In high population countries such as China and India, consider AI-based education is the only way possible for access to equal and quality

education. Therefore, implementing AI in education in such countries will grow with high speed.

Acknowledgments

References

1. Chris Dede, B., Ho, A., Mitros, P.: Big Data Analysis in Higher Education: Promises and Pitfalls. (2016).

2. Guilherme, A.: AI and education: the importance of teacher and student relations. AI Soc. 34, 47–54 (2019). https://doi.org/10.1007/s00146-017-0693-8.

3. Joshi, S., Rambola, R.K., Churi, P.: Evaluating Artificial Intelligence in Education for Next Generation. J. Phys. Conf. Ser. 1714, 012039 (2021). https://doi.org/10.1088/1742-6596/1714/1/012039.

4. Dignum, V.: The role and challenges of education for responsible AI. London Rev. Educ. 19, (2021). https://doi.org/10.14324/lre.19.1.01.

5. Octaberlina, L.R., Muslimin, A.I.: Efl students perspective towards online learning barriers and alternatives using moodle/google classroom during covid-19 pandemic. Int. J. High. Educ. 9, 1–9 (2020). https://doi.org/10.5430/ijhe.v9n6p1.

6. Davies, J.N., Verovko, M., Verovko, O., Solomakha, I.: Personalization of e-learning process using ai-powered chatbot integration. In: Advances in Intelligent Systems and Computing. pp. 209–216. Springer Science and Business Media Deutschland GmbH (2021). https://doi.org/10.1007/978-3-030-58124-4_20.

7. Chen, L.H.: Enhancement of student learning performance using personalized diagnosis and remedial learning system. Comput. Educ. 56, 289–299 (2011). https://doi.org/10.1016/j.compedu.2010.07.015.

8. Wartman, S.A., Combs, C.D.: Reimagining medical education in the age of AI. AMA J. Ethics. 21, 146–152 (2019). https://doi.org/10.1001/amajethics.2019.146.

9. Nesbit, J.C., Adesope, O.O., Liu, Q., Ma, W.: How effective are intelligent tutoring systems in computer science education? In: Proceedings - IEEE 14th International Conference on Advanced Learning Technologies, ICALT 2014. pp. 99–103. Institute of Electrical and Electronics Engineers Inc. (2014). https://doi.org/10.1109/ICALT.2014.38.

10. Raykov, Y.P., Ozer, E., Dasika, G., Boukouvalas, A., Little, M.A.: Predicting room occupancy with a single Passive infrared (PIR) sensor through behavior extraction. In: UbiComp 2016 - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp. 1016–1027. Association for Computing Machinery, Inc, New York, NY, USA (2016). https://doi.org/10.1145/2971648.2971746.

11. Conati, C.: Commentary on: "toward Computer-Based Support of MetaCognitive Skills: A Computational

Framework to Coach Self Explanation." Int. J. Artif. Intell. Educ. 26, 183–192 (2016). https://doi.org/10.1007/s40593-015-0074-8.

12. Mohamed, A., Alnaqbi, A.: The role of using AI techniques in enhancing e-learning in the military education process.

13. Salah Hashim, A., Akeel Awadh, W., Khalaf Hamoud, A.: Student Performance Prediction Model based on Supervised Machine Learning Algorithms. https://doi.org/10.1088/1757-899X/928/3/032019.

14. Siau, K.: Higher Education in the AI Age Supply Chain Management View project.

15. Koedinger, K.R., Aleven, V.: An Interview Reflection on "intelligent Tutoring Goes to School in the Big City." Int. J. Artif. Intell. Educ. 26, 13–24 (2016). https://doi.org/10.1007/s40593-015-0082-8.

16. Hao, K., Yi, Z.: China has started a grand experiment in AI education. It could reshape how the world learns.

17. Ahmed, J., Ahmed, M.: A web-based framework for classification and integration of news articles big data using ontology. Int. J. Recent Technol. Eng. 8, (2019). https://doi.org/10.35940/ijrte.C5409.098319.

18. Sijing, L., Lan, W.: Artificial intelligence education ethical problems and solutions. In: 13th International Conference on Computer Science and Education, ICCSE 2018. pp. 155–158. Institute of Electrical and Electronics Engineers Inc. (2018). https://doi.org/10.1109/ICCSE.2018.8468773.

19. Lee, H.S., Lee, J.: Applying Artificial Intelligence in Physical Education and Future Perspectives. Sustainability. 13, 351 (2021). https://doi.org/10.3390/su13010351.

20. Ahmed, M., Ahmed, J.: A Framework for Sentiment Analysis of Online News Articles . Int. J. Emerg. Technol. 11, 267–274 (2020).

21. Munezero, M., Montero, C.S., Mozgovoy, M., Sutinen, E.: Exploiting sentiment analysis to track emotions in students' learning diaries. In: ACM International Conference Proceeding Series. pp. 145–152. ACM Press, New York, New York, USA (2013). https://doi.org/10.1145/2526968.2526984.

22. Klomsri, T., Grebäck, L., Tedre, M.: Social media in everyday learning: How Facebook supports informal learning among young adults in South Africa. In: ACM International Conference Proceeding Series. pp. 135–144 (2013). https://doi.org/10.1145/2526968.2526983.

23. Liu, P., Gulla, J.A., Zhang, L.: RETRACTED ARTICLE: A joint model for analyzing topic and sentiment dynamics from large-scale online news (World Wide Web, (2018), 21, 4, (1117-1139), 10.1007/s11280-017-0474-9), (2019). https://doi.org/10.1007/s11280-018-0650-6.

24. Zhao, L., Wang, J., Huang, R., Cui, H., Qiu, X., Wang, X.: Sentiment contagion in complex networks. Phys. A Stat. Mech. its Appl. 394, 17–23 (2014). https://doi.org/10.1016/j.physa.2013.09.057.

25. Johnson, W.L., Lester, J.C.: Face-to-Face Interaction with Pedagogical Agents, Twenty Years Later. Int. J. Artif. Intell. Educ. 26, 25–36 (2016). https://doi.org/10.1007/s40593-015-0065-9.

26.   Eynon, R., Young, E.: Methodology, Legend, and Rhetoric: The Constructions of AI by Academia, Industry, and Policy Groups for Lifelong Learning. Sci. Technol. Hum. Values. 46, 166–191 (2021). https://doi.org/10.1177/0162243920906475.

27.   Selwyn, N.: Should robots replace teachers?: AI and the Future of Education. Polity Press (2019).

28.   Sampaio, B., Morgado, C., Barbosa, F.: Building collaborative quizzes. In: ACM International Conference Proceeding Series. pp. 153–159 (2013). https://doi.org/10.1145/2526968.2526985.

29.   Pedro, F., Subosa, M., Rivas, A., Valverde, P.: Challenges and Opportunities for Sustainable Development Education Sector United Nations Educational, Scientific and Cultural Organization. UNESCO (2019).

30.   Ahmed, M., Ahmed, J.: Big data and semantic web, challenges and opportunities a survey. Artic. Int. J. Eng. Technol. 7, 631–633 (2018). https://doi.org/10.14419/ijet.v7i4.5.21174.

31.   Howard, S.K., Yang, J., Ma, J., Ritz, C., Zhao, J., Wynne, K.: Using Data Mining and Machine Learning Approaches to Observe Technology-Enhanced Learning. In: Proceedings of 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2018. pp. 788–793. Institute of Electrical and Electronics Engineers Inc. (2019). https://doi.org/10.1109/TALE.2018.8615443.

32.   Math Learning Software | MATHiaU by Carnegie Learning, https://www.carnegielearning.com/products/software-platform/mathiau-learning-software/, last accessed 2021/02/08.

33.   Ahmed, J., Ahmed, M.: Ontological Based Approach of Integrating Big Data: Issues and Prospects. In: Lecture Notes in Electrical Engineering. pp. 365–378. Springer (2020). https://doi.org/10.1007/978-981-15-1420-3_38.

34.   Ahmed, J., Ahmed, M.: Semantic Web Approach of Integrating Big Data-A Review. (2018).

35.   Yan, F., Ruwase, O., He, Y., Chilimbi, T.: Performance modeling and scalability optimization of distributed deep learning systems. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1355–1364. Association for Computing Machinery (2015). https://doi.org/10.1145/2783258.2783270.

36.   Predicting Employment Through Machine Learning, https://www.naceweb.org/career-development/trends-and-predictions/predicting-employment-through-machine-learning/, last accessed 2021/02/08.

37.   Kexin, L., Yi, Q., Xiaoou, S., Yan, L.: Future Education Trend Learned from the Covid-19 Pandemic: Take ≪artificial Intelligence≫ Online Course As an Example. In: Proceedings - 2020 International Conference on Artificial Intelligence and Education, ICAIE 2020. pp. 108–111. Institute of Electrical and Electronics Engineers Inc. (2020). https://doi.org/10.1109/ICAIE50891.2020.00032.

38.   How to address new privacy issues raised by artificial intelligence and machine learning, https://www.brookings.edu/blog/techtank/2019/04/01/how-to-address-new-privacy-issues-raised-by-artificial-intelligence-and-machine-learning/,last accessed 2021/02/08.

A Dynamic Ontology Framework Design to Provide Interoperability in Agriculture IoT

P Salma Khatoon[1] and Muqeem Ahmed[2]

Department of Computer Science and Information Technology

Maulana Azad National Urdu University, Hyderabad

salmakhatoon537@gmail.com[1], muqeem.ahmed@gmail.com[2]

Abstract: *In the recent years, the usage and applications of Internet of Things (IoT) has increased exponentially. IoTnow connects multiple heterogeneous devices like sensors, micro controllers, actuators, smart devices like mobiles, watches etc. The IoT contributes significantly to the production of data in the context of Data collection, in domains such as healthcare, agriculture, military and many other fields. The diversity of possible applications at the intersection of the IoT and the web semantics has prompted many research teams to work at the interface between these two disciplines. This makes it possible to collect data and control various objects of transparent way. The challenge now lies in the use of this data. Ontologies address this challenge to meet specific data needs in the IoT filed. This paper presents the implementation of a dynamic agriculture ontology building tool that parses the ontology files to extract meaning full data and update it based on the user needs. The technology is used to create library in angular for parsing the OWL files. The proposed ontology framework would accept user defined ontologies and provide an interface for the online updating of the owl files, ensuring interoperability in the agriculture IoT.*

Keywords: Internet of Things (IoT), interoperability, dynamic agriculture ontology, OWL

1. Introduction

The development in the agricultural technology is leading to Feasibility of the vision "Agriculture 4.0" connected by an unforeseen potential for "robotization" of agriculture. Today's harvest associations can be much more efficient be used by being networked in the field and be coordinated collectively [1]. Also, can the fertilization and care be portioned exactly as required? The goal must be to help agriculture both greener and more productive shape. New processes and machines are achieved that can be flexibly adjusted and tailored for each application and geographical situation. The integration and use in agricultural activities of more intelligent and networking Systems (for example, self-driving and cooperating machining machinery, sensor masses, swarm robots, drones, farms, farm systems, ERR systems, etc) is only possible[2]. The combination of digital infrastructure, touch-screen, intelligent and easy-to-use devices, and digitalised processes

97

offers huge opportunities for productivity growth, reliability, and the creation of manufacturing transparency and a whole value chain.

In the future, completely new devices will also be in the Agricultural engineering possible that of today's technology mobile agricultural machinery with wheels and tractors differ greatly [3]. For example, instead of with Bicycles will also walk on legs in future Machines are used that are highly precise and carry out individual work, e.g., on plants and only weak point loads on the ground generate and only where none Useful plants stand. These can turn out to be difficult Slopes and high on the smallest fields Achieve efficiency. There will also be new opportunities profitability for air- and ground-based systems increase significantly, towards precision agriculture while maintaining the desired ecological boundary conditions.

The farmer or contractor will have the option of this novel semi-autonomous agricultural machinery to be able to control and monitor as an association, and no longer just each machine individually [4]. This should make it possible, even on small ones Fields to achieve maximum profitability. Depending on geographical requirement can be adapted ecological requirement of a small parcel be economically viable.

The number of networked devices is increasing by Year on year, whether in private in households or in companies. For the user should use the sensors and Smart devices collect data and Observe conditions as precisely as possible, so the environment can be described [5]. Huge amounts of data are produced from the sensors. Computer systems evaluate the Data and pass it on to the user ideally structured and sometimes semi-structured. Decisions based on the data obtained in this way automated via corresponding IT expressions or about people. The integration of the data in the real world allows the user an intuitive interface to the networked data sources in an IoT. There are numerous rich IoT platforms and specification soften the standardization of the aim at different interfaces instead of which, however, another domain represents specific "standard" [6]. Different domains (e.g., production, building technology) have different Requirements for the IoT systems and thus are also different protocols and communication standard are in hand.

At the overlap score, for example, in energy efficiency systems and automation occur, are the developments most of all because these different ones IoT systems need to work together. They focus accordingly Standardization organizations on these Areas [7]. The

differentiating between IoT systems in very extensive and large systems with many individual and sub-different sensors and devices and critical systems which often have an inposition and the applications stable communication and require case security. Through the current Dissemination of applications and ecotechnologies in the IoT environment will generate new handles such as the Internet of Nano-Things (IoNT) shaped[8]. Herewith is according to the integration of more and more, but also increasingly smaller sensors and case-specific sensors. Additional Lich is precisely for this technical development of an accessible ecosystem calls. Many standardization organizations are already working on different specifications to the IoT systems better to link and for external systems make systems more accessible. Likewise, the low-power communication in the Prioritize IoT networks. The current ones Technologies for communication inputs. In this an overview of current approaches and active in- application scenarios of IoT specifications.

A comprehensive interoperability between the standards is lacking. Through semantic Models can objects with a high Level of detail are described [9]. These detailed descriptions can however, they are already very useful for simple devices become complex. By increasing Use of sensors in all areas of life between the systems the amount of data that people use is increasing often can no longer be processed. This calls for a higher level of automation in which the People have to intervene less and less. In addition to automation, there is also the Ability to contextualise people with user-centered information and thus reduce the flood of data.

2. Interoperability

Interoperability is the ability to communicate effectively between two systems and is a key factor in the future development of IoT [10]. For the banking, medical, agriculture and other life service industries, we expect those platforms used for information exchange to communicate seamlessly when we need them. Each of us has thousands of pieces of data that are closely related to our lives, related to our health, financial situation and other important aspects of life, which also explains why the latest developments in machine learning and artificial intelligence (AI) can be used, so that we can benefit from data collaboration.

Generally, IoT systems or IoT sensor communication networks are designed to communicate and operate together using common technologies, such as Zigbee, Bluetooth, Z-wave, WiFi, etc. [11]. One of the current problems of the Internet of Things is to connect

new devices to the existing network because there are different communication protocols between the existing network and the new system.

Thus, for a scalable, flexible and seamless IoT network, the interoperability of IoT devices is becoming increasingly essential.

McKinsey's claims that IoT interoperability is very necessary to achieve the full value[12]. Interoperability between IoT systems on average, 40 percent of the value needs multiple IoT systems to collaborate and, in some contexts, almost 60 percent. Interoperability must be based on the future benefit of all.

The IoT ecosystem needs interoperability in order for different products or devices, or sensors to communicate and interact seamlessly to establish programmability or configurability [13]. Regardless of the manufacturer, model or industry, a single universal standard is required to enable devices to be communicated, operated and programmed.

The interoperability of equipment mainly includes the following aspects:

2.1 Technical interoperability:

Components, devices, and platforms of hardware/software supporting communications between machines. Mainly the contract and the facilities necessary for the agreement to be operated.

2.2 Semantic interoperability:

The ability to exchange information between two or more systems or components and to use the information that has been exchanged.

2.3 Syntax interoperability:

The details conveyed by the Communication Protocol must have a simple syntax and data format(such as XML, JSON, etc.).

The emerging issue of interoperability has many aspects

1. Various types of protocols/devices link
2. Various industry principles of interoperability like AllJoyn, OneM2M, OPC-UA, etc.
3. Multiple organizations have created interoperability for the Internet of Things In addition to abstracting low-level device protocols from end users, the service platform also needs to communicate with these interoperable IoT service platforms.

There are also, however, several organizations that build interoperability solutions through the adoption of standards and open-source technology to improve internet standardization. The current/future necessity is a successful portal integration service (middleware). It converts various low-level data protocols into interoperable IoT data protocols, which are seamlessly used by the IoT platform layer and interact with other Operating protocols work together [14]. The current systems and principles of interoperability are nevertheless highly fragmented. This seems to be one of the reasons why products do not use interoperability service systems and in many cases continue to use existing low-level technologies.

When designing the IoT ecosystem, people need to consider interoperability issues and at the same time define the IoT architecture to expand products and connect more different types of things/devices [15]. Interoperability is a strong challenge for the industry and needs more active thinking, in order that IoT network device/system solutions can be unified, reliable, stable, flexible and adaptable. Seamless interoperability of IoT devices is expected to speed up and unleash massive opportunities on the IoT market, speed up innovation within industry and enable developers and businesses to rapidly build solutions.

3. Proposed method

The term ontology comes from ancient Greek and means something like "the doctrine of being". It was originally a branch of theoretical philosophy or metaphysics and dealt with questions about the nature of being and the structure of reality. This focus shifted over time: In the more recent interpretations of philosophy, the term ontology primarily focuses on language itself. This means that there is a linguistic division into areas of knowledge and areas of life. In addition, the complex of ontology and language includes many regularities. The term is also used in computer science from this almost linguistic perspective.

Here the ontology creation is used to formalize knowledge and its relevance is steadily increasing. Because due to the intensive use of the Internet and the massive automation of various processes, it is more important than ever that different systems can communicate with one another. With the help of ontological systematizations, relevant information can be read and used by both human users and various machine systems (knowledge engineering).

When creating ontologies, not only terms and terminologies per se, but also their relationships to one another and, if necessary, derivation rules between them are recorded. In this respect, ontologies go far beyond pure linguistics in their application. They serve to facilitate communication between human and machine operators. In IT, therefore, individual ontologies are defined and used for each application or area.

3.1 Ontology and Its Components

    A. Ontologies in computer science are characterized by three factors:

    B. A formal structure so that machines can also read and use them.

    C. The knowledge they describe must be clearly identifiable.

    D. In addition, ontologies describe not only terms of a defined area, but also abstract concepts ("Specification of a Shared Conceptualization"). These concepts are related to each other.

3.2 Building an Ontology

To describe the structure of an ontology, one has to consider the concepts and relationships of ontologies more closely. There are two types of concepts. Those that describe a whole class or a specific set of individual objects. And concepts that only describe specific terms.

The relations for their part are divided into two parts: There are relations that arrange concepts hierarchically or which put concepts into any relationship with one another.

3.3 Ontologies on the web

The semantic search is one of the main applications in the Semantic Web - especially for confusing or unknown knowledge. Ontologies can be useful here in several places. On the search query side, the input can be generalized or specialized, and adjusted or corrected with regard to the content if the information sought was based on ontological structural knowledge. If documents are enriched with further descriptive (and ontologically structured) meta-data, complex contents can also be displayed in the search. Sorting and presentation rules formulated by ontologies can also be applied to the search result so that information can be optimally mapped and integrated.

This leads to the next common area of application of these classification systems, intelligent information integration. Ontologies can also help to describe not only content but also schemes. This is the case, for example, with the integration of data that have different origins (sources). Ontologies function as transformation and translation rules.

3.4 Resource Description Framework (RDF)

An RDF/XML file is considered in an OWL-RDF which constructs the OWL ontology with the format of JSON in relevant to the represented triples in the RDF. In such type of a parser, a strategy could use which describes in this paper. Specifically, we intend to provide particular data about how one can establish a parser based on the majority of OWL ontologies.

The discussion is included the OWL Lite, and OWL DL and the procedure is mentioned below by considering the parsing of OWL DL ontologies. The object and subject of the triple with relevant types both are either owl: DatatypeProperty or owl:ObjectProperty) is required in OWL DL when using rdfs:subPropertyOf. If not, an error will raise by the parser. This would allow in an OWL full parser, but it's unnecessarily clear what would be the related abstract syntax for such kind of construct.

OWL ontologies with a characterization in an abstract syntax is provided in the document of the OWL Semantics and Abstract Syntax. This is a way of description highly where the properties and classes' characteristics can define.

Additionally, a mapping to RDF triples is provided by S&AS that mentions about how the OWL ontology with abstract description can transform into the RDF triples collection. It can also represent in a concrete fashion based on RDF/XML.

This mapping is done reverse in order to parse an OWL-RDF into a certain structure near the abstract syntax. That means, what were the property and class definitions results in those specific triples determine. It should be noted that the reverse mapping is not necessarily special.

Class (a)Class (b)

SubClassOf (b a)and

Class (a)

Class (b partial a)

Both leads to the same collection of triples under the mapping:

103

a rdf:typeowl:Classb rdf:typeowl:Classb rdfs:subClassOf a

This is not an issue necessarily for different purposes, like validation of species. We expect that the parser would be a consistent in the strategy in other cases, in which editing tool has used as it produces the descriptions of abstract syntax. An OWL Lite or DL ontology may not correspond to an arbitrary RDF graph necessarily. Alternatively, it can describe as an OWL Lite or DL ontology may not be there to transform or produce the given graph through the mapping. A species validator tries to estimate whether an ontology exists. Then, a parse will attempt to build such type of an ontology.

The corresponding of an OWL ontology to an RDF graph may fail in two ways such as:

4. *In the format of abstract syntax, an OWL Lite or DL ontology is there that allows the mapping of a superset of the triples some of which are not available in the graph and have forgotten.*

5. *In abstract syntax form, the ontologies are there that map the superset of triples or triples. For membership of the OWL DL or Lite subspecies, some restrictions are violated. This would be the situation where no availability of such type of ontologies.*

3.5 Parser Implementation

Most of the XML parsers follow a streaming fashion where the elements report to the parser as file processing incrementally and encountered during the parsing. While parsing RDF models, this process is difficult to perform or at least processing a task like generating an OWL ontology with an abstract syntax from an XML or RDF. The order is not ensured where the triples process in the graph and this is the problem. Thus, the streaming sparser is reported. In the RDF file, a specific construct of syntax may divide across various locations. We need to observe that the encountered triples earlier and return to the process them later for parsing in a streaming way. A situation is considered where an owl:AnnotationPRoperty uses for making an annotation regarding the specific individual:

AnnotationProperty (hasName) Individual (fredhasName "Frederick") This

leads to the triples:

[1] hasNamerdf:typeowl:AnnotationProperty

[2] fredhasName "Frederick"

We have known that the property is an annotation when encountering before during the parsing and process as an annotation. Here, we don't know whether it's required to

```
<Declaration>
<Class IRI="#Temp"/>
</Declaration>
<Declaration>
<Class IRI="#Temper"/>
</Declaration>
<Declaration>
<Class IRI="#Temperature"/>
</Declaration>
<SubClassOf>
<Class IRI="#Temp"/>
<Class IRI="#Temperature"/>
</SubClassOf>
<SubClassOf>
<Class IRI="#Temper"/>
<Class IRI="#Temperature"/>
</SubClassOf>
```

process like an individual value or an annotation when encounter. Until we view all the triples, whether it will happen or is not known and we should have to wait till all triples have viewed before processing.Until all triples are available, the parser doesn't process anything in this strategy. When we collect the triples and process them, the parser conceptual complexity is reduced even though it's possible to process the data in a streaming way. The ramifications have included on the resources that require while parsing. Large memory amounts may need if large RDF graphs are parsing.In the created ontology, we can get access to the objects is assumed while parsing. For example, we have access to it if an ObjectProperty p. The ObjectProperty has defined with p in reference to the *ObjectProperty p*

4. Experimental Results

This section presents the experimental results carried out to validate the proposed framework. The input owl file for the experimentation consists of sensors and their alternate words. The example of the Temperature sensor is shown below

The sample OWL data provided above contains the

```
{
        "Temperature": {
                "alternate":        ["Temp",
"Temper"],
                "type": "Sensor"
        }
}
```

```
<Declaration>

<Class IRI="#Temp"/>

</Declaration>

<Declaration>

<Class IRI="#Temper"/>

</Declaration>

<Declaration>

<Class IRI="#Temperature"/>

</Declaration>

<Declaration>

<Class IRI="#t"/>

</Declaration>

<Declaration>

<Class IRI="#T"/>

</Declaration>

<SubClassOf>

<Class IRI="#Temp"/>

<Class IRI="#Temperature"/>

</SubClassOf>
```

The JSON and OWL files can both be updated using the proposed framework. New alternatives can be added to both the files as shown below:

```
{
        "Temperature": {
                "alternate":        ["Temp",
"Temper", "t", "T"],
                "type": "Sensor"
        }}
```

The above OWL data shows the addition of two new alternatives namely *t* and *T*. The same change is reflected in the JSON file*Declaration* tags and *SubClassOf* tags. The declaration tags describe the variables used in the ontology, *Temperature, Temper* and *Temp.* The relation between these variables is provided by the *SubClassOf* tags. The proposed framework extracts the information and constructs a JSON file. The resultant JSON data is shown below:

This would allow the users to define their own alternatives to the sensors and obtain accurate results. The use of the proposed framework would read the data from different frameworks. The result is show in figure 1.

Fig .1. Data parsing using the proposed ontology

The figure 1 shows the result of using the proposed ontology on the sensor data. On the left side, the sensor readings are accurate and complete with the ontology. But on the right side only pH sensor readings were extracted.

6. Conclusion

When processing the search, ontological derivation knowledge as well as ontological definitions serves to bridge possible inconsistencies in the formulation of the search and the available information. Furthermore, a similarity-based search is made possible by using background knowledge. This paper presented an ontology framework for the parsing an OWL file and converting it into JSON data which can be easily parsed by many platforms. The framework also provided means of updating the OWL file and the JSON file at the same time. The proposed framework would help reading the sensor data files accurately without missing any information

References:

[1]. Fountas, S., Espejo-García, B., Kasimati, A., Mylonas, N., & Darra, N. (2020). The future of digital agriculture: technologies and opportunities. *IT Professional*, *22*(1), 24-28.

[2]. Fountas, S., Espejo-García, B., Kasimati, A., Mylonas, N., & Darra, N. (2020). The future of digital agriculture: technologies and opportunities. *IT Professional*, *22*(1), 24-28.

[3]. Oksanen, T., Linkolehto, R., & Seilonen, I. (2016). Adapting an industrial automation protocol to remote monitoring of mobile agricultural machinery: a combine harvester with IoT. *IFAC-PapersOnLine*, *49*(16), 127-131.

[4]. Zhang, X., Geimer, M., Noack, P. O., & Grandl, L. (2010). A semi-autonomous tractor in an intelligent master–slave vehicle system. *Intelligent Service Robotics*, *3*(4), 263-269.

[5]. Fortino, G., Russo, W., Savaglio, C., Shen, W., & Zhou, M. (2017). Agent-oriented cooperative smart objects: From IoT system design to implementation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *48*(11), 1939-1956.

[6]. Stočes, M., Vaněk, J., Masner, J., & Pavlík, J. (2016). Internet of things (iot) in agriculture- selected aspects. *Agris on-line Papers in Economics and Informatics*, *8*(665-2016- 45107), 83-88.

[7]. Marcu, I., Voicu, C., Drăgulinescu, A. M. C., Fratu, O., Suciu, G., Balaceanu, C., & Andronache, M. M. (2019, March). Overview of IoT basic platforms for precision agriculture. In *International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures* (pp. 124-137). Springer, Cham.

[8]. Yang, K., Bi, D., Deng, Y., Zhang, R., Rahman, M., Ali, N. A., ... & Alomainy, A. (2019). A comprehensive survey on hybrid communication for internet of nano-things in context of body-centric communications. *arXiv preprint arXiv:1912.09424*.

[9]. Schlieder, C., Vögele, T., & Visser, U. (2001, September). Qualitative spatial representation for information retrieval by gazetteers. In *International Conference on Spatial Information Theory* (pp. 336-351). Springer, Berlin, Heidelberg.

[10]. Noura, M., Atiquzzaman, M., & Gaedke, M. (2019). Interoperability in internet of things: Taxonomies and open challenges. *Mobile Networks and Applications*, *24*(3), 796-809.

[11]. Bröring, A., Schmid, S., Schindhelm, C. K.,

Khelil, A., Käbisch, S., Kramer, D., ... & Teniente, E. (2017). Enabling IoT ecosystems through platform interoperability. *IEEEsoftware*, *34*(1), 54-61.

[12].Deichmann, J., Roggendorf, M., & Wee, D.(2015). Preparing IT systems and organizationsfor the Internet of Things.

[13].Bröring, A., Schmid, S., Schindhelm, C. K.,

Khelil, A., Käbisch, S., Kramer, D., ... & Teniente, E. (2017). Enabling IoT ecosystems through platform interoperability. *IEEEsoftware*, *34*(1), 54-61.

[14].Al-Masri, E., Kalyanam, K. R., Batts, J., Kim, J., Singh, S., Vo, T., & Yan, C. (2020).

Investigating messaging protocols for the Internet of Things (IoT). *IEEE Access*, *8*, 94880-94911.

News Articles Recommendation System: A Review

Waseem Akram Khan[1], Amandeep Kaur[2]

[1]Department of CS & Tech., Central University of Punjab, Bathinda, Punjab, India

[2]School of Engineering and Technology, Central University of Punjab, Bathinda, Punjab, India

waseemakhan4@gmail.com, amandeep.kaur@cup.edu.in

Abstract: *The recommendation system have become the way of living in our day to day life in a numerous ways. The recommendation system are currently in use in almost every areas such as e-commerce, movies, tourism, online restaurant preference, news articles. These systems have also proved to save the browsing time yet excelled in providing the preferences matching the interest of the user. The main objective of this paper is to analyze the problem of information overload and News Articles Recommendation approaches. This paper presented different challenges and work of various researcher related to the recommendation system. This paper is divided into five sections such as first section covers introduction of recommendation system, second section covers challenges, third section describes recommendation techniques, forth section covers work related to recommendation and finally, covers the conclusion of the paper.*

Keywords: *Information overload, recommendation system, recency, personalized, non-personalized, content-based, collaborative-based*

6. Introduction

Todays, usage of the internet and technologies has become more popular as compared to earlier days due to which every day, even every seconds, people are browsing, surfing, searching for the products or items over the internet depending on their needs. On an average, this information i.e., products or items are consistently increasing with the rapid growth of the internet and this led to overwhelming of information, in the virtue of this it is very problematic for the internet users who want to get preferable products or items, this problem lead to information overload also known as information overflow (Bradley, 1995). It is happening due to continuous storage of information over the internet that is generating from numerous sources, a pool of database, ambiguities in available information (Hoq, 2016) and the consequences will distract the concentration from appropriate information, consume lots of time to retrieve the relevant information, accelerate the stress and anxiety among the users (Renjith, 2017). In the early of 1990s, the first recommendation system was developed by

(David Goldberg, 1992) which was based on collaborative filtering technique and there is an electronic mailing tool called Tapestry that is used for information retrieval from massive wave of incoming message or mail from the users. Recommendation system is a very interesting and widely researched topic since nineties. Recommendation system is a way of decision-making approach (Ricci, 2011). It is a technique to handle or address the problem of information overloaded and suggesting the items to user(s) based on their searched history & preferences as well as which items are most trending & popular nowadays, for example what kind of items did he bought, what kind of movies did he watched, what kind of online news read, etc. (Borges, 2009).A traditional recommender system consist of following terminologies:

- *Users***:** Those people who are in the system for the taste of items or products as per need is called the users (Ricci, 2011). Each users may have a set of features called demographics such age, gender, email-id, name, etc. For each user model can be conclude. For example, their sort of TV serial, the type news they like to read.

- *Items***:** Those products in the system which are selecting to recommend to the users side, called items. Every items have its own feature attributes or properties for example, an actress in a TV serial, author of news articles, etc. (Ricci, 2011)**.**

- *Preferences***:** It refers to user's likes or dislikes and how many times a particular user visit the Web page (in case of news articles)**.**

- *Transactions***:** It refers to repository of both items and user behaviours and it collect the past history of user (Ricci, 2011)**.** For example items chosen by users, rating or feedback given by users, etc.

7. Challenges in News Recommendation System

- *Cold-start problem*: It means system can't recommend any items to users if the user is new because system don't have any past behaviour of the current user so, it is also known as first rater or early rater problem.

- *Gray sheep problem*: The recommender system will also work within a group of members whose preference or interest are matched to others, if there is a user who is not a part of group and preference will also not matched then this problem will occur.

- *Data Sparsity problem*: It uses sparse matrix (between users and items) to visualize this issue, when users rated limited number of news items then this problem would be occur.

- *Scalability problem*: Recommender system should have ability to handle large amount of users along with news articles and it should be fast and more processing time.

- *Recency problem*: The term recency means "Freshness" or "Fresh" or "Recent". The recency problem has adversely affected this online newsreader who are in thrust of fresh news instead of old news over a website.

- *Implicit-User feedback*: User's feedback plays an important role to recommend precise and correct news articles to the users. Suppose that, if user will not give feedback then is very tedious to understand for system, is user liked or not?

- *Cross-lingual problem*: Generally, news articles are published in a specific language, it should be recommend in bi-lingual or tri-lingual. This is one of the most important problem in recommender system and need to solve it.

8. Recommendation Techniques

Figure 1 describes different recommendation techniques which includes personalized and non-personalized recommender system. Personalized recommendation system further divided into four parts such as Collaborative-based, Content-based, Knowledge-based and Demographic-based. And collaborative-based divided into two parts such as Memory-based and Model-based.
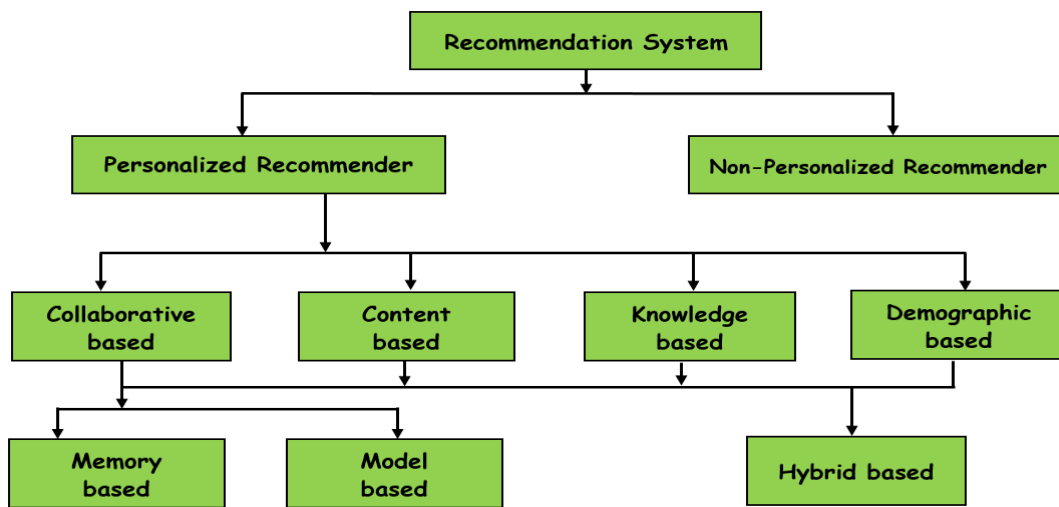
Figure 1: Different techniques of Recommendation system

3.1 Personalized Recommender System (PRS)

Personalized recommender system represents the relationship between users and items, which utilized the user's past history or profile and features attribute of the items. Recommended items are very from user to user therefore, this type of recommendation system follow one-to-one relationship strategy i.e., individual preferences (S.Ephina Thendral, 2018). General Workflow of personalized recommendation will be first step in system is to obtain the information of user through registration process, second step is system interact with user based on input information then after information pre-processing will take place after this step it form the recommendation to recommend something to user as per transactional activity (Chawan, 2015). For example Facebook friend suggestion based on profile matching such as same university, same city, phone directory, etc.

3.2 Non-Personalized Recommender System (NPRS)

It is a simple recommender system. In this recommender system information of user is not mandatory and system does not require transactional activities to recommend the item or product between the users. Non-personalized recommender system recommend items based on products or articles which are most popular and most trending therefore, recommendation process are same for everyone (Prasad, 2012). This recommender system is known as one-to-many relationship type functionality it means one item multiple user. For example, suppose that an anonymous user visit any news articles website, before that multiple of user visited and like a particular article then IT recommend that article to the anonymous user (Chawan, 2015). Netflix.com, flipkart.com, news.yahoo.com type of websites are treated as Non-personalized recommender system.

3.3 Collaborative Filtering (CF)

The word collaborating filtering is widely used and earliest proposed technique in recommender system. Collaborative filtering technique is used for collecting and analysing a huge amount of information from user's data repository. What is the user's behaviour, past history, preferences, etc. and forecast items being liked by the users based on the similarity and recommend to other user (Atisha Sachan, 2013). The first recommendation system was developed by (David Goldberg, 1992) which was based on collaborative filtering technique, i.e., is an electronic mailing tool called Tapestry which is used for information retrieval from massive wave of incoming messages or mails from users. Illustration of collaborative filtering technique as follow:

Statement1: If user U1 likes the item A, item B, item C.

Statement2: If user U2 likes item A, item C, item D.

Statement3: If user U3 likes only item A.

So, there is high probability that user U3 may like item C because, first two users (U1 & U2) likes item C which is new for user U3. Hence, Collaborative filtering help user make choice based on the mind of other users as shown in figure 3. Collaborative filtering technique is classified into two more categories such as Memory-based CF and Model-based CF.

3.3.1 Memory Based CF

In Memory based CF is used to compute the user-to-user and item-to-item correlation with the help various distance metrics, such as Pearson correlation coefficient, Cosine distance, Euclidean distance, Manhattan distance, etc. In this method (Atisha Sachan,

2013), we just compute the similarity between the user, which is based on the previously rated item given by users so, it is called user-based, neighbour-based collaborative filtering technique. The main disadvantage of this method is sparseness of data. For example, Suppose that User 1 and 2 having similar past behaviour, if user 1 likes item A then user-based CF can recommend item A to user 3 as shown in figure 2
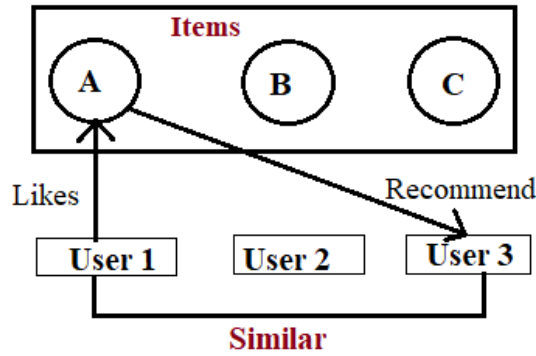

Figure 2: *User-based Collaborative Filtering*

3.3.2 Model Based CF

This method does not require user's past information (such as rating, explicit feedback, etc.) to recommend the items. The basic principle of this method is to compute the similarities between items, based on the similar features of items and rated items that is called item-based CF technique as shown in figure 3. For example, a news website having four articles as A1, A2, A3 and A4. Suppose that articles A1 and A3 are closed similar if users likes articles A1 then item-based CF can recommend article A3 to user.
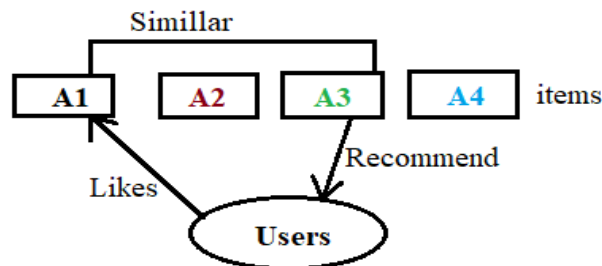
Figure 3: *Item-based Collaborative Filtering*

3.4 Content Based Filtering (CB)

This filtering technique is based on the detail/contents or some information provided for that product. The basic principle of this method lies in the mining attributes features of

recommendation items and the construction of interest based on content features and users past behaviours (Rachana R. Jadhav, 2016). General Workflow of content based recommendation is classify into three steps. Firstly, text representation of recommendation of items, it means extract the specific features from each items based on provided details. After it, we construct the model using selected features of items set and user past preferences (i.e., likes or dislikes). Finally, system prepare a set of most relevant items to recommend to the user based on the interest and their profile matching as shown in the figure 4
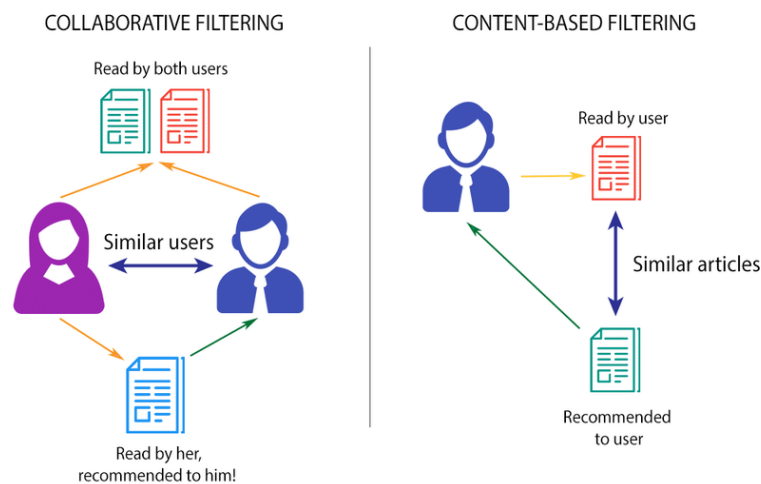
Figure 3: *Collaborative and Content-based Filtering*

3.5 Knowledge Based Filtering (KB)

The Knowledge based filtering technique is used to recommend the items based on the inferences about specific user's requirement and preferences. This technique is mainly used where both content based and collaborative based recommendation does not applicable. It tells all about how a specific type of item meets an individual user need and it uses the explicit description about item (Burke, 2002). With the help item's feature recommendation will perform among the users.

3.6 Demographic Based Filtering (DB)

The main objective of this technique is to classify the users based on personal features like age, gender, name, languages, location etc. and system attempts to recommendation using these demographic classes. This recommendation system isn't fall on the people item rating and it gives recommendations before the user appraised any item. Demographic techniques form 'user-to-user' correlations like collaborative ones, but use different data (Burke, 2002).

Main advantage of a demographic technique, it may not require a past history of user's ratings of the type needed by "collaborative and content-based" techniques.

3.7 Hybrid Based Filtering (HB)

It is a combination of two or more type of filtering techniques such as collaborative, content, knowledge and demographic. It uses all type of information and properties from both user-item interaction and user-item specification. With the help of this technique, we can easily overcome the limitation of other filtering techniques and improve the accuracy of recommendation system.

4. Literature Review

4.1 Social Media based Recommendation system

Facebook is one of the most widely used free social networking site and it is based on personalized approach. (Graham, 2012) Nowadays, over 2.7 billion of users are actively using Facebook, where we can easily communicate and share right information or thoughts to other users and playing games, making friends based on their profiles. Facebook provide us live streaming videos, recoded videos watching facility, reading news articles or blogging services, etc. (Baatarjav, 2008) Proposed "Group based recommendation system for Facebook" to overcome the problem of choosing right group to join for future entertaining with different minded people with the help of hierarchical clustering and decision tree algorithms. The accuracy was outperform after used of decision tree classification technique. (Zhai, 2009) proposed novel news recommendation system for social network community of "university of Illinois" on regular basis. In this system, content-based filtering was used for fetching the content of Facebook news articles and collaborative filtering was used for storing explicit and implicit feedback of each users. The system was computed for utility, accuracy and scope of application on a set Facebook community. Finally, the system flexible & feasible for larger communities and users found that application useful and coherent for social networking web portal.

Twitter is a personalized based free social networking website where millions of users share the information or thoughts or communicates via tweets to other users (Maclean, 2013). Twitter provide us micro-blogging services using simple messaging technique. (Nidhi, 2017) proposed the content-based twitter recommendation system with the help two algorithm called Noun-Phrase detector and Naïve Bayes filtering approach to retrieve the contents of tweets of the

users and then looked for other users who retweet the same contents as per recommendation. (Gauch, 2013) was proposed popularity (it uses ranking of articles) and relevance based news recommender system called Hybrid based. The system uses cosine similarity method to compute or check the interconnectivity of both popularity and user's profile interest. The accuracy of the system was better in hybrid case not an individual case. The evaluated result were done by 280 news articles datasets collected from BCC and CNN and 202,224 tweets datasets, retrieved from twitter platform with seven different topics like sports, crime, business, technology, politics, health and entertainment. (Phelan, 2009) proposed real-time news recommendation system using Buzzer system and system recommend topical news like user's favourite RSS (Really Simple Syndication) and twitter feeds. The Buzzer system uses content-based approach, firstly web based registration approach will record the user preferences from twitter account/timeline after that lucence approach is used to mining the content of harnessing twitter and RSS(articles) information. Finally, recommendation engine arrange the news article based on ranked or popularity with the help of TF-IDF algorithm. Buzzer has ability to act as collaborative filtering news recommender system like new RSS feeds based on the relevant people to follow twitter.

4.2 Mobile based Recommendation Systems

(Lee, 2007) the author proposed a mobile news recommendation system and the main objective was to assimilate the features of users as well as the nature of the mobile web news services and contents. who access the MONERS, firstly, system confirm that the user is a new or existing user. If the user is new then he/she temporarily placed in similar sections on the basis of demographics after it recommendation will perform for those users otherwise the recommendation for an existing user will perform based on the preferences and user can easily retrieve the news articles. The calculation of recency is based on the weight of the articles it means to compute the distance between the current time and the time an article was posted. (Athalye, 2013) proposed an android application called newsreader for those users read news online. This application mainly focuses on two things i.e., previously recorded reading habit of users and most trending news articles. Finally, application will recommend the news articles based on maximum number of user click rate and user's preferences. (A Wahana, 2019) proposed an item-based recommendation system with the purpose to recommend the latest news to interested online news reader. In this paper, the collaborative filtering was used to rank the popular news articles based on the similarity measurement. Finally, it proven that collaborative filtering has a capacity to gives good result with better

117

accuracy. (Dharma, 2019) the author proposed a novel approach to recommend the news articles using KNN algorithm. In the paper, to study the user interest based on IP address, access time, User agent, status code, page size, HTTP request method, etc. and user similarity. With the help of these user's data a personalized recommendation will work and to recommend the news articles to target users.

4.3 Recency-based Recommendation system

The term recency means "Freshness" or "Fresh" or "Recent". The recency problem has adversely affected this online newsreader who are in thrust of fresh news instead of old news over a website. It save a lots of user's time and get the recent what happening around us on time.

(Lak, 2016) proposed a methodology to analysis of news article's contents using various text analysis tools such as keyword popularity, tf-idf, word2vec. The main aim of this work is to find out correct position (which one is older or fresh) of news articles content because a lots of news articles are being published every day. Finally, analysis shows that the recent articles plays an important role among the news reader. (Chakraborty, 2019) the authors focusing to improve or optimize the performance of recency issue for non-personalized recommendation system. Author's proposed strategy is "Highest-Future-Impact" to recommend the fresh & relevant news to the users. Authors explored two methods for the prediction of lifetime-impact of news articles called ordinary least squares(OLS) and Gradient tree boosting(GTB). (Ruey-Cheng Chen, 2019)proposed recency bias issue in job recommendation system with the help of unbiased learning to rank approach and inverse propensity weighted loss function. Evaluation of recommendation engine was based on Normalized Discounted Cumulative Gain(NDCG) and Hit rate. (Moreira, 2019) proposed a Meta-architecture called CHAMELEON based on deep learning for personalized session-based news recommendation. Architecture deal with both news articles information and user behaviour information for news recommendation to a particular user. The main goal is to tackle the problem of cold-start and recency in news recommendation system. (Blaž Fortuna, 2010) proposed a support vector machine (SVM) based news recommendation system and to solve the issue of freshness of news articles over relevant articles. Due to this issue, users feel irritated and may lose their interest to read in continuing with the news fed by the system. (Wen, 2012) another approach called Hybrid recommender system for news articles recommendation on the Website to handle the age limit of news articles i.e., incorporates the

time factor during recommendation process and news articles would recommend based on user's interest or preference. (Liang, 2012) proposed Time-aware personalized topic recommendation system for micro-blogs articles. According to "Twitter" dataset, the result of hashtag recommendation performed better in many aspect like most popular, most recent, content recency, etc. (Li, 2011) designed a novel framework for personalized recommendation system called SCENE which included two-stages. The working principle of first stage is to make cluster of news articles and second stage is to recommend news articles to users based on exclusive characteristic like popularity, recency, etc. (Wanrong Gu and Dong, 2014) proposed a NEMAH system architecture for personalized recommendation system, which consist of four modules such as classification & clustering, Subclass Popularity Prediction, User profile module, recommendation module and each module has its own working principle. The process of recommend news articles based on user profile by subclass popularity and recency. The result of NEMAH was better as compared to other proposed framework.

5. Conclusion and Future work

In this review paper, we have discussed numerous of research work which have already done in the context of challenges and applications with respect to News domain by multiple researchers. Most of the News Articles recommendation system has been build using similarity measurements such as Cosine similarity, Jaccard similarity, Euclidean similarity, etc., and Machine learning Algorithms. Addressed system, recommend the News articles based on User's rating, Article's popularity, recent articles to the end users. In News recommendation system, news articles reader are facing various problems such as cross-lingual, cold-start, recency over the web portal so, developer should understand and try to find out the solution behind the addressed problems. If the system succeeds in overcoming with the above problems may prove to be a successful, robust tool and popular news website among the newsreaders. In future, to analyze News articles recommendation system using NLP model or vector space model & Deep learning model and compare the results of both.

References

A Wahana, D. S. (2019). News recommendation system using collaborative. *4th Annual Applied Science and Engineering Conference*.

Athalye, S. (2013). Recommendation System for News Reader. *Master's Projects 294*.

Atisha Sachan, V. R. (2013, 04). A Survey on Recommender Systems based on Collaborative Filtering Technique. *International Journal of innovations in Engineering and Technology(IJIET), 02*(02), 08-14.

Baatarjav, E.-A. a. (2008, 11). Group Recommendation System for Facebook. *5333*(0), 211-219. doi:10.1007/978-3-540-88875-8_41

Blaž Fortuna, C. F. (2010). Real-Time News Recommender System. *Springer Berlin Heidelberg, Machine Learning and Knowledge Discovery in Databases*.

Borges, H. &. (2009). A Survey on Recommender Systems on News Data. *Studeis in Computational Intelligence, 260*(2010), 129-151. doi:10.1007/978-3-662-04584-4_6

Bradley, M. N. (1995). GroupLens: An Open Architecture for Collaborative Filtering. *Commun. ACM, 40*(March 1995), 1-18.

Burke, R. (2002, 11). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction, 12*, 331-370. doi:10.1023/A:1021240730564

Chakraborty, A. a. (2019). Optimizing the recency-relevance-diversity trade-offs in non-personalized news recommendations. *Information Retrieval Journal, 22*, 447–475.

Chawan, P. a. (2015). Recommender System for volunteers in connection with NGO. *International Journal of Computer Science and Information Technologies, 07*(0), 11-15.

David Goldberg, D. N. (1992, 12). Using collaborative filtering to weave an information Tapestry. *Communications of the ACM, 35*, 61-70. Retrieved from http://portal.acm.org/citation.cfm?doid=138859.138867

Dharma, A. S. (2019). The User Personalization with KNN for Recommender System. *Journal Publications & Informatics Engineering Research, 3*.

Gauch, N. J. (2013). Personalized News Recommendation Using Twitter. *ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 3*(0), 21-25. doi:10.1109/WI-IAT.2013.144

Graham, R. E. (2012). A Review of Facebook Research in the Social Sciences. *Perspectives on Psychological Science, 7*(0), 203-220.

Hoq, K. M. (2016, 02). INFORMATION OVERLOAD: CAUSES, CONSEQUENCES AND REMEDIES: A STUDY. *Philosophy and Progress, 55*(2016), 49-68. doi:10.3329/pp.v55i1-2.26390

Lak, P. a. (2016). News Article Position Recommendation Based on The Analysis of Article's Content - Time Matters .

Lee, H. J. (2007, 01). MONERS: A news recommender for the mobile web. *Expert System Aplication, 32*(01), 143-150. doi:10.1016/j.eswa.2005.11.010

Li, L. a. (2011). SCENE: a scalable two-stage personalized news recommendation system. *ACM*, 125-134.

Liang, H. a. (2012). Time-aware topic recommendation based on micro-blogs. *ACM international conference on Information and knowledge*, 1657-1661.

Maclean, F. a.-L. (2013, 06). Understanding Twitter. *British Journal of Occupational Therapy, 76*(0), 295-298. doi:10.4276/030802213X13706169933021

Martin, F. a. (2011, 9). The Big Promise of Recommender Systems. *AI Magazine, 32*(0), 19-27.

Moreira, G. a. (2019). Contextual Hybrid Session-Based News Recommendation With Recurrent Neural Networks. *IEEE Access, 11*, 1-1.

Nidhi, R. a. (2017, 06). Twitter-user recommender system using tweets: A content-based approach. *International Conference on Computational Intelligence in Data Science(ICCIDS), 0*(0), 1-6. doi:10.1109/ICCIDS.2017.8272631

Phelan, O. a. (2009). Using twitter to recommend real-time topical news. *Proceedings of the third ACM conference on Recommender systems, 0*(0), 385--388. Retrieved from http://scholar.google.de/scholar.bib?q=info:Z2jn7-UDGDUJ:scholar.google.com/&output=citation&hl=de&as_sdt=0&ct=citation&cd=24

Prasad, R. (2012, 09). A Categorical Review of Recommender Systems. *International Journal of Distributed and Parallel systems, 3*(0), 73-83. doi:10.5121/ijdps.2012.3507

Rachana R. Jadhav, M. N. (2016, 12). A Survey on Recommender System. *International Journal for Research in Engineering Application & Management (IJREAM), 02*(09), 1-5.

Renjith, R. (2017). The effect of information overload in digital media news content. *Communication and Media Studies, Vol. 6 No. 1*(March 2015), 73- 85.

Ricci, F. a. (2011). *Roccomender Systems Handbook* (2nd ed.). (F. a. Ricci, Ed.) New York : Springer.

Ruey-Cheng Chen, Q. A. (2019). Correcting for Recency Bias in Job Recommendation. *CIKM '19: Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2185-2188.

S.Ephina Thendral, C. (2018). Understanding Personalization of Recommender System: A Domain Perspective. *International Journal of Applied Engineering Research, 13*(0), 12422-12428. Retrieved from https://www.ripublication.com/ijaer18/ijaerv13n15_100.pdf

Wanrong Gu and Dong, S. a. (2014). An Effective News Recommendation Method for Microblog User. *TheScientificWorldJournal*, 907515.

Wen, H. a. (2012). A hybrid approach for personalized recommendation of news on the Web. *Expert Systems with Applications: An International Journal, 39*, 5806-5814.

Zhai, M. A. (2009). An Online News Recommender System for Social Networks. *Urbana, 51*(0), 61801-61804.

Secure Multiparty Computation Applications: A Review

Nawaz Malla[1]

Institution's Affiliation: Maulana Azad National Urdu University, Hyderabad, India
mallanawaz2@gmail.com.[1]

Abstract: *From last decade, Secure multiparty computation has been an interesting area of research in applied cryptography. Secure multiparty computation started in early 1970's is gaining fame in practical world. Due to growth of internet, the data breach whether personal data or organizational data may get compromised or can be abused by hackers, also many of the companies share the user's data, which would have negative effects cooperative computations triggered tremendous opportunities where computations on private inputs can be conducted jointly by multiparty. These competitors and untrusted parties can use secure multiparty computation on each other's private inputs without revealing the inputs. To perform computations, one should know the inputs of other parties, if the computing party is not trusted, the privacy becomes a concern such type of problem is called secure multiparty problem. There has been lot of research how to provide privacy then computation take place between competitor or mutually untrusted parties. For such problems, secure multiparty computation can be used. The paper aims to review the application, some techniques used, and limitations of secure multiparty computation.*

Keywords: *secure multiparty computation, privacy.*

1. Introduction

From last few decades, the data privacy and security has become the primary concern to everyone. People were not concerned about their data than before. Due to the rise in technological advancements and internet, it has been a challenging task to provide the data security and data privacy of the data, when data is distributed over large distributed networks. As everyone is now concerned with their data, lot of research is going on how to provide the data security and privacy to the participants in the network. One of the technique which provide the solution to the problems of data security and data privacy is the secure multiparty computation.

The secure multiparty computation, also known as multiparty computation started late in 1970's. It does not gained its popularity at the time as it was not implemented practically. In 1982 it was introduced as secure two party multiplication (2PC). The secret sharing of inputs and zero knowledge proofs in malicious case can assure the bad behavior detected by majority of honest players. The dishonest person is eliminated or his input is made public (Goldreich & O. (Ed.), 2019). SMC is focusing on limited problems, but now after the concept of distributed

systems, it has underwent through the rapid change and is now concerned to a variety of problems in computation domain. There are number of problems to be addressed using secure computation like privacy preserving intrusion detection, bribing in distributed systems by malicious nodes. The major challenges are the computational overhead, high communication cost, mobile friendly SMC and scalability. In order to provide the solutions for the problem, we can use the SMC using blockchain to provide the data privacy. The secure multiparty computation is a subset of cryptography which can provide the privacy to data in blockchain. SMC is now a proven reality and can be used for secret sharing and has the power to eliminate the need of third party in the system (Choi & Butler, 2019).

2. Background

Secure multiparty computation may be defined the problem of 'n' players to compute jointly on an agreed function in a secure way on the inputs without revealing them. Secure multiparty computation started early in 1970's. it was known as multiparty computation at that time. It does not gained its popularity at that time as it was not implemented practically. in 1982's it was introduced as secure two party multiparty computation. It is used to solve lot of problems of computation without revealing the inputs to other parties. Finally it came with a name as secure multiparty computation in which the functions of different types are computed, that is the reason it is sometimes called as SFE secure function evaluation. the secure multiparty computation is used for utilization of data without compromising the privacy. It is the cryptographic subfield which helps in preserving the privacy of the data. the secure multiparty computation is very trending topic nowadays. The emerging technologies like blockchain, mobile computing, IOT ,cloud computing has resulted in rebirth of secure multiparty computation. the secure multiparty computation has become the hot area of research from last decade due to the rise of blockchain technology. The researcher are now more interested to implement the secure multiparty computation in distributed systems. We can use the secure multiparty computation as a tool for computing on private data. Unlike in centralized systems the secure multiparty computation may have better performance in distributed systems. the secure multiparty computation can be implemented using blockchain to provide solution to different problems where a set of parties compute a function by keeping the input as private.

3. Architecture

The secure multiparty computation which provides a protocol where no individual can see the other parties data while distributing the data across multiparties. It enables the data

scientists and analysts to compute privately on the distributed data without exposing it. The simple architecture of the secure multiparty computation is shown in figure below,
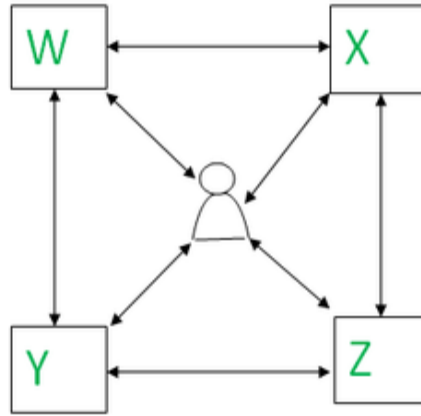

Fig: *Multiparty sharing data among each other without any third party using a specified protocol.*

The co-workers want to compute the maximum salary without revealing their individual salary to others. To perform such a computation, secure multiparty computation is implemented to calculate the maximum salary. The parties in a distributed manner jointly perform on a function to calculate it without revealing the salary. Using secure multiparty computation, we need not to worry or drop the features for preserving the privacy of data. Data in use is kept in encrypted form , broken up and distributed across parties, that no chances of qutam attacks.

It is impossible to have a trusted party in real world, as all parties communicate each other in one or the other way In such scenario, the parties may get corrupted. The corrupted parties have the behavior like semi honest and malicious (Evans & Kolesnikov, 2017). A semi-honest opponent is one who follows the specified protocol, but makes the parties corrupted. The protocol is run honestly, but they try to extract information from the messages exchanged between parties. A malicious adversary makes an attempt to breach the security and do not follow the specified protocol. The adversary can make the changes during the execution process of the protocol (Evans & Kolesnikov, 2017). While using multiparty computation, we assume the party is honest which follows all the protocols.

4. Applications

Secure Multiparty computation is used for wide range of applications. Secure Multiparty computation is used by distrusting data owners to compute functions jointly. Some of them are listed below:

Privacy preserving intrusion detection

To prevent fraudulent intrusion into the computing systems of the organizations, we need to collect the data of hackers and their behavior. We can cannot mix the different databases of profile and behavior of hackers, thus we use secure multiparty computation for data mining and maintaining privacy for intrusion detection. Privacy preserving data mining. For databases the decision trees cannot be made as these will not provide privacy as the algorithms have default setting that the database is fully available. ID3 algorithm was developed by lindell and pinkas for generating decision tree providing privacy to multiple databases using secure multiparty computation

Privacy preserving geometric computation

Two parties having two figures, without knowing the shape of figure, we can compute the intersection of two figures while preserving privacy by using the secure multiparty computation without learning the relative position of the shapes.

Secure machine learning

The machine learning systems are provided with data during training phase and inference phase. Inference phase allows the client to submit requests while keeping the data as private from the server and other clients.

Secure auctions

Providing the privacy in auctions, secure multiparty computation is used. It provides the privacy to both seller and the bidder, so that one player may not learn the other players bid.

Yao's millionaire problem

It was implemented in 1982 to find who is richer among two, without revealing the input. Yao's Millionaire Problem was the first problem using multiparty computation.

5. Deployment

Practically the secure multiparty computation has been used in different field. some of the deployments of secure multiparty computation are;

Sugar beet auction

It was the first commercial application of secure multiparty computation. The bidding platform was created for sugar beet production contracts by Danish researchers in

collaboration with the government and stakeholders. the platform provided good bidding privacy and security to auction participants.

Estonian students study

Secure multiparty computation was used to for analysis to find the correlation between working students during studies and failure to graduate on time. There was no correlation found between the same. More education was correlated with higher income and these results were calculated using secure multiparty computation.

Boston wage equity study

A web based multiparty computation aggregation tool was designed by researchers of boston university, through which the salary data of employees was collected privately and with legal protection for study purpose.

Key management

Security to sensitive data of the organization is the biggest issue. The sensitive data of the organization is stored in encrypted form and then distributed for computation. Key management performs the role in encrypting and decrypting the same.

6. Techniques

There are number of techniques developed for Secure Multiparty computation protocol construction having different features. Some of the techniques used in Secure Multiparty computation are listed below;

Shamir Secret Sharing

The secret sharing is utilized as the basic tool, when there is honest majority in secure Multiparty computation. A secret sharing scheme is that a secret s is shared among 'n' parties, such that t+1 parties or more come together to reconstruct the secret. The parties lesser than '*t*' cannot get any information or reconstruct the secret. The scheme which fulfils the requirements of *t+1* out of n is called threshold secret sharing scheme.

Input Sharing

Every party shares the input using Shamir secret sharing. The circuit is being provided the input for computation. Every party keeps his input private by adding some random number to the input and finally after getting the output the random number is known to the party is removed and we get the output.

Circuit Evaluation

The circuit is evaluated by parties one gate at a time .The gates are evaluated serially from input to output. The evaluation consists of computation of addition and multiplication gates. For inputs a(x) and b(x), the output of addition for the ith party is calculated as $c(i)=a(i)+b(i)$. Similarly the output of multiplication for the i_{th} party is calculated as $c(i)=a(i).b(i)$.

Private set intersection

The private set intersection protocol is very efficient for two party problem. Two parties who wish to find the elements of intersection with private set of inputs without revealing the input, the private set intersection is better approach for both honest and dishonest adversaries.

Threshold cryptography

The threshold cryptography aims to carry out the cryptographic operations for set of parties without holding the secret by any of the single party. RSA algorithm is used for the scheme where the basic function is $y=x^e \mod n$. RSA is used for encrypting the secrets or messages.

Dishonest majority MPC

In Secure Multiparty computation there can be both honest and dishonest parties. The Secure Multiparty computation is secure as long as there is honest majority. If the adversaries are corrupt more than majority, new approaches are required for security. For dishonest majority there are protocols like GMW oblivious transfer, grabled circuit, Tiny oz and many more protocols.

7. Limitations

Secure multiparty computation being used for solving different problems, there are few limitations. One of the limitation is the computational overhead and high communication costs.

Computational Overhead

To provide the security we need to generate the random numbers, the random number generation requires the more computation overhead which slows down runtime.

High communication costs

Distributing the data to multiple parties for computation over the networks leads to higher costs of communication.

8. Conclusion

The data breach whether personal data or organizational data may get compromised or can be abused by hackers, also many of the companies share the user's data, which would have negative effects.

Secure multiparty computation a vibrant field provides versatile primitive for designing privacy preserving applications. Secure multiparty computation emerged as a powerful cryptographic tool with opportunity to create, develop and apply in different areas

Reference:

1. Canetti, R. (2000). Security and composition of multiparty cryptographic protocols. *Journal of CRYPTOLOGY*, *13*(1), 143-202.

2. Evans, D., Kolesnikov, V., & Rosulek, M. (2017). A pragmatic introduction to secure multi-party computation. *Foundations and Trends® in Privacy and Security*, *2*(2-3).

3. Goldreich, O. (Ed.). (2019). *Providing sound foundations for cryptography: on the work of Shafi Goldwasser and Silvio Micali*. Morgan & Claypool.

4. Choi, J. I., & Butler, K. R. (2019). Secure multiparty computation and trusted hardware: Examining adoption challenges and opportunities. *Security and Communication Networks*, *2019*.

5. Wigderson, A., Or, M. B., & Goldwasser, S. (1988). Completeness theorems for noncryptographic fault-tolerant distributed computations. In *Proceedings of the 20th Annual Symposium on the Theory of Computing (STOC'88)* (pp. 1-10).

6. Bellare, M., & Rogaway, P. (1993, December). Random oracles are practical: A paradigm for designing efficient protocols. In *Proceedings of the 1st ACM Conference on Computer and Communications Security* (pp. 62-73).

7. Yao, A. C. C. (1986, October). How to generate and exchange secrets. In *27th Annual Symposium on Foundations of Computer Science (sfcs 1986)* (pp. 162-167). IEEE.

8. Ishai, Y., Kushilevitz, E., Lindell, Y., & Petrank, E. (2006, August). On combining privacy with guaranteed output delivery in secure multiparty computation. In *Annual International Cryptology Conference* (pp. 483-500). Springer, Berlin, Heidelberg.

9. Goldreich, O., & Lindell, Y. (2001, August). Session-key generation using human passwords only. In *Annual International Cryptology Conference* (pp. 408-432). Springer, Berlin, Heidelberg.

A Comparative Study of Indian Higher Education Institutes of National Importance Websites on the Challenges regarding Web Accessibility

Shaikh Afzal V[1] and Muqeem Ahmed[2]

Department of Computer Science and Information Technology,

Maulana Azad National Urdu University, Gachibowli, Hyderabad[1, 2]

bagbanafzal@gmail.com[1], muqeem.ahmed@gmail.com [2]

Abstract: *The internet became an integral part of our life; one of the major applications of the internet is the web. The Web has transformed our everyday life, and the way of getting information. It is easy now to get any information about anything from the internet; even a person generally relies on the official website of a particular organization. Educational services on the website are growing rapidly so it is necessary for the end user, regardless of user's any disabilities, will be access to these official websites in an efficient way, but the homepages of higher education institutes in India still doesn't fulfill web accessibility benchmarks. The paper expresses issues of the web accessibility exist in 116 higher education institutes in India funded by the central government. Based on the result It was found that higher education institutes' websites have one common frequent issue is linked to a absence of alternative text of the picture. Also, we found that websites violate present Web Content Accessibility Guidelines 2.0 of World Wide Web Consortium's standards. Numerous other issues also recognized regarding Website accessibility which needs to follow Web accessibility policies to make websites more usable.*

Keywords: *Accessibility, Web accessibility, Institutes of National Importance, India, university, Website, Web Content, Accessibility Guidelines, WCAG 2.0, IIT, NIT, IISERs, Central Universities*

1. Introduction

Internet technology affected many areas of life, including educational system (S. Luján-Mora, 2013). Due to availability of internet educational websites has increased gradually. Report of Global Digital shows that netizens in January 2021 hit, up by 316 million (7.3%) since this time last year (Digital 2021: Global Overview Report). The world's population stood at 7.83 billion at the beginning of 2021. According to World Health Organization, nearly 15% of the populace abides with a disability (WHO 2011).

There are many higher educational websites exist on internet, each website have its own form and style. But the problem is not every websites follows the recommendation provided by the W3C (World Wide Web Consortium). The Web Consortium give suggestions that the website should function effective with optimal feature should be there in website to access it. A well developed and designed website should follow recommendation given by the W3C. The W3C recommendations are made to allow websites to be accessed by maximum users and visits.

Web accessibility means to design websites such a way that maximum people should understand, perceive, operate and it should support different technologies (devices and browsers). W3C made the Web Content Accessibility Guidelines (WCAG) 2.0, which make and give guidelines to web developer and designer how to make website more accessible by the individuals regardless of their physical disability or constraint , their surroundings, physical condition, situation, people use different gadget to access website and eliminate accessibility errors (W3C WCAG 2.0 2008)

A functional accessible web page abides by accessibility recommendations are better appear correctly on any device and any web client or browser. Even though web designers and developers think that it will increases the budget and cost estimation for making web pages as per W3C guidelines but after getting the necessary skills and training which is initial investment the cost of developing and sustaining accessible web sites is less costly as compare to those of which not follow guidelines (E. Fgee, 2010)

Web pages comply with WCAG2.0 are small and take very less time to load also pages which are separated with Cascading Style Sheets (CSS) for presentation also load faster and take very less size.

WCAG 2.0 covers many recommendations which make website more accessible not only for disable people including complete blindness, myopia or amaurosis, deafness and hearing impairment, learning or reading disabilities, intellectual and speech disabilities, limited movement, or combinations of these, disability due to age and condition but also for normal people. In addition, we can improve search engine optimization ranking by making web accessible.

Educational websites plays major role to connect students, faculties, teachers, non-teaching and administrative staff, so website must have accessibility features. As a user's first impression of an educational website is very crucial; therefore, we conducted comparative study on Indian higher education institutes of national importance which includes Indian Institutes of Science Education and Research (IISERs), Indian Institutes of Technology (IITs), central universities, Indian Institute of Science (IISc) and National Institutes of Technology (NITs), to find out the accessibility problems of above said institutes, and assessment of these websites according to W3C guidelines regarding website accessibility. To check and analyze web accessibility we used automatic tools to verify websites according to WCAG 2.0.

These institutes are national importance and funded by central government for higher education and research and Indian student's first prefer these institutes for many reasons. Most of these institutes have global recognition and ranking so it is very important that these websites must be easily accessible as per guidelines of WCAG 2.0 and according to recommendations by Indian government to be followed by Indian Government Websites.

Following is the overview of guidelines of WCAG 2.0 (WCAG 2.0, 2008)

Principles - There are four standards at the top that give the outline to Web accessibility: perceivable, operable, understandable, and robust.

Everybody who likes to operate the Web must have included that is:

- Perceivable - User interface (UI) elements and data must be neat for users in ways perceivable form.
- Operable – Navigation in different page and User interface elements must be operable and feasible.
- Understandable - Functioning of user interface and instruction should be understandable and comprehensible.
- Robust - Contents and elements must be sturdy and sturdy enough that it will be transcribed smoothly by a different variety of user agents of any web client, including assistive technologies.

In the event that any of these are false, users with inabilities won't utilize the Web.

Guidelines. 12 rules give the fundamental objectives that authors / developer should run after to make content more available to users with different disabilities.

Success Criteria - To address the issues of various gatherings and various circumstances, three degrees of conformance are characterized: AAA (highest), AA (moderate), and A (lowest).

There is a great deal of examination completed on institutions of higher education as far as web accessibility assessments. Like, A recent report showed that Brazilian university sites introduced issues identified with the utilize of forms; the primary issue is seen when use tables which didn't follow WCAG 2.0 standards. These forms create problem in web site to access for individuals without disabilities (T. J. Bittar, 2012).

In 2015, an article revealed data concerning an entrance examination of a few Colombian online government sites. The examination uncovered an absence of information on the principles and accessibility rules of the WCAG (J. A. Riano Herrera, 2015).

A recent report's outcomes introduced the web accessibility determination for few sub domains of the institute or university online interface and revealed the shortfall of consistence rules expected to arrive at the consistence level AA standardized by WCAG 2.0 (Veronica Segarra, 2017).

Exploratory investigations was completed by researchers regarding accessibility of Indian Universities home pages by utilizing distinctive web accessibility assessment automated tools based on WCAG 2.0 guidelines. (Ismail and Kuppusamy, 2018)

Global Digital report shows that in January 2021 there were 624.0 million netizens in India. The quantity of netizens in India expanded by 47 million (+8.2%) somewhere in the range of 2020 and 2021. Indian Internet penetration is remained at 45.0% in January 2021. (Digital 2021: India).

2. Materials and Methods

The assessment of educational websites was completed as per the WCAG 2.0. During the assessment, tools were utilized to recognize a portion of the web accessibility issues. Utilizing this approach, 116 Indian higher education institutes were assessed; only the sites' home or main pages were investigated.

Assessment of the educational websites was completed as per the WCAG 2.0. During the assessment, tools were utilized to recognize a portion of the web accessibility issues. Utilizing this approach, 116 Indian higher education institutes were assessed; the only sites' home main pages were investigated. An automated tool didn't know that alternative text is appropriate for image and tools can know that the text is for an image but can't know context and content of image. There are many tools are available to evaluate website so first we analyzed tools which covers most of WCAG 2.0 guidelines after that evaluated which are easy to use and the most suitable. The tools accessible for dissecting web accessibility are DYNO Mapper, AccessMonitor, A11Y Compliance Platform, AChecker, AATT (Automated Accessibility Testing Tool), eXaminator, Accessibility Checker, TAW, Accessibility Checklist, Tenon, accessibility management platform (AMP) and WAVE..

There are many parameters we considered to choice a tool like is it free or paid service, require an account to use it, the response time, which license type they used, and results reported format like text, visual, comma-separated value (CSV), if is it have web browser add-on or plug in. Based on analyzing all tools we selected WAVE tool for our study, following is the screenshot of WAVE plug-in which must be installed before analyzing website. Because

WAVE tool is free to use, doesn't required an account, it have browser plug-in and result is in visual form.

We selected 116 websites from Ministry of Education website in which we have 23 IITs , 31 NITs, 54 central universities, 7 IISERs and on IISc. We analyzed these website one by one using WAVE tool plug in for chrome browser then we get following results as shown in table 1.
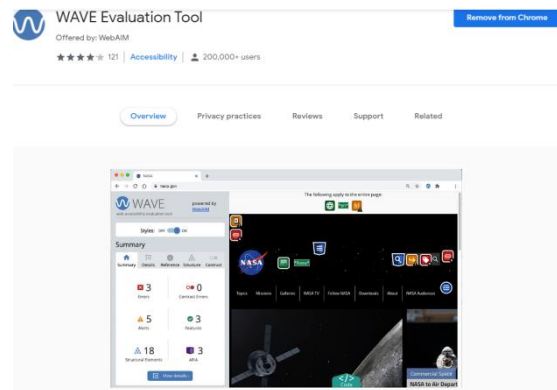
Figure 1: WAVE plug-in for Google chrome

Due to space constraint we made average of all IITs, NITs, central universities and IISERs as given table 1.

3. Results and Discussion

Following is the different parameters used by WAVE tools to analyses the website,

Errors: Basic underlying issues, which forestall screen readers from having the option to peruse the content in a significant manner.

- Missing Alternative text: Add an alt attribute to the picture. The quality worth ought to precisely and briefly present the content and working of the picture.

- Empty button: Spot text content inside the <button> component or give the <input> component a value attributes.

- Empty Link: Eliminate the unfilled link or give text inside the link that portrays the usefulness as well as target of that link.

Contrast Errors: Increment the difference between the forefront (text) color and the foundation color. Enormous text content (bigger than 18 point or 14 point bold) doesn't need as much difference as more modest content.

Alerts: Not basic for perusing the page, enhancements would help the screen readers to deal with the page better, repetitive alternative text, dubious alternative text, long alternative text, broken same-page interface, Link to PDF report

ARIA: Accessible Rich Internet Applications (ARIA) gives improved semantics and accessibility to web content.

Sr No.	Name of Institute(s)	Errors	Contrast Errors	Alerts	Features	ARIA
1	IITs – 23	54.73	61.47	128.60	34.39	26.73
2	NITs – 31	52.25	45.80	155.2	29.61	90.83
3	Central Universities - 54	76.79	76.59	381.12	24.72	97.12
4	IISERs – 7	26.57	14.42	35.14	9.714	10.28
5	IISc – 1	64	236	141	22	19

Table 1: Result from WAVE tool

Based on our study we found that IITs have average 54.73 errors in which IIT Madras have only 1 error where as IIT Dharwad have 192 errors same in NITs , NIT Tiruchirappalli have only 6 errors and NIT Kurukshetra have 292 errors in home page. Central universities average error is 76.79 out of which Guru Ghasidas Vishwavidyalaya and Shri Lal Bahadur Shastri National Sanskrit University don't have any error where as Central University of Haryana have maximum errors 2096 in this category. In IISERs average error is 26.57 in which IISER Thiruvananthapuram have only 8 errors and IISER Bhopal have maximum 86 errors in home page. Study shows that every single website is needed to work on web accessibility guidelines.

4. Conclusions

The outcomes got by this accessibility assessment of institutes websites show that no web site arrived at a worthy accessibility level. Subsequently, it is important to survey the mistakes that need revising as per the WCAG 2.0.

To enhance this assessment, it's proposed that evaluation should be performed with large number of different users to get results related to their particular limits and technology which is assistive.

Disclosure Policy

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

References

4   Abid Ismail, K.S. Kuppusamy (2018) Accessibility of Indian universities' homepages: An exploratory study, Journal of King Saud University - Computer and Information Sciences

5   Central Universities (2021) https://www.ugc.ac.in/centraluniversity.aspx

6   Digital India (2021) https://datareportal.com/reports/digital-2021-india

7   Fgee, H. A. Abakar and A. Elhounie (2010) Enhancement of Educational Institutions Dynamic Websites by Adding Security and Accesibility, Fourth International Conference on Next Generation Mobile Applications, Services and Technologies

8   Guidelines for Indian Government Websites https://web.guidelines.gov.in

9   Global Overview Report. (2021) Datareportal. https://datareportal.com/reports/digital-2021-global-overview-report

10   List of IITs (2021) https://www.iitsystem.ac.in

11   List of NITs (2021) https://www.nitcouncil.org.in

12   J. A. Riano Herrera and J. A. Ballesteros Ricaurte (2015) Web accessibility: Study web accessibility in public places of the Colombian State, Latin American Computing Conference (CLEI).

13   J. Carter and M. Markel (2001) Web accessibility for people with disabilities: an introduction for Web developers, in IEEE Transactions on Professional Communication

14   T. J. Bittar, P. de Mattos Fortes F. B. Faria, L. A. do Amaral and R. (2012) An assessment of accessibility in contact forms of Brazilian public universities, 7th Iberian Conference on Information Systems and Technologies (CISTI 2012).

15   Veronica Jorge Campoverde and Segarra-Faggioni 2017. Evaluation of the web accessibility of university web portal: Case Study: Universidad Tecnica Particular de Loja. In Proceedings of the 2017 9th International Conference on Education Technology and Computers (ICETC 2017). Association for Computing Machinery, New York, NY, USA,

16   WAVE Tool (2021) https://wave.webaim.org/

17   Web Content Accessibility Guidelines (WCAG) 2.0. (2008). W3C. https://www.w3.org/TR/WCAG20/

18   World Report on Disability (2011) https://www.who.int/teams/noncommunicable-diseases/sensory-functions-disability-and-rehabilitation/world-report-on-disability

Recognition of Offline Handwritten Characters: A Survey

Md Faizur Rahman[1] and Mohatesham Pasha Quadri[2]

Maulana Azad National Urdu University

mdfaizurrahman9@gmail.com [1], syedmohtesham@manuu.edu.in [2]

Abstract: *In the present era, there are many challenges in the visualization of the data which can be quite difficult to solve as they are different in nature in terms of region. But Machine learning and deep learning offers many algorithms to overcome and solve these problems, like in pattern recognition, image or text recognition will give better results. Handwritten Character Recognition (HCR) is one of the problems of pattern recognition. This can solve by using machine learning and deep learning algorithm with higher accuracy. Offline HCR is still an open research area because of getting the higher accuracy and variations of written characters by writers. In this paper, present a comprehensive review of HCR and different languages like English, Arabic, Urdu, Devanagari, Bengali, etc. In this, we proposed a comparative analysis of different algorithms (Feature Extraction (FE), Convolutional Neural Network (CNN), Support Vector machine (SVM), K-Nearest Neighbour (KNN), Neural Network (NN) etc.,) that were already in different languages with their variation of accuracy.*

Keywords: HCR, Offline HCR, Preprocessing, Feature Extraction, CNN, SVM, KNN, NN

1. Introduction

A large number of researches already have been done to achieve the solution for character recognition. Recognition of characters is a procedure of changing over the computerized picture of content into machine-encoded content. It is widely used for extracting the information from digital/printed/handwritten records. Artificial intelligence, Signal processing, Pattern recognition, and machine vision are average instances of fields of investigates that include character recognition. Comprehensively, character recognition can be utilized for an immense scope of utilizations that incorporates whatever dependent on the possibility of change of anything humanly clear into machine manipulatable portrayal. It is broadly divided into two parts (a) Online Handwritten character recognition (b) Offline handwritten character recognition. According to the research study, Online HCR is a process of recognition of character which is written by using a special pen on the digital device like Palm PDA, Google Handwrite. When a pen writes something on the digital device or

electronic surface of the device then it generates the series of electronic signals and stores in the memory. After that Online handwritten characters recognize by using some features like Stroke pressure, velocity trajectory, etc. Offline HCR is still a state of art. It is more difficult than Online HCR. Again Offline HCR s divided into two parts (a) Printed Offline HCR (b)Handwritten Offline HCR. In printed Offline HCR, There are equal sizes of printed characters and also they have the same style. Handwritten Offline HCR has a different size and style of handwritten character because handwritten characters are written by different writers. [1]

This paper is further divided into these sections. Systematic Literature Review, Finding, Results, and Conclusion.
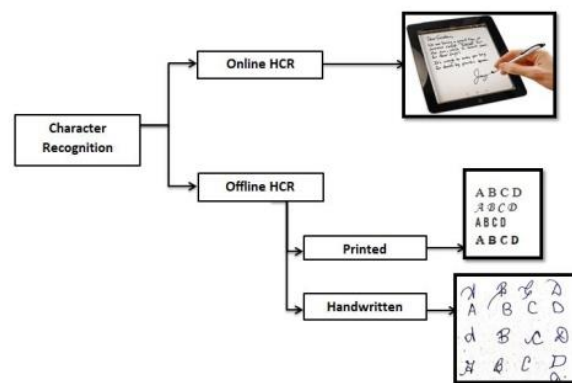
Fig: Classification of Character Recognition

2. Motivation

Most associations use reports to secure data from clients. These reports are by and large transcribed. Such archives can be structures, checks, and so on. For their simpler recovery or data assortment archives are changed and put away in computerized designs. Basic practice to deal with that data is physically filling the same information into PC. It would be tedious and tedious to deal with such archives physically. Henceforth the prerequisite of an exceptional Handwritten Character Recognition Software emerges which will naturally perceive writings from the picture of records. The procedure of extricating information from the manually written archives and putting away it in electronic arrangements has made simple by Written by hand Character Recognition (HCR) Software.

Banking areas, Health care businesses and some such associations where manually written records are utilized normally. HCR frameworks additionally discover applications in recently

rising regions where handwriting information passage is required, for example, the advancement of electronic libraries, interactive media database and so on.

3. Literature review

Features and types of classifier algorithms are base for achieving higher accuracy in different languages. Here the following review of literature for handwritten character recognition using machine learning and deep learning algorithm.

A. English characters

i) Vinoj Jayasundara et.al [2] had presented a strategy for expanding the size of a dataset by exploiting the ideas in CapsNets and exhibit the presence of this method on notable written by hand character datasets. Initially, there are 3 Convolutional layers in the Capsule Network along with 4 layers which will be the primary Capsule layer, and finally the last layer act as a character capsule layer which is already associated with the capsule layer. It utilized two decoders associated with the same number of loss functions and joined two reconstructed character images together for getting an improved quality image. Accuracy for EMNIST (letters, balanced, digits) and MNIST datasets with 200 samples per class is 92.79, 87.82, 98.96, 98.68.

ii) Dewi Nasien, Habibollah Haron Sophiayati Yuhaniz [3] Proposed a model for identification of handwritten English characters. For preprocessing just includes a thinning procedure and utilizations Freeman chain code (FCC) as the portrayal method of an image character. In the FCC technique that utilizes 8-neighborhood that start as 1 to 8 from direction labeled as. To build the FCC a randomized algorithm is used. The support vector machine is choosen for the classification characters and the feature Radial basis is utilized to construct an SVM model. For the experimental result, it uses data from the NIST database. Accuracy of the uppercase letter is 86%, a lowercase character is 88% and combined uppercase and lowercase is 73%.

iii) Pritam dhande, Reena kharat [4] Proposed the development of the first horizontal histogram for a sample image. The normal stature of an ascending region is considered as a threshold. The status of each rising segment will then be tested whether it is more prominent or equal to the threshold, at that point every line is segmented by a binary image and vertical projection methods for segmentation and Convex hull algorithm, zoning system used for extraction of features. Support Vector Machine is used for classification and recognition. CCC benchmark dataset is used for all experiments. .

iv) Ms. Harita Dave1, Mitesh Patel [5] They proposed a handwritten English character recognition system. For preprocessing steps, they utilized Gray-Scaling, Noise removal, Binarization, Inverted picture, Skeletonization, and Normalization. From that point onward, they segmented the individual character from the word utilizing Bounding Box, Edge identification. They used a Multilayer Perceptron Neural Network for classification and recognition which gives over 70% precision for manually written English characters. Just as the proposed system diminished the preparation time and cost to distinguish handwritten characters.

B. Arabic characters

i) Hashem Ghaleb, P. Nagabhushan, Umapada Pal [6] proposed an algorithm that gets the idiosyncrasies of the Arabic context into thought. By utilizing graph-theoretic modeling firstly extracted the sub-words from content line images or word images. To identifies segmented points for constituent letters they introduced two algorithms first one is average stroke thickness and the second one is the statistical model of stroke thickness and lastly they merged the segmented points that come from both algorithms based on some rules. The proposed calculation is tried on a lot of 450 written by hand images comprising of 3579 characters browsed the IFN/ENIT dataset. Finally got 72% of characters are accurately segmented, another 19% of characters are over-segmented, and the remaining characters come under segmented.

ii) Supriana, I. and Nasution, A. [7] developed a framework for Arabic Optical Character Recognition (AOCR) that has five phases named preprocessing, feature extraction, thinning, segmentation, and classification. Used Hilditch thinning algorithm joined by two layouts and Zidouri algorithm for segmentation. A decision tree created for classification using C4.5 algorithm. The line segmentation has an accuracy rate of 99.9%. Standard deviation is shown can reduce over-segmentation and quasi-line. Standard deviation can be shown to reduce quasi-line and over-segmentation. Segmentation of letters had an exactness of 74%, tried on six unique textual styles. The classification segments had an exactness of 82%, checked by cross validation. Lamentably, the system's overall efficiency was just 48.3%.

iii) Ahmed El-sawy, Mohamed Loey [8] They created their own Arabic Handwritten Character dataset included 16800 characters from Alef to yeh which was composed of 60 members. For preprocessing the dataset they utilized RGB to grayscale, smoothing, and filtering. They utilized the Convolutional Neural Network for recognizing the Arabic character. For training purposes, they used 13440 samples from 16800 characters and the remaining 3360 samples for

testing purpose, After testing the model they achieved 94.9% accuracy and 5.1% misclassification error.

iv) Almansari O, Hashim Nik [9] They proposed a model in deep learning with a Multilayer perceptron(MLP) and Convolutional Neural Network (CNN). They utilized open dataset name as AHCD for preparing and testing the model which comprises 16800 manually written characters. At the point when utilized MLP model they accomplished 72.08% precision however they confronted overfitting issue. So MLP was neglected to perceiving the dataset with high precision and no overfitting. In the CNN model, they used 25 epoch, 23 batch size, 5×5 filter matrix, and 0.2 dropouts for getting the most noteworthy precision 95.34% and the least test loss of 0.29112.

Recommended font sizes are shown in Table 1.

C. Devanagari Characters

i) Sonawane, P., Shelke, S. [10] effectively attempted to classify written by hand Devanagari characters utilizing transfer learning mechanisms with the assistance of Alaxnet. Used transfer learning for training the model with little training dataset which gives better what's more, quicker outcomes. Alexnet, a Convolutional neural, which is prepared over an own 224x224 input size around 16870 samples of 22 consonants Devanagari content which gives noteworthy outcomes filter matrix is used for extracting features for trained the model, the whole dataset is categorized into 56% for training, 24% for validation and 20% for testing and finally got 94.49% accuracy for validation and 95.46% accuracy from testing.

ii) Madhuri Yadav, Dr. Ravindra Purwar [11] attempts to automate the recognition of written by hand Hindi detached characters utilizing different classifiers. They used two histograms named as oriented gradients and profile projection for extracting the feature. Assessed transcribed character precision utilizing five unique classifiers Bagged trees, Subspace segregates, k-NN, Quadratic SVM, and weighted k-NN reasoned that the presentation of classifiers utilizing proposed features is over 93% and even Quadratic SVM gives an exactness of 96.6%.

iii) Nikita Singh [12] utilizes a technique called a piecewise histogram of oriented gradients(HOG) for partitioning the images in similar sizes for feature extraction. A feature vector includes HOG characteristics of all segments for setting up the neural framework. The suggested method achieves a 99.27% classification exactness in preparation and can recognize the distinctive manually written Devanagari characters with a standard 97.06% accuracy. The

suggested method could be beneficial to peruse the written by hand material in the application to daze individuals.

iv) Deepika Gupta, Soumen [13] the approach proposed uses the polygonal approximation first to obtain a digitally straight line portion of the term. In these portions, apply the traverse algorithm that results in character segmentation. After character segmentation, the upper and lower segmentation of the modifiers is performed using the word width statistical information. This has the favorable position of segmenting all of the modifiers using this particular logic. The approach is checked on Hindi manually written text and has produced promising outcomes with 95.70% exactness. This proposed approach is fit to dealing with form variations and diverse writing styles.

D. Urdu Characters

i) Shahbaz Hassan, Ayesha Irfan, Ali Mirza [14] presents a successful written by hand cursive Urdu content recognition system. The entire procedure utilizes content lines and depends on segmented the character implicitly for recognition. By using Convolutional Neural Network extracted the optimized features. After that bi-directional long short term memory (LSTM) network is utilized these optimized features for classification. They gathered 6000 one of a kind book lines from 600 writers as a dataset. This dataset utilized 4000 samples for prepared the model and the remaining 2000 samples are similarly partitioned for Validation and testing. The blend of CNN-LSTM achieved 83.69% of the character recognition rate on the 1000 testing line samples.

ii) Nawaz, Calefati, A., et al [15] made another written by hand Urdu characters dataset associated with classes that are appropriate for deep learning. For his research two methods of deep metric learning relate to performance: one is a Siamese network and the other is the triplet network. A Siamese network or a triplet network both learn distance metrics, in which comparable models are often mapped near each other furthermore, disparate models are mapped more remote separated. After experimenting with each network they show that Siamese network built on alexnet good performance when using MNIST (60000 training and 10000 testings) handwritten digit dataset and 20324 segmented Urdu characters but triplet network give better performance.

iii) Saad Bin Ahmed, Saeeda Naz [16] They make another and far-reaching transcribed Urdu character dataset which is known as Nasta'liq manually written dataset(UNHD) composed by 500 essayists. Some preprocessing techniques are applied on UNHD which incorporates a

median filter for the expulsion of baseline and noise, gray-scaling, the horizontal projection for skew correction, and for segmented the content line utilized projection profile. For prepared the model utilized 7200 content line, validate the model by 4800 examples and the staying 2400 examples for the test line. They utilized a two-dimensional BLSTM classifier for assessing the performance and accomplished generally excellent precision which is around 6.04 to 7.93% error rate.

E. Bangla Characters

i) C Saha, H Faisal, et al [17] proposed a new Deep Convolutional Neural Network(DCNN) model named as BBCNET-15 for recognizing the manually written Bangla essential character. The model comprises of 6 each in the Convolutional and max-pooling layers with 2 completely connected layers followed by the softmax. Dropout regularization is utilized to abstain from overfitting. For implementing this model they utilized the CMATERdb dataset. At last, they achieved 98.86% training accuracy with the least loss value 0.03434 and 96.40 % test accuracy with 0.1669 loss value.

ii) Purkaystha, B., Datta, T. and Islam, M. [18] proposed a deep convolutional neural system that consists of 2 convolutional layers and 3 dense layers for recognizing isolated Bangla characters. For finding a valuable set of features used kernel and local receptive /responsive fields and afterward they utilized densely associated layers for the segregation task. For recognizing Bangla characters they used the BanglaLekha-Isolated database. This accomplishes 98.66% precision in 10 numeral classes, 94.99% in 11 vowel character classes, 91.60% precision in 20 compound letter classes, 91.23% precision in 50 letter character classes in order, and 89.93% precision in practically all 80 Bangla characters.

iii) Mujadded Al Rabbani, Sabbir Ahmed [19] proposed a modified architecture called ResNet-18 (Residual Network) that is particularly robust in classifying Bangla handwritten characters. For preparing and estimating the model execution they utilized two huge datasets initial one is BanglaLekha isolated and another is CMATERdb datasets. In this model, they consolidated a dropout layer in every one of the ResNet modules which further upgrade the ability of ResNet architecture to improved the performance of classifying Bangla HCR by a significant margin. At last, they accomplish 95.10% accuracy which is critical 0.51% execution improved when they utilized modified ResNet-18 architecture.

iv) S M Azizul Hakim, Asaduzzaman [20] proposed a consecutive Convolutional Neural network model with six Conv layers two completely associated layers, and one output layer

likewise utilized a 0.2 dropout layer after each Convolutional and completely associated layer for recognition of Bangla numerals and characters. Right now to perceive 60 Bangla manually written characters(10 numerals and 50 basic characters). This accomplishes 99.44 % exactness on the BanglaLekha-Isolated dataset and got 95.16% accuracy on a prepared dataset.

Classifier Used	Features / tools	Dataset	Reference	Accuracy
Convolutional Neural Network	SCNN, ECNN	___	Alizadehashraf, B. and Roohi, S., [24]	97.1%, 96.3%
	Siamese network, Triplet network	MNIST, OWN	Nawaz, S., Calefati, A., Ahmed, N. and Gallo, I. [15]	96.23, 98.23
	used kernel and local receptive fields	BanglaLekha-Isolated dataset	Purkaystha, B., Datta, T. and Islam, M. [18]	89.93%
	___	BanglaLekha-Isolated dataset, Prepared dataset	S M Azizul Hakim, Asaduzzaman [20]	99.44 %, 95.16%
	linear interpolation and a constant thickening factor	HP Labs India.	Kavitha B.R., Srimathi C[21]	95.16%, 97.7%
	Filter Matrix	___	Sonawane, P. and Shelke, S. [10]	validation 94.49% , Testing 95.46%

F. Tamil Characters

i) Kavitha B.R., Srimathi C [21] has proposed a model to recognize Tamil characters using CNN. The dataset contains 82,929 pictures in the tiff or png position created by HP Labs India. These pictures are gotten from the online form utilizing straightforward piecewise direct addition and a consistent thickening element. Pictures size standardized to 64x64 utilizing bilinear insertion method and scaled to 0, 1 territory. There are 9 layers in this CNN, while five Convolutionary layers, two max-pooling layers and two completely connected layers. The performance layer is a 156-class, softmax classifier. In this 156 unique Tamil characters are used. The model gives preparing precision is 95.16% and the testing exactness is 97.7%.

4. Algorithm and Findings

Table V Algorithm Wise Findings – A Comparison Table

Classifier Used	Features / tools	Dataset	Reference	Accuracy
Support Vector Machine	vertical projection, Convex hull	CCC benchmark dataset	Pritam Dhande, Reena Kharat [4]	___
	Freeman chain code	NIST dataset	Dewi Nasien, Habibollah Haron Sophiayati Yuhaniz [3]	86%, 88%, 73%
	graph distance theory	Chars74k numerals database	Sahare, P. and Dhok, S.B [22]	99.84%
Neural Network	Piecewise histogram of oriented gradients (HOG)	Nikita Singh [12]		97.06%

Table 2 Overall Findings

Author	Classifier /Algorithm	Languages	Database	Result (Accuracy)
Ghazanfar Latif et.al [23]	CNN	English, Arabic, Persian, Devanagari	Madbase, Mnist, Hoda, Pmudb, Dhcd	99.26%
Vinoj Jayasundara et.al [2]	CapsNets (Capsule Network)	English Character and digits	EMNIST-balanced, EMNIST-digits, and MNIST	90.46%
Pritam Dhande, Reena Kharat [4]	Convex hull algorithm and SVM	cursive English script	CCC benchmark dataset	-
Alizadehashraf, B. and Roohi, S., [24]	SCNN,ECNN	Persian	HODA	97.1%,96.3%
Supriana, I. and Nasution, A. [7]	Hilditch and Zidouri algorithm	Arabic	Take documents image from article titled "Kingdom of Saudi Arabia"-	48.3%

144

Madhuri Yadav, Dr. Ravindra Purwar [11]	Quadratic SVM, k-NN, weighted k-NN, Subspace discriminate, and Bagged trees,	Hindi	data from different authors	more than 93% and Quadratic SVM 96.6%
Deepika Gupta, Soumen [13]	traversing algorithm	Hindi Text	-	95.70%
Nikita Singh [12]	ANN and histogram of oriented gradients (HOG)	Devanagari	-	97.06%
Kavitha B.R., Srimathi C [21]	CNN	Tamil	HP labs India	Training 95.16% Testing 97.7%
S M Azizul Hakim, Asaduzzaman [20]	CNN	Bangla numerals and characters	Bangla Lekha-Isolated, Prepared dataset	99.44%, 95.16%

5. Conclusion

In this review exploration, Deep learning algorithm Convolutional Neural Network and other Neural Networks algorithms are proposed for the recognition of Hand Written Character (HCR) and offline Hand Written Character for the languages English, Arabic, Urdu, Devanagari, Bengali, etc. Here different algorithms applied for single language hand written character and got the better accuracy. In one the tool is used to get less accuracy. Presently a day's CNN is nearly utilized in each exploration which depends on pattern and image recognition. It has been shown that the higher the identification rate, the more the layers concealed. We plan to investigate a deep learning design for universal written by hand character recognition with better precision. This will be likewise extended for consolidated numerals and characters recognition in various languages too.

6. References

[1]   A. S. Abdalkafor, "Survey for Databases on Arabic Off-line Handwritten Characters Recognition System," 1st Int. Conf. Comput. Appl. Inf. Secur. ICCAIS 2018, pp. 1–6, 2018, doi: 10.1109/CAIS.2018.8442001.

[2]   V. Jayasundara, S. Jayasekara, H. Jayasekara, J. Rajasegaran, S. Seneviratne, and R. Rodrigo, "TextCaps: Handwritten character recognition with very small datasets," Proc. - 2019 IEEE Winter Conf. Appl. Comput. Vision, WACV 2019, pp. 254–262, 2019, doi: 10.1109/WACV.2019.00033.

[3]   D. Nasien, H. Haron, and S. S. Yuhaniz, "Support Vector Machine (SVM) for english handwritten character recognition," 2010 2nd Int. Conf. Comput. Eng. Appl. ICCEA 2010, vol. 1, pp. 249–252, 2010, doi: 10.1109/ICCEA.2010.56.

[4]   P. Dhande and R. Kharat, "Recognition of cursive English handwritten characters," Proc. - Int. Conf. Trends Electron. Informatics, ICEI 2017, vol. 2018-January, pp. 199–203, 2018, doi: 10.1109/ICOEI.2017.8300915.

[5]   M. H. Dave and A. P. M. Patel, "Handwritten English Character Recognition using Multilayer Perceptron Neural Network," no. 3, pp. 713–719, 2017.

[6]   H. Ghaleb, P. Nagabhushan, and U. Pal, "Segmentation of offline handwritten Arabic text," pp. 41–45, 2017, doi: 10.1109/asar.2017.8067757.

[7]   I. Supriana and A. Nasution, "Arabic Character Recognition System Development," Procedia Technol., vol. 11, no. Iceei, pp. 334–341, 2013, doi: 10.1016/j.protcy.2013.12.199.

[8]   A. El-sawy and M. Loey, "Arabic Handwritten Characters Recognition using Convolutional Neural Network," vol. 5, 2017.

[9]   O. A. Almansari, "Recognition of Isolated Handwritten Arabic Characters," 2019 7th Int. Conf. Mechatronics Eng., pp. 1–5, 2019, doi: 10.1109/ICOM47790.2019.8952035.

[10] P. K. Sonawane and S. Shelke, "Handwritten Devanagari Character Classification using Deep Learning.," 2018 Int. Conf. Information, Commun. Eng. Technol. ICICET 2018, pp. 1–4, 2018, doi: 10.1109/ICICET.2018.8533703.

[11] M. Yadav and R. Purwar, "Hindi handwritten character recognition using multiple classifiers," Proc. 7th Int. Conf. Conflu. 2017 Cloud Comput. Data Sci. Eng., pp. 149–154, 2017, doi: 10.1109/CONFLUENCE.2017.7943140.

[12] N. Singh, "An Efficient Approach for Handwritten Devanagari Character Recognition based on Artificial Neural Network," 2018 5th Int. Conf. Signal Process. Integr. Networks, SPIN 2018, pp. 894–897, 2018, doi: 10.1109/SPIN.2018.8474282.

[13] D. Gupta and S. Bag, "An Efficient Character Segmentation Approach for Handwritten Hindi Text," 2018 5th Int. Conf. Signal Process. Integr. Networks, SPIN 2018, pp. 730–734, 2018, doi: 10.1109/SPIN.2018.8474047.

[14] S. Hassan and A. Mirza, "Cursive Handwritten Text Recognition using Bi-Directional LSTMs : A case study on Urdu Handwriting," 2019 Int. Conf. Deep Learn. Mach. Learn. Emerg. Appl., pp. 67–72, doi: 10.1109/Deep-ML.2019.00021.

[15] S. Nawaz, A. Calefati, N. Ahmed, and I. Gallo, "Hand written characters recognition via deep metric learning," Proc. - 13th IAPR Int. Work. Doc. Anal. Syst. DAS 2018, pp. 417–422, 2018, doi: 10.1109/DAS.2018.18.

[16] S. Bin, A. Saeeda, S. Swati, and M. Imran, "Handwritten Urdu character recognition using one-dimensional BLSTM classifier," Neural Comput. Appl., no. 1, 2017, doi: 10.1007/s00521-017-3146-x.

[17] C. Saha, R. H. Faisal, and M. M. Rahman, "Bangla handwritten basic character recognition using deep convolutional neural network," 2019 Jt. 8th Int. Conf. Informatics, Electron. Vision, ICIEV 2019 3rd Int. Conf. Imaging, Vis. Pattern Recognition, icIVPR 2019 with Int. Conf. Act. Behav. Comput. ABC 2019, pp. 190–195, 2019, doi: 10.1109/ICIEV.2019.8858575.

[18] B. Purkaystha, T. Datta, and M. S. Islam, "Bengali handwritten character recognition using deep convolutional neural network," 20th Int. Conf. Comput. Inf. Technol. ICCIT 2017, vol. 2018-January, pp. 1–5, 2018, doi: 10.1109/ICCITECHN.2017.8281853.

[19] M. Al, R. Alif, and S. Ahmed, "Isolated Bangla Handwritten Character Recognition with Convolutional Neural Network," pp. 22–24, 2017.

[20] S. Azizul Hakim and Asaduzzaman, "Handwritten Bangla Numeral and Basic Character Recognition Using Deep Convolutional Neural Network," Int. Conf. Electr. Comput. Commun. Eng., p. 6, 2019.

[21] B. R. Kavitha and C. Srimathi, "Benchmarking on offline Handwritten Tamil Character Recognition using convolutional neural networks," J. King Saud Univ. - Comput. Inf. Sci., no. xxxx, 2019, doi: 10.1016/j.jksuci.2019.06.004.

[22] P. Sahare and S. B. Dhok, "Multilingual Character Segmentation and Recognition Schemes for Indian Document Images," IEEE Access, vol. 6, no. c, pp. 10603–10617, 2018, doi: 10.1109/ACCESS.2018.2795104.

[23] G. Latif, J. Alghazo, L. Alzubaidi, M. M. Naseer, and Y. Alghazo, "Deep Convolutional Neural Network for Recognition of Unified Multi-Language Handwritten Numerals," 2nd IEEE Int. Work. Arab. Deriv. Scr. Anal. Recognition, ASAR 2018, pp. 90–95, 2018, doi: 10.1109/ASAR.2018.8480289.

[24] Alizadehashraf, , B. and Roohi, S., 2017, November. Persian handwritten character recognition using convolutional neural network In 2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP) (pp. 247-251). IEEE.

Regionally Disparate Recovery Rates and Rising Innovations Amidst the Pandemic

Shreya Upadhyaya[1], Parul Kalra[2]and Deepti Mehrotra[3]

Amity University Uttar Pradesh

*shreya.upadhyaya1506@gmail.com[1], parulkalra18@gmail.com[2]

Abstract: *Large parts of the world have countered the impact of Covid -19 through rigorous research and technological innovations. The pandemic has made developing nations like India smarter in varied spheres leading us into an era of physical isolation but of mentalistic association. Though nations across the world have overcome the deadly waves of Coronavirus, India is still amidst the same making all efforts to alter and control the situation. The fight against the pandemic in India emphasizes upon the need to delve deeper into the positive aspects of the combat situation to ensure human survival which is contingent more on the mental strength and other external factors rather than just the physical strength of an individual. This makes our research focus on the regionally disparate recovery rates to provide an understanding that the variations in salient steps taken by different states and nations in terms of physical isolation, depth of research and innovation (R&I), early detection, physical assistance and mental support to afflicted ones has led to great differences in recovery rates. It clarifies that the canvass of recovery doesn't follow a theme of complete darkness rather it provides hope to those who are still combating its implications.*

Keywords: *Machine learning, disparate recovery, mentalistic association*

1. Introduction

The COVID-19 pandemic has unfolded as a cataclysm in different phases varying in terms of severity, symptoms, rate of recovery, sphere of Influence, morbidity burden and strains as agents of attack. The afflicted nations witnessed their masses crippling in both economic and social spheres while going through a potential loss of Humans as a resource. The severity of COVID-19 aggravated further in the wake of second wave of coronavirus thatcrippled the largest democracy of the world -India, at a stage when the world is set to undergo a Global Demographic Transition (GDT) while India marched steadily to enter in the 37- years old surging phase of working age population.[1] This necessitates us to understand the various factors which triggered the disparate recoveries in different regions, revealing specific infirmities inherent in the system that are rarely measured with accuracy and precision. Further, this also led to the "COVID-19 triggered revolution" which gave India a smarter edge, Multi-Dimensional Vital Expertise (MDiVE) and distinguished, experiential learning in areas of education, research and innovation(R&I). The current levels of growth in such areas have

shown the power of Need for Discovery (NfD) which can have far-reaching and a revolutionizing impact in the decades ahead. [5,6]

Further, the pandemic conditioned the humans to develop Personally- Characterized Precautious Behavior (PCPB) patterned by individuals to counter the spread of COVID-19 and the evolution of Personally- Characterized Combat Behavior (PCCB) developed by individuals battling against the fatal implications of COVID-19. A broader generalization of these characteristics highlighted the salience of physical isolation, preference to virtual mode of working, enriched dietary practices, induced social withdrawal, masked interaction and sanitized hospitality. But all these failed to invert the peaking fatality curve as broken mentalistic associations aggravated the problems of the infected individuals. This accentuated the importance of mental health and its inherent linkages to physical health. [2,3]

For the study, machine learning was preferred to discover the significant attributes and obtain the results. This paper is organized as follows, section 2 deals with the description of methodology-based flow, section 3 deals withResults and discussion, section 4 deals with conclusion.

2. Methodology

The Covid-19 data containing salient attributes relevant to the focus of research was pre-processed in steps sequentially proceeding from importing the dataset, filling missing values, encoding categorical data, splitting the data into two datasets-test and the training, finally scaling the features to prevent anomalies in prediction. [4,7] Then, the supervised and unsupervised machine learning (ML) algorithms were applied to obtain useful results.[4] Clustering was applied prior to category based encoding, data splitting and feature scaling while the regression and classification was applied later. The cross-validation of the results obtained on two datasets was done and the output was found relevant and accurate. The methodological framework is given below in Fig.1.

The growing mortality burden suggests the need to understand the confound relationships existing among different attributes necessary to make important predictions in regard to the massive impact of Covid-19.[8]

Cluster Dendrogram

Figure 2.represents state recovery- based cluster dendrogram.

distance
hclust (*, "complete")
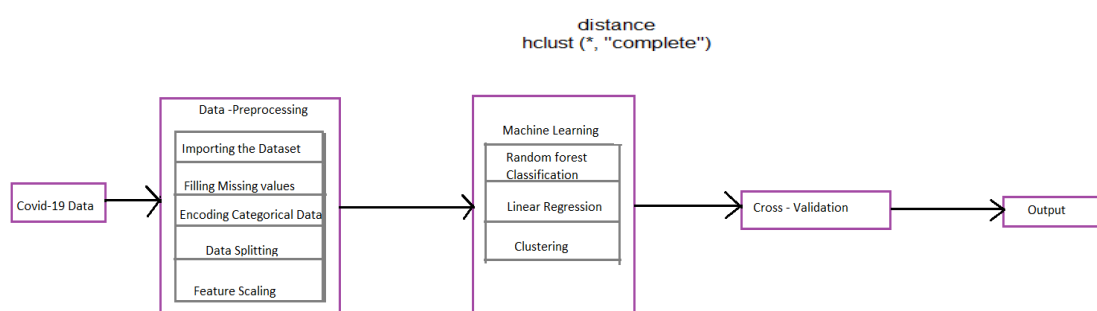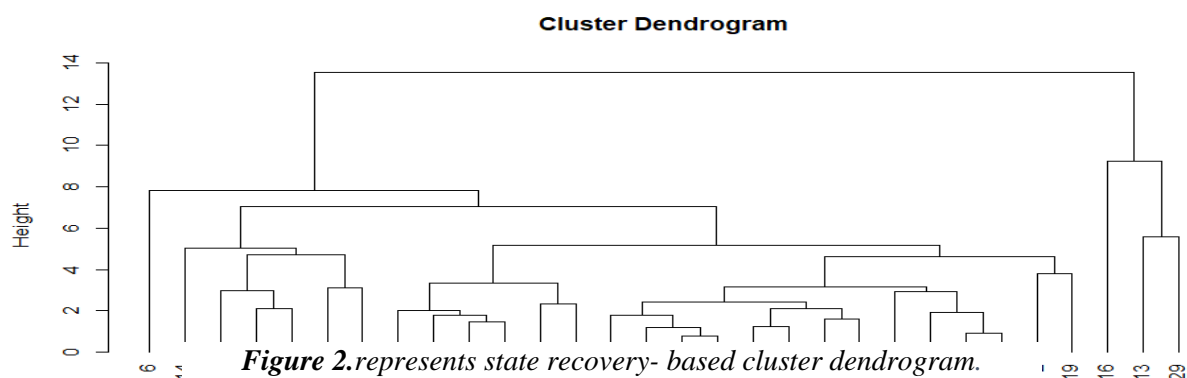
Figure 3. represents the flow of methodology.

\

3. Results and Discussion

The study schematically presents a COVID-19 recovery-based clustering and regression which revealed the existence of mentalistic association and embedded psychological factors that played a significant role in the recovery of infected individuals. Since, the pandemic has been discussed often, more as a phase of mortality rather than recovery, this made our research focus the frame of working on recovery rate. The results proved that the recovery is significantly related to health sector manpower, health infrastructure, total cured cases but not on population and percentage literacy of the state.[9,10] The significant metrics obtained are:

Table 1. The metrics show the significance of health infrastructure in recovery.

Metrics	Obtained Values
Estimate	6.639
Standard Error	1.418
t-Value	4.682
p-value	6.61 e^{-05}

149

Table 2. *The metrics show the significance of health sector manpower in recovery.*

Metrics	Obtained Values
Estimate	23951.399
Standard Error	5217.244
t-Value	4.591
p-value	0.0002446

Table 3. *The metrics for the state wise - population in recovery.*

Metrics	Obtained Values
Estimate	75.99
Standard Error	85.65
t-Value	0.887
p-value	0.3825

Table 4. *The metrics for the percentage literacy in recovery.*

Metrics	Obtained Values
Estimate	13857
Standard Error	23312
t-Value	0.594
p-value	0.557

Table 5. *The metrics for Random forest classification based on recovery rates.*

Metrics	Obtained Values
Number of trees	500
m-try	6
OOB estimate of Error rate (percentage)	26.67

4. Conclusion

The study was able to extract the significant attributes that led to the regionally disparate recovery of India from COVID-19. This suggested that India needs to expand its health infrastructure, boost the manpower in health sector in terms of number which can significantly improve the potential capacity of our nation in fighting the pandemic and other such disastrous events in a better and efficient way. Further, the innovations associated to the COVID-19

pandemic termed as "coronovation" and theresearches therein, termed "coviresearch" will be discussed in the extended paper.

References

1. Li, W. T., Ma, J., Shende, N., Castaneda, G., Chakladar, J., Tsai, J. C., &Ongkeko, W. M. (2020). Using machine learning of clinical data to diagnose COVID-19: a systematic review and meta-analysis. *BMC medical informatics and decision making*, *20*(1), 1-13.

2. Alballa, N., & Al-Turaiki, I. (2021). Machine learning approaches in COVID-19 diagnosis,

3. mortality, and severity risk prediction: A review. *Informatics in Medicine Unlocked*, 100564.

4. Huang, S., Yang, J., Fong, S., & Zhao, Q. (2021). Artificial intelligence in the diagnosis of COVID-19: challenges and perspectives. *International Journal of Biological Sciences*, *17*(6), 1581.

5. Hussein, H. A., & Abdulazeez, A. M. (2021). COVID-19 PANDEMIC DATASETS BASED ON MACHINE LEARNING CLUSTERING ALGORITHMS: A REVIEW. *PalArch's Journal of Archaeology of Egypt/Egyptology*, *18*(4), 2672-2700.

6. Moezzi, M., Shirbandi, K., Shahvandi, H. K., Arjmand, B., & Rahim, F. (2021). The diagnostic accuracy of Artificial Intelligence-

7. Assisted CT imaging in COVID-19 disease: A systematic review and meta-analysis. *Informatics in medicine unlocked*, 100591.

8. Deng, C. X. (2021). The continued global battle against SARS-CoV-2 and COVID-19. *International Journal of Biological Sciences*, *17*(6), 1440.

9. Tayarani-N, M. H. (2020). Applications of artificial intelligence in battling against Covid-19: a literature review. *Chaos, Solitons & Fractals*, 110338.

10. Li, W. T., Ma, J., Shende, N., Castaneda, G., Chakladar, J., Tsai, J. C., ... &Ongkeko, W. M. (2020). Using Machine Learning of Clinical Data to Diagnose COVID-19.

Gender Prediction from Facial Images: A Comparative Analysis of Classifiers

Rekha Gope and Mudassir Rafi[*]

Department of Computer Applications, NIT, Jamshedpur, India

*mudassir.ca@nitjsr.ac.in

Abstract: *Automatic prediction of gender has recently gained a lot of attention. With social media and various free platforms, the problem has become much more relevant. In the present paper, we have presented, an automatic gender classification problem from frontal facial images. The features have been extracted by using Local binary pattern and Gray level co-occurrence matrix. The extracted features are applied to a series of classifiers namely, k nearest neighbor (instance based), Support Vector machine (statistical based) and neural networks (connectionist based) algorithms. The results have been compared and reported in the form of a comparative analysis. We have also identified, the best and worst classifier working for the respective problem.*

Keywords: *Gender Prediction, LBP, GLCM, Classification*

1. Introduction

The identification of gender plays a pivotal role during social interactions. With the advent of social media, the problem of automatic gender classification has become much more relevant not only as a means of social interaction but it has also been extended to further applications including visual surveillance, human-computer interaction, biometrics, demographic data collection, security surveillance system and recently as a pre-processing step in face recognition systems. For human visual system (HVS), the identification of gender is very obvious and precise but the same task proves to be one of the challenging tasks for a machine. Thus, recently gender prediction and identification has gained much interest in computer vision and machine learning community.

In the present paper, the problem of gender classification has been presented as a binary classification problem. In this task, the facial images are taken as input, to be identified as one of the two classes either a male or a female. In this task, the third gender has not been included intentionally, because of the unavailability of the corresponding image dataset. Feature extraction and classification can be used to solve the problem of similar classification task. Feature extraction is the process which involves the automatic extraction of the facial features, and then it expresses the result as aim of the features. Selecting the feature is one of the most challenging tasks as its goal is to select the optimal subset of feature that represent the face in effective and robust manner.

The paper has been organized into five sections: In the first section a formal definition and the problem of gender detection has been given, after which a brief summary of the work published has been given in the Related work. In section 3, the two feature descriptors have been described, in section 4, the classifiers used has been given. Section 5, contains the experimental steps and a summarization of the obtained results. At last paper ends with a conclusive summary.
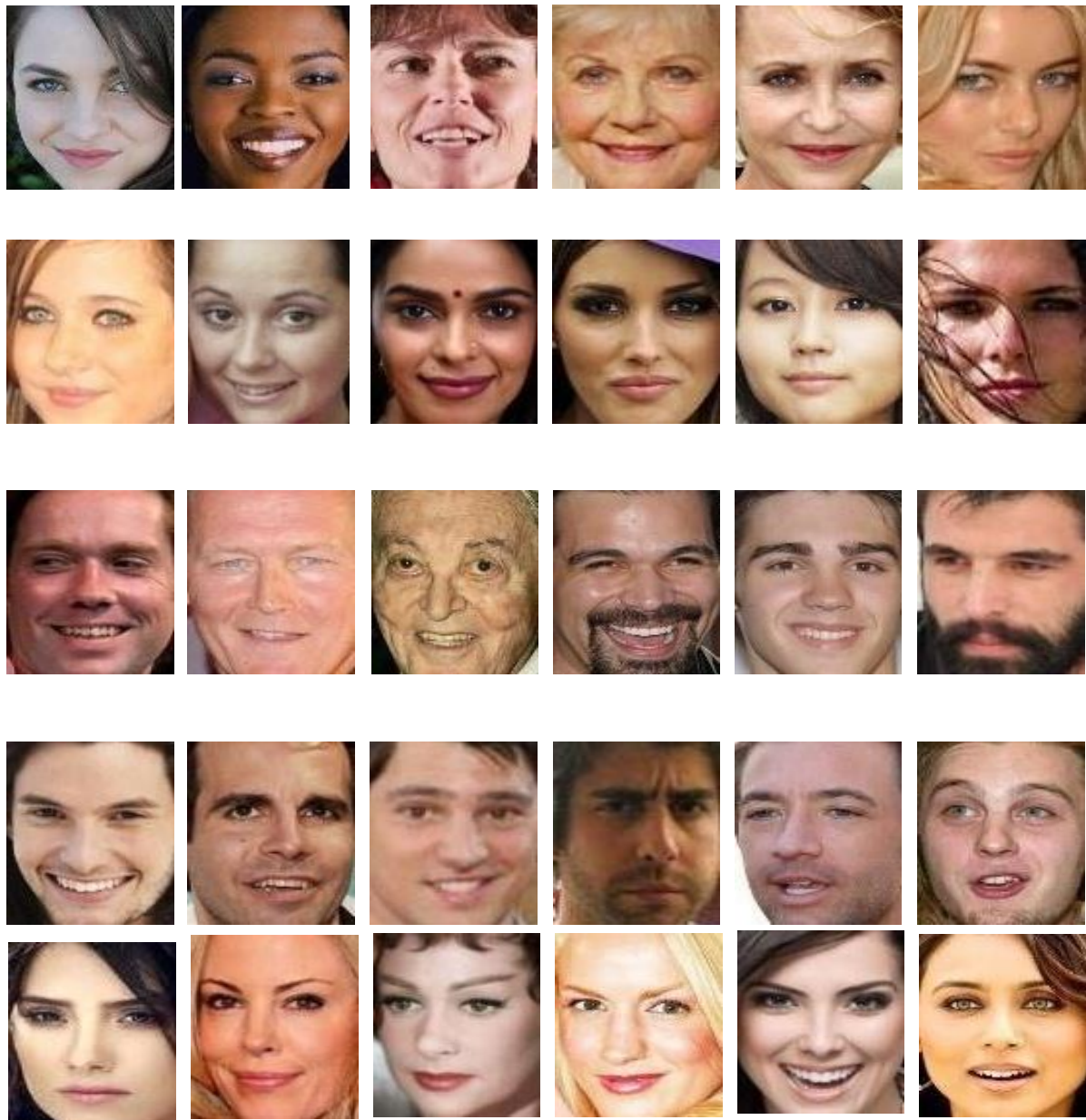
Fig.1 Samples from Training Dataset

2. Related Work

In 1990's computer vision-based techniques were used first time for gender classification. Since then, many such techniques have been reported in the literature. These techniques were found to be differ either in the representation of facial features employing direct gray values to more complex features or on the use of classifier ranging from nearest neighbor classifier to Fisher Discriminant, neural network, SVM and Adaboost classifiers. Moghaddam and Yang [1] was the first to use the direct pixel values into SVM for classification purpose whereas Baluja and Rowley [2] employed AdaBoost classifier to combine the weak classifiers and convert them into the strong classifier. For this, they have used comparison of pixels. In [3], a survey related to gender classification can be obtained. Representation of faces in terms of geometrical features Bruneli and Poggio [4] proposed a geometrical method for identification of genders. They have devised a set of 16

geometrical features from frontal face images and having no facial hair. At last, the representation has been used to train two competing hyper basis function networks; one for male and one for female. Colomb [5] has used a fully connected back propagation network. The images were resized as per convenience and rotated to match the position of eye and mouth in each image. The output of the hidden unit has been used as input to final neural network which produces values 1 for male and 0 for female. Tolba [6] has proposed gender identification using various neural network classifiers. In the back propagation network, the images are scaled to 30 x 30 size images and rotated in the way that the position of eyes and mouth are similar in every image. The output of the hidden unit can be used as input for final neural network which represent 1 for male and 0 for female. For gender identification, two different neural network classifiers can be used i.e., radial basis function network (RBF) and learning vector quantization (LVQ). Baback [7] has used SVM for low resolution images and compared its performance in the same database with other classifiers.

3. Feature Descriptors

Local Binary Pattern

Local Binary Patterns is a simple and very efficient feature descriptor. It labels the pixels of an image by thresholding the neighborhood of each pixel and the result can be considered as a binary number. It was introduced in 1994 and has been found to be a powerful feature for specifically for texture

classification. It has also been used widely for face detection in many of the novel works. The procedure of the LBP has been shown in the image given below.
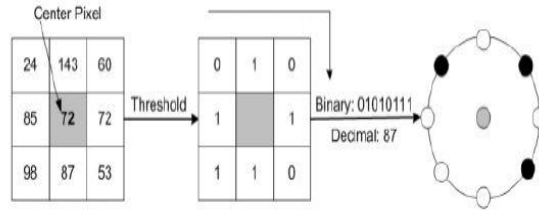
Fig. 3 *Generalized LBP operator*

In Fig. 3, the LBP for the central pixel can be computed by considering the threshold value as 72, which is the value of the central pixel. All the values have been thresholded accordingly and the pattern has been started to form from upper left topmost value. After computing the corresponding LBP values, a histogram of these values can also be formed and can be used as a feature descriptor. The generalized LBP function has been given in MATLAB, and has been employed for experiment purpose in its standard form. As the dataset used has all the images in RGB, so we have used only one channel instead of all the three-color channels.

Gray Level Co-occurrence Matrix (GLCM)

A gray-level co-occurrence matrix [8] considers both brightness information along with spatial arrangement. Mathematically, consider a matrix G whose elements $f(i,j)$ are representing the frequencies of pixel pairs with intensity values z_i and z_j occurring in the image $g(r,c)$ at the location specified by Q, where Q is an operator defining the positions of the two pixels relative to each other and $1 \leq i,j \leq L$. The matrix dimension depends upon the number of gray levels present in the image. Conventionally, the number of gray levels are quantized in order to reduce the size of the matrix. After computing all the values of the matrix, a normalization step is applied using the equation as given below:

$$N_{(p,q)} = \frac{f(p,q)}{\sum_{i=0}^{L} \sum_{j=0}^{L} f(i,j)}, p,q = 0,\dots L$$

where $N_{(p,q)}$ is the normalized matrix, $f(p,q)$ is the number of pixel pairs having values z_i and z_j, $f(p,q)$ is the entry in the co-occurrence matrix and L is the maximum number of gray levels. Haralick et al. [15] proposed a set of 14 statistical measures based on the co-occurrence matrix. These include angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, measures of correlation and maximal correlation coefficient. Out of these

14 measures in this work authors have employed 4 because of the computational overhead, these are Contrast, Correlation, Energy and Homogeneity.

Contrast is a quantitative measure of the intensity contrast between a pixel and its neighbour over the whole size of the image. The values are ranges from 0 to $(L-1)^2$.

$$Contrast = \sum_{i=1}^{L}\sum_{j=1}^{L}(i-j)^2 \, N_{(i,j)}$$

Correlation is a measure of how a pixel is correlated to its neighbor over the entire size of the image. Values are ranges from -1 to +1. If the correlation is perfect +1 is assigned whereas -1 is used for a perfect negative correlation.

Correlation $= \sum_{i=1}^{L}\sum_{j=1}^{L}\frac{(i-m_r)(j-m_r)}{\sigma_r \sigma_c}$ where $\sigma_r \neq 0$, $\sigma_c \neq 0$ and

$$m_r = \sum_{i=1}^{L} i \sum_{j=1}^{L} N_{(i,j)}$$

$$m_c = \sum_{i=1}^{L} j \sum_{j=1}^{L} N_{(i,j)}$$

Energy is a measure of regularity and lies in the range [0,1]. It is 1 for a constant image.

$$Energy = \sum_{i=1}^{L}\sum_{j=1}^{L} N_{(i,j)}^2$$

Homogeneity measures that how much the distribution of elements in the matrix G are closer to its diagonal. The range of values lies in r0; 1s, with the highest being reached when G is a diagonal matrix.

$$Homogeneity = \sum_{i=1}^{L}\sum_{j=1}^{L} N_{(i,j)}^2$$

4. Classifiers

Nearest Neighbor It is an instance-based learning algorithm. It simply stores the training examples and no description of target function is explicitly constructed. It assumes that every instance is a poin in n-dimensional space and nearest neighbors are taken by computing Euclidean distance.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n}\left(a_r(x_i) - a_r(x_j)\right)^2}$$

Table 1. Classification Results

Classifiers	GLCM			LBP		
	Training (%)	Testing (%)	Cross Validation (%)	Training (%)	Testing (%)	Cross Validation (%)
Nearest Neighbor (k=1)	100	58.9	58.6	100	61.87	60.31
Nearest Neighbor (k=2)	100	58.3	58.1	100	61.43	60.04
Nearest Neighbor (k=3)	100	61.4	61.3	100	65.05	63.51
AdaBoost M	100	62.14	62.7	100	62.66	63.22
Neural Network	100	65.2	64.0	100	71.91	70.77
Random Forest	100	65.3	65.0	100	71.56	70.31
SVM	100	63.0	63.0	100	70.73	70.13

where an instance x is defined by $1 < a_1(x), a_2(x), \dots a_n(x) >$. The algorithm is simple and supervised in nature. It is used to solve both classification and regression problems. In the current work, we have used with three different values of k, starting from 1 to 3.

AdaboostM is a boosting algorithm proposed in 1996 and converts a set of week classifier in strong one. It combines the classifiers with the training set of every iteration. The weightage of each trained classifier in the iteration depends upon the accuracy. The equation for classification can be represented as

$$F(x) = Sign\left(\sum_{m=1}^{M} \theta_m f_m(x)\right)$$

For experimental purpose we have used the adaboost with its standard setting as given in Weka3.8.5.

Neural Network are a family of classifiers used predominantly in machine learning and cognitive science. These are inspired by biological neural network and employed for approximating those functions that might be depend on a large number of inputs. These are generally organized in layers. Layers are consisting of nodes connected to each other. There

may be one or more hidden layers. The hidden layers are in turn connected to the output layer. In the present work, we have used multilayer perceptron with one hidden layer.
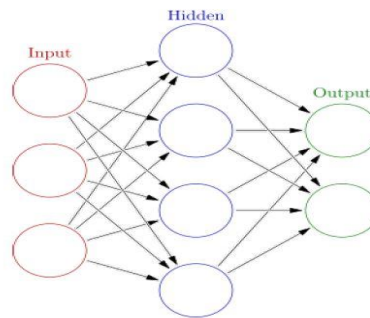
Fig. 4 Neural Network

Random Forest is a popular machine learning algorithm and belongs to the class of supervised learning algorithms. It can be used for both Classification and Regression problems in ML and is based upon the concept of ensemble learning. It uses a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting. There are few advantages of using random forest, first of all, It takes less training time as compared to other algorithms and predicts output with high accuracy, even for the large dataset it runs efficiently. Lastly, It can also maintain accuracy for a large proportion of missing data. In the present work, Random forest has been used with standard setting provided in weka.

Support Vector Machine (SVM) The main objective of support vector machine is to find a best separating hyperplane in n-dimensional space where n represents number of features. It is mainly a binary classifier and used the margins from support vectors. IN SVM classification, we give some training data D, a set of n points of the form

$$D = \left\{(x_i, y_i) \mid x_i \in R^P, y \in \{-1,1\}\right\}_{i=1}^{n}$$

Where y_i represents the class to which point x_i belongs. x_i is a p dimensional real vector. In weka, it is given as SMO, which stands for Sequential Minimal Optimization. It is used for the training of support vector machines on a dataset and used to solve the quadratic programming problem arises during the training of SVM, as the methods for SVM training are more complex. It can handle a very large training set, because the memory required for SMO is linear to the size of the training set. In the current work, SMO has been used with standard parameters given in freely available weka 3.8.5.

5. Experimental Results and Discussions

The present section describes the experimental steps employed for feature extraction, classification and the results obtained. As the current work classifies into resulting features thus obtained have been applied to numerous classifiers. All the experimental works have been evaluated and executed using MATLAB-2016, running on intel core i-5 processor with 4 GB RAM. The LBP and GLCM, has been used with standard setting as given in MATLAB. For GLCM we have used only 4 Haralick features instead of 14 because of the computational simplicity. The four features have been computed for two offset values [2 0] and [0 2] respectively. The resulting features has been used simultaneously for classification purpose. Thus, a set of eight features have been used. As both the feature descriptors used works for gray level images, instead of converting the complete data set into corresponding gray images we have used only the first channel of RGB images. As the conversion might further compress the images and one channel out of the three have nearly the same information as the gray images.The dataset[9] has been downloaded from Kagle as it has more than 40,000 images having persons from various ethnicities. The following steps have been employed for carrying out the experiments:

1. LBP and GLCM descriptors are applied on all the images present in the dataset, first on training set then on validation set, the results thus obtained have been stored in the form of datasheet.

2. Each row of the datasheet represents one face image of the dataset, in this way we have one data sheet for training and one for testing, the last attribute represents for male and female. For experimental purpose, the female has been represented as 1 whereas the male has been denoted by 0.

3. Attributes in the datasheets have been normalized so that no single feature dominates during the classification phase.

4. The results have been randomized for an unbiased classification performance.

5. The two files corresponding to training and validation sets, for each descriptor namely LBP and GLCM, first of all has been converted to .arff files which is the only format to be uploaded in weka.

6. For each classifier, first the results for 10 folds cross validation have been obtained by using training dataset then the same dataset has been used to train the respective

classifier, after successful formation of the model the results have been obtained and reported.

7. In table 1, the classification results obtained by the classifiers has been mentioned. The classification percentage, varies from 58 to 70 percent. For, all the classifiers applied, the classification results are found to be a little bit large for LBP rather than GLCM, however, the results for both the descriptors are not encouraging at all. As cross validation takes the average of all the 10 folds, thus the classification results are less in comparison with testing.

6. Conclusion

With the advancement in social media and free platforms, the automatic identification of gender has gained much popularity. In the present paper, we have studied the classification accuracy of various classifiers for the features extracted from the two simple and popular descriptors namely Local binary pattern (LBP) and Gray level co-occurrence matrix (GLCM). The extracted features are applied to a series of classifiers namely, k nearest neighbor, Adaboost, Neural network, Random Forest and Support Vector Machine. The classification accuracies are found to be very less however for neural network-based classifiers, the classification percentage is a little bit higher but it is too less to be used for practical purposes. Thus, LBP and GLCM are not the appropriate descriptors to be employed for feature extraction. Moreover, the application of segmentation before feature extraction might increase the final classification percentage.

Reference:

[1] Moghaddam, B., & Yang, M. H. (2002). Learning gender with support faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), 707-711.

[2] Baluja, S., & Rowley, H. A. (2007). Boosting sex identification performance. *International Journal of computer vision*, *71*(1), 111-119.

[3] Makinen, E., & Raisamo, R. (2008). Evaluation of gender classification methods with automatically detected and aligned faces. IEEE transactions on pattern analysis and machine intelligence, 30(3), 541-547.

[4] Brunelli, R., & Poggio, T. A. (1991, January). HyperBF Networks for Real Object Recognition. In IJCAI (pp. 1278-1285).

[5] Golomb, B. A., Lawrence, D. T., & Sejnowski, T. J. (1990, November). SEXNET: A Neural Network Identifies Sex From Human Faces. In NIPS (Vol. 1, p. 2).

[6] Tolba, A. S. (2001). Invariant gender identification. Digital Signal Processing, 11(3), 222-240.

[7] Moghaddam, B., & Yang, M. H. (2000, March). Gender classification with support vector machines. In Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580) (pp. 306-311). IEEE.

[8] Rafi, M., & Mukhopadhyay, S. (2018). Texture description using multi-scale morphological GLCM. Multimedia Tools and Applications, 77(23), 30505-30532.

[9] https://www.kaggle.com/cashutosh/gender-classification-dataset

Batch Processing on Spatial and Temporal-DBSCAN for Big Data

Syed Abdul Malik[1] and Pradeep Kumar [2]

Maulana Azad National Urdu University, Hyderabad, India

*sd.abd54@gmail.com[1], pradeep@manuu.edu.in[2]

Abstract: *Clustering is an unsupervised learning process, as there is no target variable. In clustering, the data points of similar features try to group together in one cluster and dissimilar data points in other clusters. The objective of clustering is to create different grouped of similar data points. Spatial-Temporal clustering could be used in many fields such as, for analyzing geographical location with respect to time, traffic analysis of junctions, human mobility & crime analysis and medical imaging. As there are number of clustering algorithm available to use, but there are very few algorithms that can perform clustering well on spatial and temporal data. However, these algorithms do not achieve clusters on large data. As the data size augments the computational cost also rises, if the data cannot fit into primary memory, it throws memory error. This research proposes an algorithm to cluster large data of space-time features. Data used in this study consists of location (latitude and longitude) and timestamp. The algorithm has been implemented in Python and evaluates the clusters by plotting. The results demonstrate that the batch processing method well clustered the large spatial temporal data.*

Keywords: *Batch processing, Clustering, Spatial-temporal*

1. Introduction

Clustering is an unsupervised learning process, as there is no target variable. In clustering, the data points of similar features try to group together in one clusters and dissimilar data points in other clusters. The objective of clustering is to create different grouped of similar data points. Spatial-Temporal clustering could be used in many fields such as, for analyzing geographical location with respect to time, traffic analysis of junctions, human mobility & crime analysis and medical imaging.

Spatial-Temporal data denote spatial data which is collected and stored with respect to time series data. Spatial-Temporal data is increasing rapidly in the big data era, as the availability of huge spatial-temporal datasets such as remote sensing images, geographical locations changing with time and climate change data.

The algorithm ST-DBSCAN, that is based on density based algorithm that is Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [1]. DBSCAN relying on a density based idea of clusters which is designed to find out clusters of inconsistent shape with noise points. DBSCAN algorithm requires two input parameter Eps and MinPoints. Eps is

neighborhood of a point or radius of the clusters and MinPoints are the least number of data points or density in each clusters and assists the user in deciding a suitable value for it [1]. Each cluster is created with data points above a certain threshold. In ST-DBSCAN, it requires one additional parameter Eps2 that is for the computation of the resemblance of non-spatial values [2]. These two algorithms do not require to be given number of clusters. Few researchers have proven the ability of DBSCAN algorithm in processing very large databases [1, 3, 5]. The ST-DBSCAN algorithm works well, creates clusters based on spatial and temporal data with few thousands of values. However, it cannot handle more than tens of thousands of values. When we run it with few millions of values (rows) it throws memory error. As the data size gets enlarge, the computational cost also accelerates.

Our algorithm requires all the parameters of the ST-DBSCAN algorithm and one extra parameter which is batch size must be given. We have improved ST-DBSCAN algorithm only for processing large data. The proposed technique used batch processing to solve the problem of memory error and cluster data based on ST-DBSCAN algorithm.

The whole paper is divided into 5 segments. They are as follows. Segment 2 gives explanation about the fundamental ideas of density based clustering algorithm along with spatial-temporal density based clustering algorithm. Segment 3 outlines the limitations of existing ST-DBSCAN algorithm and our proposed solution. Segment 4 demonstrates our algorithm in detail and its implementation. Segment 5 deals with conclusion and future improvement.

2. Fundamental Ideas of Density Based Algorithms

This section deals with basics concept of density based clustering algorithm and ST-DBSCAN algorithm.

Density based clustering refers to unsupervised learning algorithm in which data are clustered or grouped together based on the high data point density. The DBSCAN algorithm is very essential in finding non-linear shapes of clusters based on the high density of data points and it can also separate the outlier or noise from the clusters. It takes two parameters: one is radius value Eps (ε), radius can be defined by user depending on the domain knowledge and MinPts. MinPts is the value of the least number of data points to be considered as a cluster. The DBSCAN algorithm uses the idea of **density reachability** and **density connectivity.**

Core Object: It refers to a data points inside of each cluster within a given radius (ε).

Border Object: A data point on the border of each cluster. If the data point is not a core object and it is density-reachable from another core object of other clusters.

Density reachability: "A point p is *density-reachable* from a point q wrt. Eps and MinPts if there is a chain of points Pl,…,Pn, Pl = q, Pn = P such that Pi+l is directly density-reachable from Pi. Density-reachability is a canonical extension of direct density-reachability. This relation is transitive, but it is not symmetric" [12]. Figure 1 "depicts the relations of some sample points and, in particular, the asymmetric case. Although not symmetric in general, it is obvious that density-reachability is symmetric for core points. Two border points of the same cluster C are possibly not density reachable from each other because the core point condition might not hold for both of them. However, there must be a core point in C from which both border points of C are density-reachable. Therefore, we introduce the notion of density-connectivity which covers this relation of border points". [12]

Density connectivity: "A point p is *density connected* to a point q wrt. Eps and MinPts if there is a point o such that both, p and q are density-reachable from o wrt. Eps and MinPts.

Density-connectivity is a symmetric relation. For density reachable points, the relation of density-connectivity is also reflexive. (shown in Figure 1)

Now, we are able to define our density-based notion of a cluster. Intuitively, a cluster is defined to be a set of density connected points which is maximal wrt. density-reachability. Noise will be defined relative to a given set of clusters. Noise is simply the set of points in D not belonging to any of its clusters." [12]
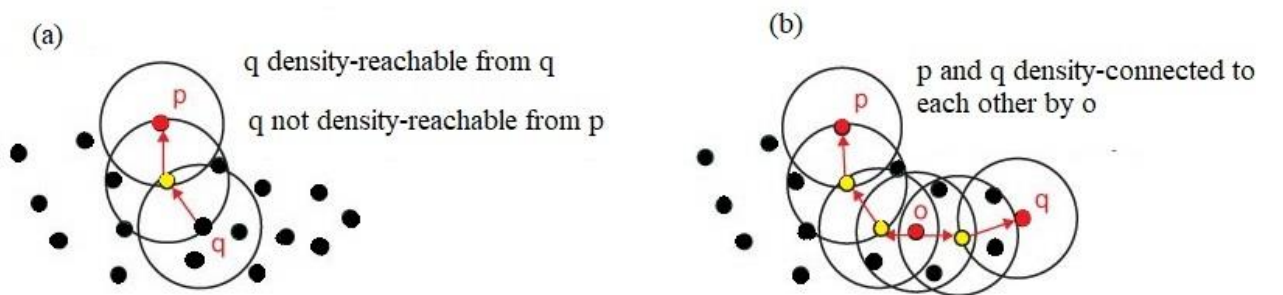
Fig. 1. (a) Density reachability and (b) Density connectivity

The algorithm commences with the first data object p and within Eps distance, it fetches all neighbors of p. If p is a core object and the total numbers of fetched data points are more than the MinPts, a cluster is formed. The adjacent points and the object pare put inside the newly

created cluster. After that it collects all the adjoining points within ε distance from the core point. Above process is rerun for the data points.

ST-DBSCAN means spatial-temporal DBSCAN. It works similar to DBSCAN with few differences that it clustered data based on spatial and temporal features. It obligatory to pass four parameters ε1, ε2, MinPts, Δε. ε1 is the distance between spatial parameters and ε2 is the distance between non-spatial parameters. It uses Euclidean, Manhattan or Minkowsski distance metric to calculate distance of spatial and non-spatial features. MinPts is the minimal number of data points or objects within ε1 and ε2. "The ST-DBSCAN algorithm begins with the foremost point p in dataset or database. The algorithm fetches all the data points which are density-reachable from p wrt. E1 and ε2. If p becomes a core point, then a new cluster is generated. If p becomes a border object, no points will be density-reachable from p. The algorithm processes the adjacent data points of the dataset. The process is rerun until all the data points have been processed". [2].

There are other algorithms that include temporal information in their methods. Wang et al [6] uses spatial and temporal information into two different distance computations to define the neighborhoods that create the clusters. Agarwal [7] also uses identical method in their algorithm for space-time extension of the OPTICS.

In few other algorithms, the spatial distance is given more weight by the temporal features and there are some methods which combines both spatial and non-spatial or temporal features to form one spatial-temporal distance [8,9]. Andrienko and Andrienko method [10] converts temporal or non-spatial feature into a uniform spatial distance, after that Euclidean distance is used to integrate the both spatial and temporal features. Oliveira [11] adopted the shared-nearest-neighbor(SNN) density-based clustering method to combine spatial-temporal and non-spatial data by using a weighted union of the three features (SNN-4D+) [12].

3. Limitation of Existing Algorithm

The limitation of the ST-DBSCAN algorithm is the main memory run out of space when the data size increases. When we run the algorithm on a very large dataset of few millions of rows it throws memory error because the computation cost of the algorithm increases and due to memory complexity. If the datasets fit in the main memory, it will not face any issues. When the data size increases gradually and the main memory capacity is not increased then memory complexity occurs and also computational cost increases.

4. Proposed Technique and Results

The technique which we use to solve the memory complexity issue is the batch processing technique. The proposed algorithm requires all the parameters of ST-DBSCAN algorithm with an extra argument of batch size. The batch size should be provided to the proposed algorithm for computation.

a. Description of the algorithm:

The algorithm requires 5 input parameters, four parameters which are same as ST-DBSCAN are Eps1, Eps2, MinPts and $\Delta\varepsilon$ [2]. And one extra parameter is batch size. In Fig.2, 'D' is the set of points or data. Eps1 is used to calculate the geographical distance and Eps2 for temporal distance value. The MinPts are the maximum number of data points must be within Eps1($\varepsilon1$) and Eps2($\varepsilon2$). $\Delta\varepsilon$ is the cut-off value to prevent clusters from combining each other. The last one is the batch size. The batch size is the maximum number of rows or values in a batch to process it at once.

In this technique, firstly, all the data is sorted based upon the spatial and temporal features. Then the data is split into different size of chunks, so that the chunk or batch of data can fit into main memory for computation. The total batches depend upon the batch size; as the data size grows the number of batches also rises.

ST-DBSCAN algorithm is applied on each batch of data, the cluster output is store in an array 'arr' then calculate the centroid of each cluster and store it in another array 'arr_centroid'. Again the ST-DBSCAN algorithm is applied on the array 'arr_centroid', the output of the algorithm is mapped to arr_centroid and 'arr' to assign all the data points to their respective clusters. And the last step is to assign new cluster label to each unique cluster.

Algorithm
batch_process_stdbscan (D, Eps1, Eps2, MinPts, Δε, batch_size)
    Inputs: the first four parameters are of that STDBSCAN algorithm
    D: data set
    Eps1: geographical or spatial distance value
    Eps2: non-spatial distance value
    MinPts: Maximum number of points within Eps1 and Eps2
    Δε: threshold value to prevent clusters from combining
    Batch_size: Maximum number of rows or values in a batch

    Output: set of clusters C

    Sort data by spatial and non-spatial values

    total_batches = Total length of D/batch_size

    arr = []
    for i from 1 to total_batches
        applied stdbscan_algorithm on each batch
        clust_batch= stdbscan_algorithm (D, Eps1, Eps2, MinPts, Δε)
        add clust_batch to arr

    arr_centroid = []
    for i from 1 to length.arr
        cent_i=calculate the centroid of i
        add cent_i to arr_centroid

    D1=convert arr_centroid to structured data set

    Apply: stdbscan_algorithm (D1, Eps1, Eps2, MinPts, Δε)

    C=store all the new clusters of D1

    Map C with arr_centroid and arr

    Assign new clusters id to each point

    END

Fig.2

b. Results:

Fig.4

Fig.5

Fig.6

Fig.7
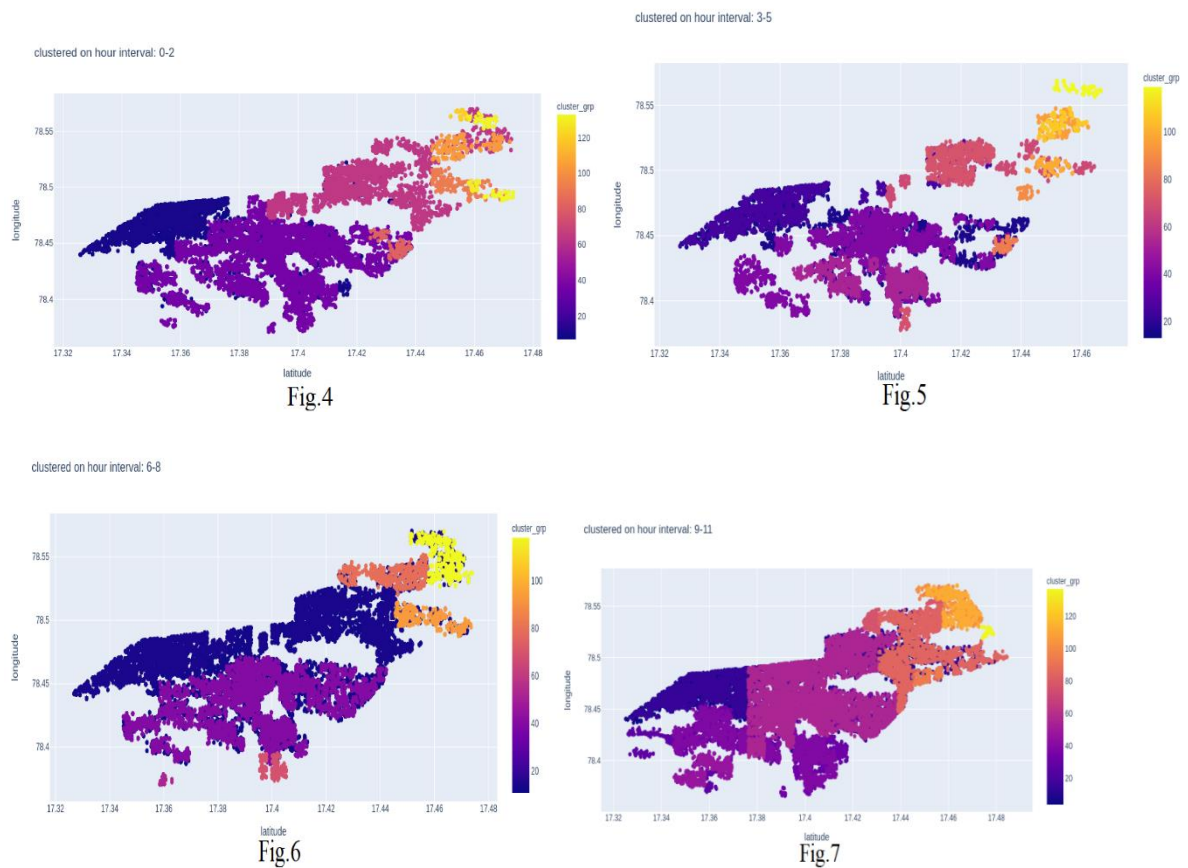
The proposed algorithm is run on a dataset consists of latitude, longitude and timestamp. The data consists of almost six millions of rows. The experiment is run on a server. The server configuration was 90GB of RAM and multicore processor. By trial and error, the batch size is taken to be 21000 of rows. The total time taken to run the algorithm on the whole dataset is 90 minutes. The Eps1 value used in the experiment was 0.5km and Eps2 value was 1. The temporal features are converted from Unix timestamp to just hour and it is used for Eps2.Though, there were some overlapping clusters in the plots but the results are satisfactorily as it runs on large datasets and the data has been clustered well on both spatial and temporal features.

Fig. 4,5,6,7 are the clustered plots of data in a particular region across the different time intervals. Fig.4 is the clustered plot across hourly interval from 0-2 hours. Fig.5 shows the cluster groups from 3-5 hours. Similarly, Fig.6 and Fig.7 show the hour interval of 6-8 and 9-11 respectively.

5. Conclusion

Clustering is an integral part of data mining and knowledge discovery, pattern recognition, image analysis and machine learning. The reason for modification of ST-DBSCAN algorithm is it fails when the data set is huge and the computation cost also increases. This study overcomes its limitation by proposing a new method for very large data sets. As the dataset size increases, we need more main memory in order to fit the data in physical memory. To overcome this defect, a batch processing technique is used, the proposed method is successfully applied on a very large dataset consists few millions of rows. So as to show the real-time application of our algorithm, a spatial-temporal dataset is used to illustrate that the proposed algorithm is very effective in creating cluster on big data. In future, the proposed algorithm can be implemented using Spark to decrease its processing time. And the Spark is one hundred times faster than the traditional techniques for processing big data.

6. References

[1] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, No. 34, pp. 226-231).

[2] Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & knowledge engineering*, *60*(1), 208-221.

[3] Zhou, A., Zhou, S., Cao, J., Fan, Y., & Hu, Y. (2000). Approaches for scaling DBSCAN algorithm to large spatial databases. *Journal of computer science and technology*, *15*(6), 509-526.

[4] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1998). Clustering for mining in large spatial databases. *KI*, *12*(1), 18-24.

[5] Khalilian, M., Mustapha, N., & Sulaiman, N. (2016). Data stream clustering by divide and conquer approach based on vector model. *Journal of Big Data*, *3*(1), 1-21.

[6] Wang, M., Wang, A., & Li, A. (2006, August). Mining spatial-temporal clusters from geo-databases. In *International Conference on Advanced Data Mining and Applications* (pp. 263-270). Springer, Berlin, Heidelberg.

[7] Agrawal, K. P., Garg, S., Sharma, S., & Patel, P. (2016). Development and validation of OPTICS based spatio-temporal clustering technique. *Information Sciences*, *369*, 388-401.

[8] Kisilevich, S., Mansmann, F., Nanni, M., & Rinzivillo, S. (2009). Spatio-temporal clustering. In *Data mining and knowledge discovery handbook* (pp. 855-874). Springer, Boston, MA.

[9] Wardlaw, R. L., Frohlich, C., & Davis, S. D. (1990). Evaluation of precursory seismic quiescence in sixteen subduction zones using single-link cluster analysis. *pure and applied geophysics*, *134*(1), 57-78.

[10] Lamb, D. S., Downs, J., & Reader, S. (2020). Space-time hierarchical clustering for identifying clusters in spatiotemporal point data. *ISPRS International Journal of Geo-Information*, *9*(2), 85.

[11] Bermingham, L., & Lee, I. (2017). A framework of spatio-temporal trajectory simplification methods. *International Journal of Geographical Information Science*, *31*(6), 1128-1153.

[12] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, No. 34, pp. 226-231).

Software Defined Networks in IoT Devices: A Comparative Study

Mohommad Rashid[1], Irfan Alam[2]

[1]Department of Computer Science and Information Technology

Maulana Azad National Urdu University, Hyderabad

[2]University of Hyderabad, Hyderabad

[1]mdrashidyusuf@gmail.com, [2]irfanrmamu@gmail.com

Abstract: In the era of rapid advancement in information service innovation, Internet of Things (IoT) and Software Defined Networks (SDN) are common buzz words we come across. The study begins with the introduction to SDN and to the IoT. The different layers that are present in the SDN architecture are then explored.The study then proceeds with the differentiation between Traditional networks and Software Defined Networks. Further the various types of SDN models that are used to implement SDNs are studied. We then explain how and why SDNs can help in improving IoT architecture. We then study some techniques of how SDN can be implemented in some real-life examples in IoT devices. We then conclude the study by mentioning the benefits and scope of SDNs in IoT devices.

Keywords: Software Defined Networks, Internet of Things (IoT), IoT Devices

1. Introduction:
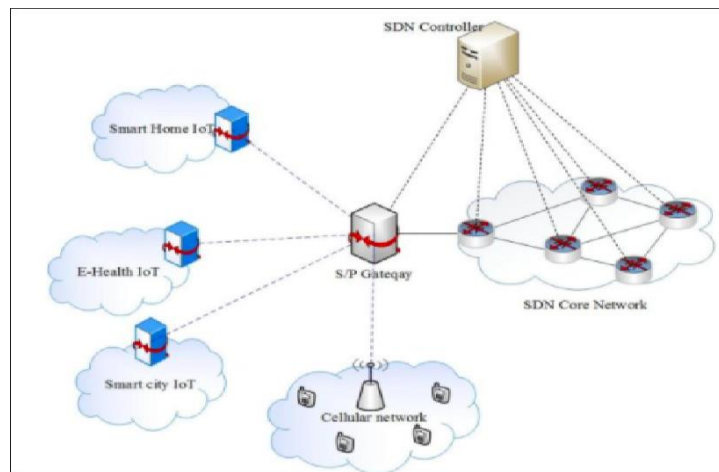
***Figure 1:** Illustrate the general view of SDN architecture [1]*

The Internet of Things (IoT) expedite billions of devices to be empowered with network connectivity to collect and exchange real-time information for affording intelligent services. The Internet of Things (IoT) enables billions of appliance should be empowered with community connectivity to acquire and swap real-time records for affording sensible services. Till the year 2020, IoT consisted of an estimated 30 billion devices [3]. Figure.2 shows complete view of

Internet of Things [5]. IoT lets inrelated appliances to be restrained and accessed remotely within side the presence of adequate network framework. Miserably, conventional network technologies which includes company networks and traditional timeout-primarily based totally shipping protocols aren' table to managing such necessities of IoT in an efficient, extensible, coherent, and cost-powerful aspect [10]. A network structure technique is called Software-Defined Networking (SDN)which permits the network to be intelligently and centrally manipulated, or 'programmed,' the use of software program applications [4]. It enables operators manipulate the whole network continually , irrespective of the underlying network technology. It makes use of software program (primarily based totally controllers) such as application programming interfaces (APIs) to deal with underlying hardware device setup and direct site visitors.

2. Preliminaries

SDN Architecture

SDN allows a brand new manner to monitor the flow of data packets over a consolidated networks. SDN allows the programming of network in a centrally managed way like software program and the use of open APIs. By commencing up conventionally stuffed network platforms and executing a typical SDN control layer, operators can manage the complete network and its gadget perpetually, no matter the complicacy of the underlying network technology. In [6], Kreutz et.al explains all the aspects of SDN in their survey paper.

From figure 1, we may see that the architecture separates control plane from data plane and consolidates the network controller. The separation of manipulate aircraft and information aircraft allows greater network limberness, manageability and adaptability for the elevated-transmission capacity vital character for updated applications. The Open

Networking Foundation (ONF) holds the main function in SDN normalization as illustrated in figure. 3. SDN architecture has been detailed written in [7] and [8].

SDN architecture model consist of 3 layers

Application Layer:

It includes the end-consumer enterprise packages that devour the SDN communications facilities.

Control Layer:

This layer includes SDN controllers associated with control plane. It affords the reasonably centralized manage capability that super intends community forwarding conduct via a public interface.
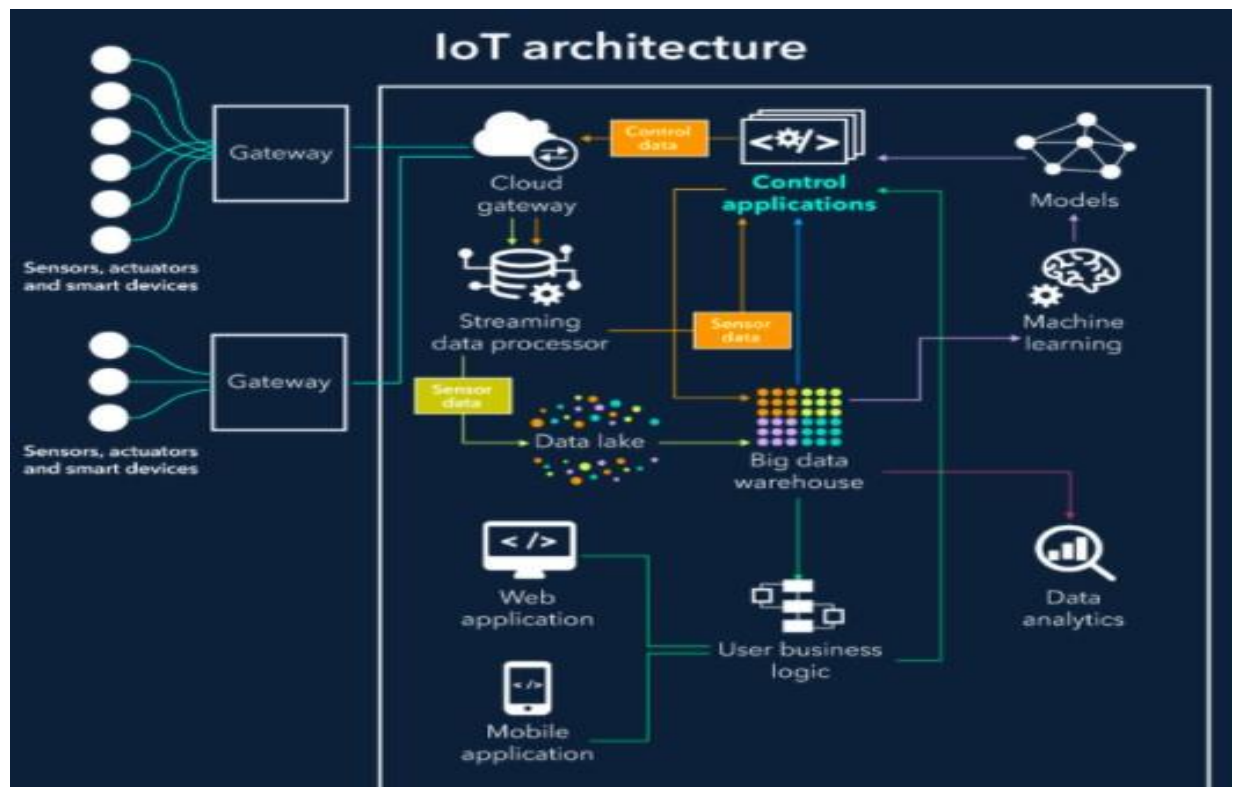
Figure 2 IOT Architecture

- **Infrastructure Layer:**

  This Layer includes the network elements (NE) and gadgets that associate to the data plane. It is accountable for packet management such as how to forward the packets and which path is to be taken to switch by packet.
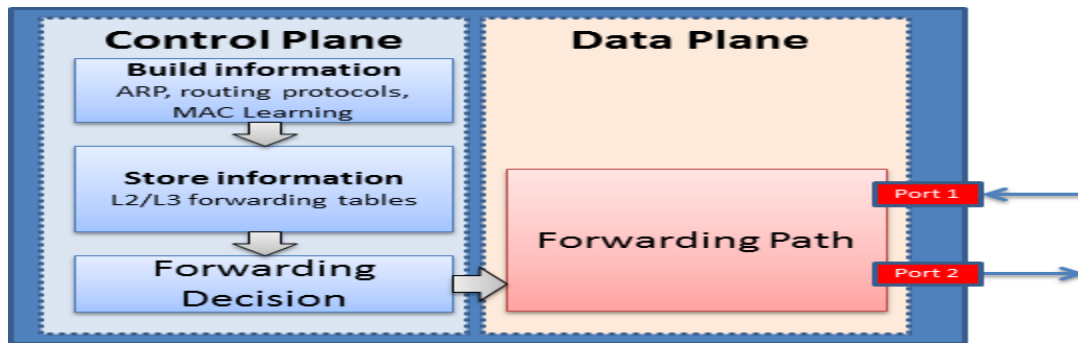
Figure3. *Infrastructure Layer*
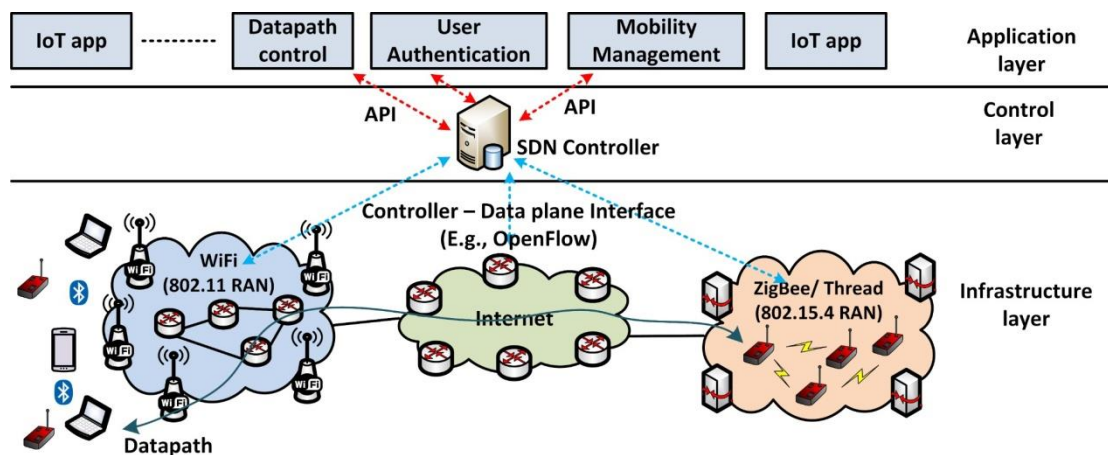
Figure 4. *SDN Architecture*

3. Difference between conventional networks and Software Defined Networks

- **Conventional Networks are hardware based.**

Conventional Network devices consist of Routers/ switches that have a Control Plane and a Data plane.

Control Plane – It mentions all of the features and procedures that decide which route apply to ship the packet or frame.

Data Plane - Data plane mentions to all of the features and procedures that ahead packets/frames from one interface to another interface on the basis of control plane logic.

- SDN are software based, instead of every router making decisions based on its control plane, the task is centralized to a common controller called the SDN Controller that instructs each router on how to send the data from one destination to another.

173

- In conventional networks the routers make decisions based only on their neighboring connected devices.

- In SDN, the SDN controller looks on the network as a whole (by accessing the traffic and load on all routes it can make the best-informed decision for message transfer)and then makes the decision. Conventional networking calls for new hardware to growth
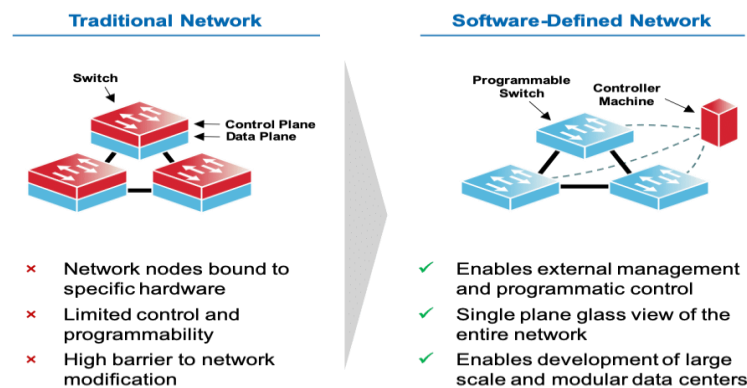
its network capacity.

- SDN have become a famous opportunity to standard networking as it lets IT directors' essential sources and traffic capacity as wished without requiring a funding of extra bodily framework. For more clear view we have put different pictures of conventional n/w and SDN in Figure 5, 6 and 7. MODEL'S TYPE

An idea of centralized software program is to monitor switches and routers as well as all software program related to management of network. Two kind fashions of SDN are below
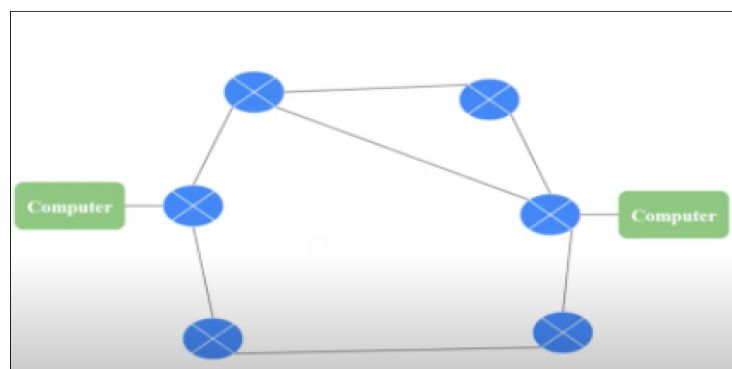
Figure 5. Conventional Network 1

1) **Open SDN or controlling by API:** A protocol like Open Flow is used by Network administrators to govern the functioning of switches (both logical and physical) at the data plane point. Open flow protocol allows the SDN controller to be able to directly connect with all the routers. Routers support Open Flow protocol.A network director employs a protocol such as Open Flow **to manipulate** the **conduct** of **digital** and

bodily switches **on the statistics aircraft** point. Open flow protocol allows the SDN controller to be able to directly connect with all the routers. Routers support Open Flow protocol.

- South Bound Interfaces = Connection between physical network devices and SDN controller

- North Bound Interface = Connection interface between SDN controller and Application Layer

Instead of the usage of a single protocol, utility programming interfaces manage statistical actions via community associated with every gadget.

Open Flow: It firstly describes the **verbal exchange** protocol in SDN architectures which facilitated the SDN controller to **immediately engage** with the forwarding **aircraft** of **community gadgets along with** switches and routers as shown in figure [8].

Figure 7.1 *Figure 7.2*
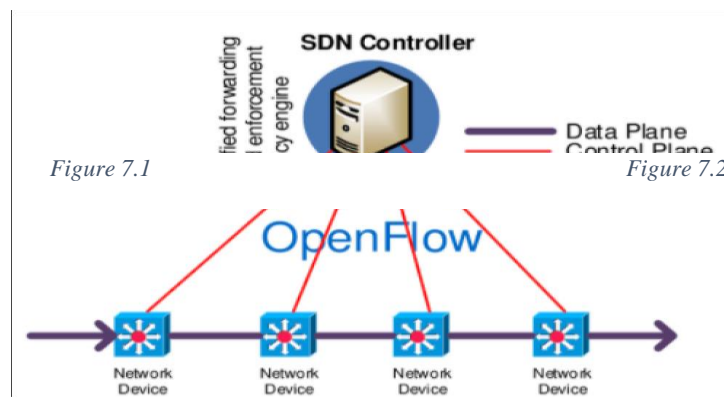
a) It is **one of the** first software-**described** networking (SDN) **requirements** and **described** the **conversation** protocol **among** SDN controllers and the forwarding **aircraft** of
Figure 8

b) Benefits **consist of** its programmability, centralized intelligence, and **the way it** abstracts **community** architecture

c) The **unique idea started out** at Stanford University in 2008 **however** has since been **controlled through** the **Open Networking Foundation (ONF).**

2) **SDN Overlay Model:** A virtual network (VN) is run by Software-defined networking on the top of a current equipment foundation. It makes dynamic passages to various data servers available and remote or on-premise. The VN allots data transfer capacity

over an assortment of channels and relegates gadgets to each channel, while the actual network remains immaculate.
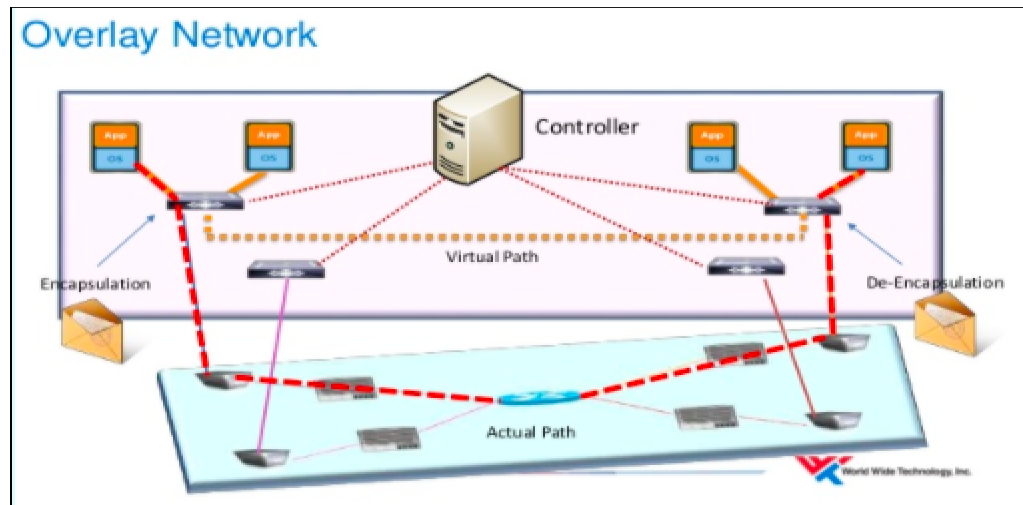
Figure 9 *SDN Overlay Model*

3) **Hybrid SDN:** To support diverse capabilities on a network, this paradigm combines software-defined networking with traditional systems administration principles in a single environment. Standard system administration conventions continue to guide some traffic, while SDN assumes responsibility for other traffic, allowing network directors to gradually familiarize themselves with SDN. This model consolidates programming characterized coordinating with an inheritance climate.
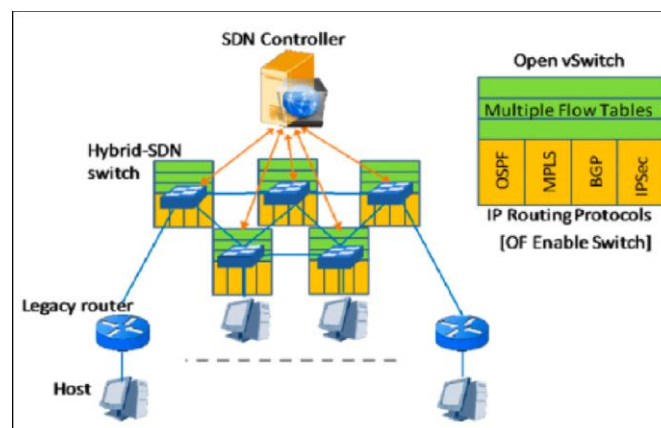
Figure 10 *Hybrid SDN*

4. Few Proposed Ideas / Techniques:

1. Software Defined Wireless Sensor Networking (SDWSN)

It makes use of techniques for green aid control and assured QoS aid have been used to enhance community performance. The techniques used are aimed toward simplifying QoS

provisioning for those constrained technology. The use of programming strategies to increase WSNs packages and their implementations.

2. Software-defined LAN

Software-described nearby location community (SD-LAN) is the facts centre thing of a Software-described Networking (SDN) solution. By deploying SD-LAN software program to every unmarried server within side the facts centre, we can automate server-to-server communication and substantially enhance the overall performance and ease of center networks. It approaches the community can now function hroughout facts centres or even more than one clouds.

3. IoT Traffic Security Upgradation:
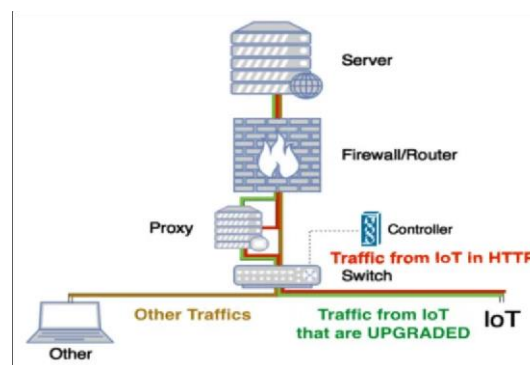
Figure 11 IoT Traffic Separation

Deep packet inspection (DPI) and visitor's separation strategies are implemented using SDN. To enforce SDN, Raspberry Pi as IoT appliance, "Kodi Media Center" on the gadgets as the software media center, and OpenFlow Protocol are used. The arrangement provides security measures such as confidentiality and integrity. It also relieves different risks without the need to alter the IoT gadgets. Vulnerability of Kodi show it utilizes Hypertext Transfer Protocol (HTTP) for additional items downloads, hence it becomes vulnerable to man-in-the-middle (MITM) attack.. The solution to the Man in the Middle attack is to utilize HTTP Secure (HTTPS), which added a transport layer security (TLS) layer on top of the transport control protocol (TCP) to provide both secrecy and integrity.

5. Advantages and disadvantages of inclusion of SDN in IoT

A thorough study of the topic revealed the following pros and cons of Software Defined Networking in IoT:

5.1 Pros

❏ With the exponential boom of gadgets related to the Internet, keeping and securing the sort of massive scale and heterogeneous community is a challenging task.

❏ To create a successful and robust IoT, there needs to be agility and elasticity in the network which can be achieved using SDN.

❏ The traditional networks are not equipped to deal with Big Data and this is where SDN helps with the right kind of data management.

❏ IoT connectivity services rely on elementary functions such as forwarding and routing, quality of service (QoS), and security.With a huge number of IoT enabled devices, their interconnectivity using a software would increase the security. The central program or the SDN Controller would deal with allocating resources and bandwidth flexibly to effectively deal with the increased load.

❏ In the SDN architecture, the controller plays out a key job. It offers with networking offerings consisting of routing, firewall, load balancing, QoS, charging for the entire community. A centralized controller for a community can takes benefit of world view and manipulate of the entire community that is an applicable feature for the framework execution and advancement.

❏ In the case of portable devices which would have to be connected to new spots, having a central program would help to retain the original network policies and controls in all different spots.

When there is a lot of traffic, the SDN controller can isolate the IoT traffic from the other traffic by analyzing the source MAC address and using the SDN switch to send the IoT traffic to a proxy.. This ensures higher security as unauthorized access is not possible and there is no leak of information. Also, it enables a smoother and less time-consuming allocation of resources.

In recent years, SDN related activities have mostly centered on how a logically centralized SDN computation logic, often designated as an SDN controller or orchestrator, can provide network devices with configuration information pertaining to the various features required to deliver a (connectivity) service.

Also, some IoT services, such as dynamic monitoring of biometric data, manipulation of sensitive information, and privacy needs to be safeguarded whenever this information is forwarded over the underlying IoT network infrastructure. SDN introduces a high degree of automation in service delivery and operation—from dynamic IoT service parameter exposure and negotiation to resource allocation, service fulfillment, and assurance.

5.2 Cons

However, nothing comes with infinite pros and that is true in case of SDN as well. It is obvious that a centralised controller cannot supply all networking functions as the network grows in size, as this can result in longer response times and inefficient resource allocation. Therefore, at the culmination of our research we conclude that Software Defined Networks can turn out to be ground breaking and would be extensively used in IoT devices in the future.

6. . Conclusion and Future work

Concepts of Internet of things and SDN are broadly studied in this paper. Along with explanations, the comparison of conventional networks and software defined networks are listed in clear pictures.

Some novel techniques are also proposed to be used to overcome the shortcomings in inclusion of SDN in IoT. In the last pros and cons of SDN in IoT are written in very clear way. In future many new ideas can be proposed based on this useful studies and comparisons.

References:

[1] Tayyaba, S. K., Shah, M. A., Khan, O. A., & Ahmed, A. W. (2017, July). Software defined network (sdn) based internet of things (iot) a road ahead. In *Proceedings of the International Conference on Future Networks and Distributed Systems* (pp. 1-8).

[2] Amin, R., Kumar, N., Biswas, G. P., Iqbal, R., & Chang, V. . A lightweight authentication protocol for IoT-enabled devices in distributed Cloud Computing environment. Future Generation Computer Systems, 78, 1005-1019 2018.

[3] Salman, Ola, et al. "IoT survey: An SDN and fog computing perspective." *Computer Networks* 143 (2018): 221-246.

[4] F. Graur, "Dynamic network configuration in the Internet of Things," 2017 5th International Symposium on Digital Forensic and Security (ISDFS), TirguMures, 2017, pp. 1-4.

[5] Internet of Things (IoT) - Cisco. [Online]. Available: http://www.cisco.com/c/en/us/solutions/internet-of-things/iot-products/solutions.html. [Accessed: November 2017].

[6] D. Kreutz, F. M. V. Ramos, P. Verissimo, C. Rothenberg, S. Azodolmolky, S. Uhlig, "Software-Defined Networking: A Comprehensive Survey," http://arxiv.org/abs/1406.0440, in submission, 2015.

[7] Open Networking Foundation, SDN Architecture, Issue 1, ONF TR-502, June 2014.

[8] H.F. Xavier and S. Seol, "A Comparative Study on Control Models of Software-Defined Networking (SDN)," Contemporary Engineering Sciences, vol. 7, no. 32, pp. 1747 - 1753, 2014.

[9] J. Rak, "'Resilience of Future Internet Communications," In Resilient Routing in Communication Networks. Springer International Publishing, 2015, pp. 45- 83.

[10] P. P. Ray, A Survey on Internet of Things Architectures, EAI Endorsed Transactions on Internet of Things, ICST, 2016.

[11] Abdullah Al Hayajneh , Md Zakirul Alam Bhuiyan and Ian McAndrew " Improving Internet of Things (IoT) Security with Software-Defined Networking (SDN) " 2020

[12] https://www.ciena.com/insights/what-is/What-Is-SDN.html

[13]https://reswww.zte.com.cn/mediares/magazine/publication/com_en/article/201601/448976/P020160311298781218045.pdf

[14]https://ieeexplore.ieee.org/document/8017556%22

[15]https://www.researchgate.net/publication/327955401_Software_Defined_Wireless_Sensor_Networks_SDWSN_A_Review_on_Efficient_Resources_Applications_and_Technologies

[16] https://medium.com/@Infosys/software-defined-networking-a-critical-enabler-of-iot-eb4e6e4b411f

[17]https://www.ibm.com/services/network/sdn-versus-traditional-networking#:~:text=The%20most%20notable%20difference%20between,virtually%20throughout%20the%20control%20plane

[18] https://www.mdpi.com/2073-431X/9/1/8/htm

[19]https://cse.iitkgp.ac.in/~smisra/theme_pages/sdn/index.html

[20]https://www.vmware.com/topics/glossary/content/software-defined-networking

[21]https://www.researchgate.net/publication/319602888_Software_Defined_Network_SDN_Based_Internet_of_Things_IoT_A_Road_Ahead

Detection of Phishing Attacks in Web Environment using Unsupervised Machine Learning

Sheo Kumar, Rajesh Tiwari, Neethu Choudhary

Computer Science and Engineering

CMR Engineering College

Hyderabad, Telangana

sheo2008@gmail.com, drrajeshtiwari20@gmail.com, nc.gottipati@gmail.com

Abstract: *Phishing has no fixed complete solution. Phishing means getting information from a person without their knowledgesuch as user's id, passwords, credentials, etc.,Phishing attacks occur in over every day-to-day life, where one person who does not know the outside world threats give their personal information through emails send by some attackers and through SMS send by spammers etc., There are different types of phishingthey are vishing (means through phone calls), whaling(attacking a group of targeted people), search engine phishing(using the webfor searching online), etc.In this research we perform Phishing attacks detection using machine algorithms include Support vector Machine(SVM) and we constructed dataset from UCI machine Repository of phishing sites and dataset is trained and tested and then apply these classification and regression algorithms to find accuracy of detecting phishing sites. And tools such as Phish Tank are used whether the site is phished or not. MX Toolbox is used to blacklist the phished sites and WHOISis used to know the details of phished sites. Counter Phishing technique is used to know the unauthorized loins of that websites.This result shows that the Unauthorized logins and by applying various machine learning algorithms the SVM shows highest accuracy in detecting the phished sites.*

Keywords: Logistic regression, machine learning, Spam detection and E-mail.

1. Introduction

Phishing is a type of fraud in which criminals impersonate legitimate websites in order to get personal and sensitive information from consumers. Email login credentials, transac-tion one-time passwords, usernames and passwords for users' bank accounts, debit card and credit card pin numbers, and other information may be included. In phishing, the attacker poses as a respected entity and dupes the target into divulging personal information. Phishing includes duping the user into sharing personal information with the attacker, making it aless difficult method of breaching a computer's defence systemthan hacking. Phishing attacks are typically carried out using e-mails including faked logos and malicious links that appear to be genuine to the naïve user.

According to Yahaya et. al. [1] a new phishing website is generated on the internet every 20 seconds. Furthermore, receivers open 70% of the scam emails they get. According to

S. Laidlaw et. al. [2] 0.470 percent of bank account holdersare victims of phishing attempts each year, resulting in losses ranging from $2.4 million to $9.4 million per million

con- sumers. These figures demonstrate how easily attackers may target naive individuals and the need of having an effective phishing attack detection system.

A phishing assault is carried out in the following manner. The attacker imitates a prominent website's login page and registers it with a URL that appears extremely similar to a legitimate one. The user is then given an email with a link to the phishing website. The e-body mail's is disguised to make it appear real to the recipient. The user clicks the link and logs in with his orher credentials. A script on the backend of the cloned website's login page grabs the unknowing user's credentials and makes them available to the attacker. The attacker can then exploit the user on a legitimate website using these credentials. The method is depicted in Fig. 1.

Fig. 1. *Phishing Mechanism*

The disguising of the website URL is a typical feature of all phishing assaults. The URL of the phishing website is most often used to deceive the victims of phishing attempts. Cybersquatting and Typosquatting are two strategies employed by attackers. URL hijacking is known as cybersquatting. The attacker purchases the domain name of a well-known organisation that does not have a website associated with it. Typosquatting is the practise of buying a website URL that appears and performs exactly like a legitimate website but has a typographical error. Google.com and goggle.com are two examples of this. When inputting the website URL, internet users frequently make typing errors, which are exploited by attackers. Aside from these choices, the attacker might alter subdomain names, query durations, add redirect requests, or make the URL too long. Because phishing data is

easily avail-able in phishing databases such as Phishtank and OpenPhish, an attacker can easily change a website's URL by altering sub-domain names to create a new website when it is suspected of being tied to phishing. As a result, an intelligent approach for detecting phishing URLs and reducing phishing attempts is required. The classification of website URLs into phishing and legal URLs can be aided by data mining techniques. The following are the paper's main contributions:

- Present a broad idea of phishing attacks and different possible shield designs using SVM classification in this study.
- Create a method with a minimal computational cost and evaluate its performance in various contexts.
- To increase future performance and accuracy in terms of phishing e-mail classification and prediction.
- To shorten the time required to classify the email with infinite learning and optimise memory consumption in the classifier process, while the characteristics of phishing e-mail features have changed.
- To compare and contrast the proposed framework with existing methodologies for phishing email detection.
- To compare and contrast the proposed framework with existing methodologies for phishing email detection.

The remainder of the work is organised as follows: section 2 discusses several literatures and their shortcomings. Section 3 delves into the specifics of the suggested method's implementation. Section 4 deals with the comparative analysis and interpretation of the outcomes. Section 5 of the article finishes with suggestions for further improvements.

2. Related Work

Routhu Srinivasa Rao et al. [3] used a variety of machine learning algorithms to identify phishing websites, and then compared the accuracy of the various systems. According to their findings, the Random Forest algorithm has the best accuracy, precision, and recall. Their analysis also illustrates the main indicators that phishing is occurring. Namrata Singh et. al. [4] proposes a categorization approach to classify phishing assaults. The UCI Irvine ML collection was used to extract features from numerous sites. The experiment was carried out in MATLAB, and the Extreme Learning method was shown to have the greatest accuracy of 95.34 percent.

Ratinder Kaur et. al. [5] demonstrated how to use NLP and machine learning to detect phishing email attempts. They used text semantic analysis to detect any nefarious activities. To parse sentences, NLP was utilised. The majority of the work was done in Python, and the

findings show a 95 percent accuracy in phishing website classification. To increase the dataset's quality, N. M. Shekokar et. al. [6] suggested feature selection techniques for phishing detection. They compared the accuracy of their algorithm to that of other regularly used categorization systems. Tree algorithms did not do well on small datasets, according to their research. The findings of the Lazy K Star algorithm were the best. Weka was used to conduct the entire study. To classify phishing URLs, a model was developed using the Random Forest method in paper [7]. To assess the feature set, the URLs were parsed. On the test dataset, their method had a 95% accuracy rate. Without any prior knowledge of the URL, A. Mishr at. Al. [8] employed neural networks to extract properties from it. With Adam optimizer, an accuracy of 94.18 percent was achieved. Neda Abdelhamid, et. al. [9] compared logistic regression approaches with bigrams with deep learning approaches such as CNN and CNN-LSTM architectures. Tensorflow and Keras were used in the research, and the dataset was compiled using OpenPhish and Phishtank. Using CNN-LSTM architecture, the authors reported a 98 percent accuracy rate. The authors proposed a phishing detection technique for Chinese websites in [10]. By mining the semantic properties of words in Chinese web pages, the model's performance was investigated. On the dataset, different machine learning techniques such as RandomForest, Adaboost, and Bagging were compared. D. K. Singh et. al. [11] created a Google Chrome plugin that uses machine learning algorithms to detect phishing websites. This was accomplished using the UCI ML Repository. The problem of this expansion is that the number of malicious websites is increasing every day, with new ones arriving every day, and the training set for the study is too small. The writers of [12] explained the different URL characteristics such as primary domain, subdomain, and website ranking for phishing detection. PhishScore was built by the inventors of [13], a tool that does lexical analysis on URLs to detect phishing. In their research, they devised a strategy based on the relatedness of URLs. The authors of [14] described the many types of online spoofing attacks and sought to identify phishing URLs using domain name attributes. The authors of [15] investigated the accuracy of various classifiers for spam email prediction. They had 2289 phishing emails with legal and phishing domains in their database. Text mining was used to extract a set of terms from the email body, which were then classified using basic machine learning methods such as random forests, decision trees, SVM, BART, and neural networks. Among these classifiers, the Random Forest approach was shown to be the most accurate. The authors gathered 2456 websites that may be classed as legitimate, questionable, or phishing in [16]. To classify the website into one of the above-mentioned categories, they employed data mining techniques such as Random Forest, Neural Network,

Decision Tree, and Neural Networks. The work's findings were compared to those of other researchers working on similar and dissimilar datasets. The lack of a description of the phishing features taken into account for machine learning purposes is a weakness of this study.

3.   Machine Learning In E-Mail Classification

The problem of determining whether a given website URL is phishing or not is a binary classification problem that can be solved using labelled data and supervised learning. This challenge necessitates the collecting of recent website URLs from both phishing and genuine sources. The dataset is then prepared by extracting essential variables that aid in the differentiation of phishing websites from authentic ones. In order to provide input to the machine learning algorithm, the features must be processed. The model is then trained on the training set before being tested on the testing set to verify its correctness. The process is depicted in the flow chart shown in Fig. 2.

After collecting the dataset, each feature has a threshold value, by using it we can compare the values .If the value is in range then it is indicated as "1". Or else "0" in binary vector and this is further used in testing. There are six parts-

- In this the data is transformed in to tree structure format.
- Rule generation is applied it means applying the "if then else" rule to the tree.
- Bayes classifier is used to remove the rules and get the remaining generated rules to next stage.
- Test dataset contains both the legitimate and phishing URLs and calculate their performance.
- Here classification is done it declares accuracy of the URL through rules and class label for each URL.
- According to accuracy and performance whether the sites is phished or not is declared

A. Feature extraction

1) URL Based Features::

- **IP Address:** A phishing website may utilise an IP address or hexadecimal characters in the domain of the URL instead of a textual domain name. http://102.24.134.12/page.html is an example of an IP address in a URL. According to a study in [17], the usage of IP addresses or hexadecimal characters has been connected to phishing websites and suspicious activities in 46.66 percent of cases.
- **Use of the @ sign in a URL:** When you use the @ symbol in a URL, the browser ignores everything before the symbol. The '@' symbol is frequently used in the phishing website address. Because consumers rarely read the complete URL, this is frequently used to

186

exploit phishing. According to a study published in [17], the letter '@' can be present on 20

- **HTTPS in the middle of the URL:** Phishers use the presence of "https" in the domain of the URL to deceive people. Users are tricked by http in
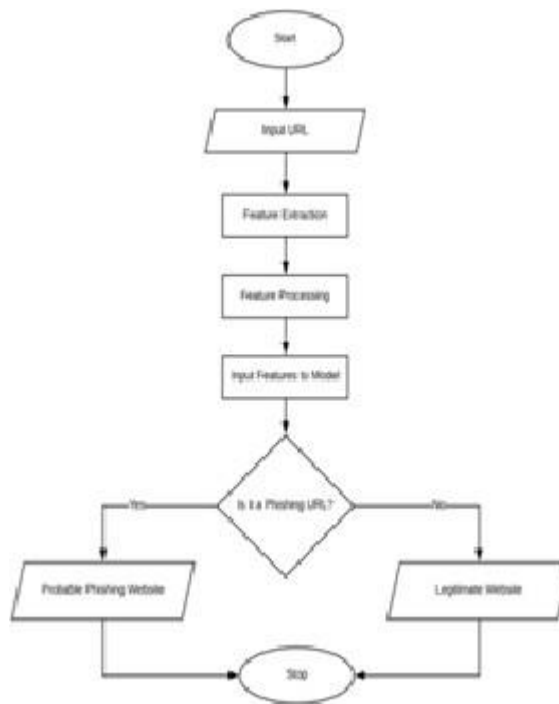
Fig. 2. Methodology

URL domains. https.paypal.com-account-update.

e3d3idw3k4securityalert.cenksen.com.tr/paypal.com/ is an example.

- **Dashes in the URL:** The presence of a '-' in the website URL indicates phishing, as reputable web page URLs do not contain dashes. Phishers frequently utilise this technique to deceive consumers into believing they are dealing with legitimate websites. In our analysis, the average number of dashes on legal websites was 0.06, whereas the average number of dashes on phishing websites was 0.49, indicating that dashes are used mostly on phishing websites.

- **Long URL:** A phishing URL that is unusually long is usually a symptom of phishing, since phishers utilise it to hide the suspicious component of the URL. The average number of characters in all of the URLs in our database was 69.3. The average length of a phishing website URL was 92.7 characters, while the average length of a reputable website URL was 36.63 characters. Our variation in length is a key element that distinguishes authentic

websites from phishing sites, and it was taken into account in this research. The author discovered that

73.33 percent of phishing URLs have an unreasonably long URL in his study [17].

- **Delimiter Characters:** Because delimiter characters are commonly encountered in suspicious URLs, the existence of an unusual number of delimiter characters such as "#,

, ,&, %" has been taken into account. _

- **Query Length:** In comparison to ordinary legitimate websites, the query length component of phishing web-sites is much longer, which is another parameter em-ployed in our study to classify phishing and non-phishing websites.

  - **Dot Count:** The number of dot symbols in the URL's domain can be used as an identifier to identify phishing websites. A phishing website's domain section usually has more dots than a reputable website's domain section. To count the dots in the website's URL, the Urllib Python package was used in conjunction with Pandas. When calculating the number of dots in a URL, the protocol and TLD sections are removed. In our research, the average number of dots in the URL's domain was 0.55. The average number of dots on phishing websites, on the other hand, was 0.87.

  - **Number of sub-domains in the URL:** Phishing websites are frequently identified by the number of sub-domains in the URL. A website with an unusually large number of sub-domains is most likely a phishing site. In our research, the average number of subdomains was 1.29. In phishing websites, however, this number was close to 1.5.

  - **Redirect Request:** The presence of a forward slash (//) in a website's URL is often indicative of a redirect request. "http://www.legitimate.com//http://www.phishing.com" is an example of this. The redirect request is placed after the double slash in this case. This function is useful for detecting phishing websites.

2) Domain Based Features:

- **Website Page Rank:** The value of a website is typically determined by its page rank. Alexa maintains a database of websites that is used to determine a website's page rank. In comparison to legitimate websites, a phishing website is frequently unranked or very lowly ranked. This is another significant distinction between phishing and non-phishing pages.

188

- **Domain Age:** In comparison to legal websites, which are normally online for a long period, phishing websites have a relatively brief lifespan. The Whois API may be used to ascertain the domain's age, which is an important indicator for distinguishing between phishing and non- phishing websites.

- **Website Validity:** The Google Whois API can be used to determine whether or not a website is still functioning. The majority of phishing sites have a brief lifespan and are taken down once suspicious activity is recognised. Validity is a key distinguishing factor between real and fraudulent websites.

B. Classification

Support Vector Machine: SVM is very similar to the decision tree i.e., it is supervised machine leaning algorithm. It is also used for both the classification and regression. In this we are using SVM for classification purpose. For this we plot the dataset in the dimensional space format with each value to the feature being in particular. Then we perform classification and find the differences in 2 layers. In this, we consider the dataset in two parts: testing dataset and training dataset. In training state, we give 33 featured values and are organized in particular format in SVM. At the testing state, we compare the values in two layers. The output values indicate whether the URLs are phished or not. If the result is equal to "-1" then it is phished site or if the result is equal to "1" then it is non-phished site.

C. Description of Tools:

The tools that are used in this paper are WHOIS, MX Toolbox, Phish Tank. In this paper, the main aim of tools is to tell which tool is better to use at which time. WHOIS: who is a query and response protocol tool. It is used mostly to know whether a website is legal or illegal. As in whois.com, many users register their website or their internet resources init. If we want to know any domain information about a legal website like the IP address etc., It provides the information in fraction of seconds in human readable format. The main advantage of who is, it tells whether a website is legal or not and gives information about a legal website.

MX Toolbox: MX Toolbox is a phishing tool that provides information likewise HOSI, but it also provides protection. It checks if any website is under blacklist, if true or having any problem with email delivery and not getting any help. MX Toolbox provides a solution by offering blacklist protection. In this we can add blacklists that are newly added or created.

Phish Tank: Phish Tank is an anti-phishing site. It is based on community verification i.e., in this site we upload a website and others users in the site vote whether it is a phished site or not. If maximum people vote that it is a phished site then it comes under phished site or else the site is non-phished site.

D. Technical Approach:

In phishing, there are different types of phishing attacks without the knowledge of the people by just one click. There are different ways of obtaining personal information from users. As technology is increasing day-by-day cybercrime is also increasing drastically. To prevent from being phishedpeople should have the knowledge of phishing techniquesin present world and how they work. There are different types of phishing techniques some of them are Spear-phishing,email spam, etc., in this paper, counter phishing techniqueis demonstrated on how personal information is gone to attackers.In counter phishing, a duplicate link is created and sent to a lot of random people through mails and without knowledge users input their personal details like email Id,passwords. This is a lot of information to the attackers to attack.

4. Experimental Analysis

A. Data set

The dataset is been taken from the website i.e., UCI machinelearning repository. The file contains Dataset of phishing websites which is available in .arff format and needs to be converted to .csv format for the classification. In the dataset there are 32 columns and 11,056 rows. Each row represents an attribute that can be a part of each phished websites. And the values 1 refers to the success state, 0 refers to the not detected state , 1 refers to the failure state. Each attribute representsits classification taken from the phished sites. Some of the attributes are as:

- **Using Ip address-** In case if an IP address is used as an alternative for a website URL such as, https://192.16.10.564.html means a phished website that can be used for stealing the user information.
- **Tiny URL-** these are used in some of the sites such as https://bit.ly.3dfe6sd.html which seems to be redirecting to the original site but can lead to phished site.
- **HTTPs token-** phishers may add the https token in the URL domain which seems to be normal website to trick the users.
- **Domain registration length-** based on the present sce- narios, each phished site can only be livedfor a very shorttime, so secure domains pay a high amount to be usedfor a long time periods.

- **Abnormal URL-**host name which needs to be added to each website to gain the trust worthiness. So, if not addedcan be considered as a part of the malicious site.

- **Website forwarding-** few websites in the internet makesuser to redirect from page to page when selects a link,but redirects to another website which can use some issueto the system and can access users information.

- **Popup window-** it is unusual to ask user credentials in a popup window. On the other hand, this feature has been used to gain.

B. Performance classification

Seven machine learning methods were tried on the datasetin this study. The findings of the various algorithms are sum-marised in the table below. The accuracy of several algorithms was determined for both the original feature set and the PCAapplied feature set, as shown in Table I. With a score of 95.82percent, the Random Forest approach was deemed to be the most accurate after PCA. Among all of the algorithms, logis-tic regression performed the worst. Decision Tree, GradientBoosting, Fuzzy Pattern Classifier, and Adaboost all performed similarly to Random Forest Algorithm in terms of accuracy. The accuracy of the original feature set and the PCA applied dataset were determined in the same way as the recall and precision of the various approaches. The Random Forestmethod has the best precision and recall of all the algorithmsexamined, at 96 percent. The accuracy and recall of thedataset were enhanced by 1% when PCA was applied to it,compared to the baseline precision and recall. Among all thealgorithms studied, the Random Forest method had the highestF1 score. Based on the findings, Random Forest appears to bea reasonably strong model for classifying website URLs into phishing and authentic domains websites into legitimate and phishing.TABLE

Algorithm	Accuracy	Sensiti-vity	Specific-ity	Precision	Recal	F1
SVM	95.33	95.82	95.00	96.00	95.00	96.00
Decision Tree	94.09	94.26	94.00	94.00	94.00	94.00
Gradient Boosting	92.19	92.22	92.00	92.00	92.00	92.00
Fuzzy Pat-tern Tree	91.22	92.23	91.00	92.00	91.00	91.00
Adaboost	91.00	90.64	91.00	91.00	91.00	91.00
Gaussian NB	83.28	85.17	83.00	85.00	84.00	86.00
Logistic Regression	73.78	82.89	74.00	83.00	74.00	84.00

PERFORMANCE COMPARISON

5. Conclusions

This research paper describes the current security issues in these days digital world with regard to phishing, as well as the mechanism by which phishing is carried out by attackers cleverly disguised phishing emails. This study focuses on identifying properties useful for detecting phishing websites based solely on the URL, and then categorising the website based on additional properties such as meta tags, favicons, pop-up windows, and redirect links on the web page, among others, using machine learning algorithms. In order to make the model more robust, the number of training samples can also be increased. Phishing will continue to be a problem, and the web browser postponement that blocks phishing websites can be quite useful for users.

References

[1] Yahaya, Saudi, Madihah ,Abdullah, Ismail. "A Review and Proof of Concept for Phishing Scam Detection and Response using Apoptosis". International Journal of Advanced Computer Science and Applications (IJACSA) ,Volume 8,Issue 6.

[2] S .Laidlaw and M .Hillick, "Profiling cyber threats detected in a target environment and automatically generating one or more rule bases for an expert system usable to profile cyber threats detected in a target environment". U.S. Patent 9,503,472. Cyberlytic Limited, 2016.

[3] Routhu Srinivasa Rao and Syed Taqi Ali, "PhishShield: A Desktop Application to Detect Phishing Webpages through Heuristic Approach", Eleventh International Multi-Conference on Information Processing- 2015 (IMCIP-2015). Volume 54, Pages 147-156

[4] Namrata Singh, Nihar Ranjan Roy, "A Hybrid Approach to Detect Zero Day Phishing Websites", International Journal of Information Compu- tation Technology. ISSN 0974-2239 Volume 4, Number 17 (2014), pp. 1761-1770

[5] Ratinder Kaur and Maninder Singh, "A Hybrid Real-time Zero-day Attack Detection and Analysis System", I. J. Computer Network and Information Security, 2015, Volume 9, 19-31.

[6] N. M. Shekokar, C. Shah, M. Mahajan, S. Rachh, "An Ideal Approach for Detection and Prevention of Phishing Attacks", Procedia Computer Science Volumn 49 (2015) page no. 82 – 91.

[7] Mouna Jouini, Latifa Ben Arfa Rabai, Anis Ben Aissa, "Classification of security threats in information systems", 5th International Conference on Ambient Systems, Networks and Technologies (ANT-2014).Volumn 32,page no. 489-496

[8] A. Mishr and B. B. Gupta, "Hybrid Solution to Detect and Filter Zero- day Phishing Attacks", Emerging research in computing, information, communication and applications, ERCICA 2014. page no.373-379

[9] Neda Abdelhamid, "Multi-label rules for phishing classification", Ap- plied Computing and Informatics (2015) volume 11, page no.29–46.

[10] MAAWG (2011). Messaging Anti-Abuse Working Group (MAAWG) Email Metrics Program. 15. third Quarter.

[11] Singh, D. K., Ashraf, M. (2019). Detect the phishing websites in the contex of internet security by using machine learning approach. International Journal of Advanced Science and Technology, 27(1), 104- 111.

[12] APWG (2010). "Phishing Activity Trends Report".From http://www.antiphishing.org/reports/apwg report Q1 2010.pdf.

[13] GARTNER (2007). "Gartner Survey Shows Phishing Attacks Escalated in 2007; More than $3 Billion Lost to These Attacks." Retrieved December 17,from http://www.gartner.com/it/page.jsp?id=565125.

[14] Bimal Parmar, F. (2012). "Protecting against spear-phishing." Computer Fraud Security 2012(1): page no 8-11.

[15] Steve Sheng,1 Mandy ,Holbrook, Ponnurangam Kumaraguru, Lorrie Cranor,Julie Downs, "Who Falls for Phish? A Demographic Analysis of Phishing Susceptibility and Effectiveness of Interventions", Copyright 2010 ACM.

[16] Christy Jeba Malar, A., Kanmani, R., Vijayavarman, R., Praveen Kumar, R., Poorna Bharathi, G. (2020). Implementation of phishing detection using SVM. Test Engineering and Management, 83, 3287-3295.

[17] Ali Darwish, Ahmed El Zarka and Fadi Aloul," Towards Understanding Phishing Victims' Profile",2013 IEEE.

Road Crack Detection and Segmentation for Autonomous Driving

Alisha Raza

Department of Computer Science and Information Technology

Maulana Azad National Urdu University, Hyderabad, India

alisha.raza23@gmail.com

Abstract: *I am working on pothole detection on road for self-driving vehicles is based on deep learning techniques, This paper addresses the estimation of Road crack problems for autonomous driving , it detects crack and potholes using texture and spatial features Road image data capturing from Mobile camera , In this study we discuss Deep learning techniques Convolutional Neural Network Faster-Region convolutional neural network (Faster-RCNN) It deployed on GPU based and image processing board, It processing the input image density of Road crack that will be measured the potholes density and provide us information if any road is crack by fluid, heavy traffic, weather conditions The model that I will create which determine the road crack properly and accuracy will enhance It deploy the problem of autonomous driving by solving the Road cracks In Existing Techniques of Road crack, pothole counting, and tracking suffer from low accuracy , methods were too slow and computationally very expensive It will be better because we have to implement a technique that we achieve better accuracy to solve existing problem here I use only 40 EPOCS and got the accuracy 0.92 which is very good for our proposed work.*

Keywords: *Faster-RCNN, GPU*

1. Introduction

Crack detection is a critical role in transportation maintenance for ensuring driving safety. Therefore, due to the intensity in Road cracks and the difficulty of the backside, such as poor drainage with pavement and darkness of related amount, it remains a difficult task Road crack occurs due to heavy rainfall, transportation, weather changing, cyclone etc,This paper proposes a deep-learning based approach for crack detection, motivated by current progress in applying deep learning to computer vision to identify each image through the collected image dataset, a supervised deep convolutional neural network is equipped. A data set is subjected to a quantitative assessment In increasing world, Road crack is a major problem for autonomous driving there are so many accident occur In our country there are more population is increasing as well as road crack, within the developed industrial will double, with this increase within the wide variety of transportation, road crack is bound to come up as a extreme trouble over the previous few years, In India there are more cities has more population and transportation that make the reason of road cracks some other place, the introduction of smart towns, and the development is being made in the direction of demands for the optimization and green operation of social systems additionally many recognition packages have performed for green

194

and secure. crack roads and asphalts are increasingly subjected to heavy rainfall and shifting environment conditions, resulting in reformist disintegration and corruption, causing transportation, jeopardizing suburban residents' well-being, and reducing profitability. In **figure-1** I will take input as a road crack than I will apply CNN and FCNN algorithm and the machine will predict where is the road and diminishing the security level on the Road However, the highest expects to offer the capacity to consume the stun and maintain the accessibility of the capacity if an unexpected event occurs the crucial elements in the Road crack control system to ensure.
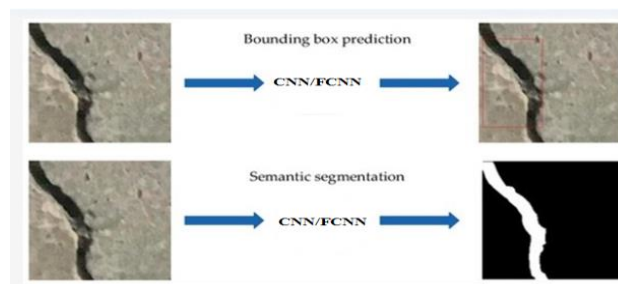
Figure 4 Road crack detection

After a lot of work to enhance the presentation, a comprehensive audit of process vision techniques for road crack was conducted. on the other hand, the review consistently fails to be tailored to local circumstances, making the suggested strategy only useful in a prearranged For several countries, environmental road cracks are big problem because they cause road because of heavy snowfall, and growth, both of which have a negative impact on the economy, However, there are several drawbacks to this approach, including high costs and the need for all zoographical areas, nation's states and other country and cities, executing innovation is an optional solution for reducing road crack issues. In rush hour gridlock intersections, there is an alternate stream that deals with the issue. Finding cracks and potholes over large perspective shifts in the presence of noise and obstruction is a difficult task. Our proposed methodology reliably identifies road cracks and potholes by using an Algorithm for road division and other pre-processing techniques use to create up-and-comers that are extremely referred to as road cracks, this paper's main goal is to study road crack detection methods that include image processing with a GPU and a camera. It will take normal image and pothole images by camera, laser camera to identify the cracks. The proposed work we extract input images in a collection of normal and potholes images and we decide whether there is a road crack or not based on prediction and learning rate, and images are converted from RGB (Red-Green-Blue) to grey scale, which allows the system to process digital data it is very versatile, accurate, and cost-

effective. In This paper we present image pre-processing technique for road crack detection method focusing on faster-region convolution neural network(Faster-CNN)method which shown in **figure-2**, that represent the Flow work of my road crack detection model, Firstly I will take image normal and potholes images and then image pre-processing of those images because in data there is more noisy data, irrelevant data when I preprocess the images all the noisy data will remove all the images come in same size than I will load model which I create than classify the image either there is crack or not and result will be show with better accuracy concludes with a brief analysis of future directions my scope of this Research is that In many areas related to detecting differences in regular patterns of roads and surfaces, Road Crack Detection has its tremendous use, so that effective measures against it can take place on time.

A few places where road crack detection can be used:

➢ Road scrutiny

➢ Status of roads in extreme weather conditions such as quake, tsunami, etc. occur.

➢ Recognition of breaks on the roads for Autonomous Driving Cars Crack Detection, Detection of Traffic Light Pole, Houses, Number of Adjacent moving vehicles and distance between contiguous moving vehicles and so forth.
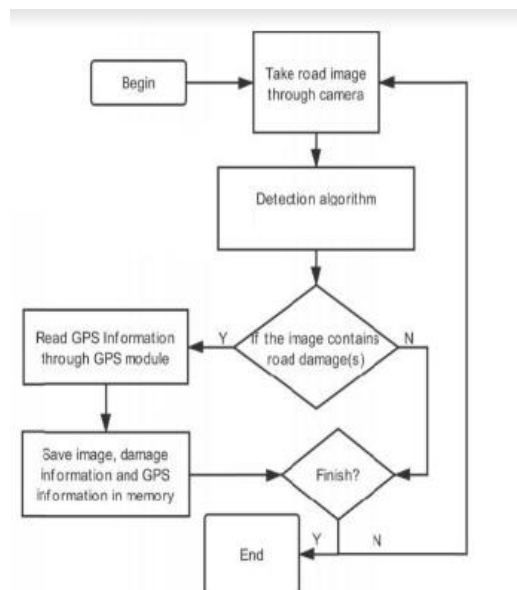
Figure 5 Work flow

There is some problem statement which I found after studied so many papers that is In today's globe, there are more modes of transportation that connect all environmental regions, country, nations, towns, and villages are connect with atmosphere train, metro stand etc.. Highway transport is the unorganized way to connect all the points of road or exterior failure because of cracks "weighty snowstorm, low contrast and group of weighty transport etc. It's all the reason

of road exterior deterioration and also the reason of cracks we can detect the crack by using the pre-processing techniques and take review about tools After find problem statement that I decide how to overcome of these problem statement so I make my objectives which I cover in my research thesis

➢ To monitor and detect the crack on a road
➢ To augment the model by using relevant images/dataset
➢ To design a model to achieve a accuracy
➢ To implement the model on dataset for validation of result

In my research there is so many challenges that I see which I include in my paper and that is This Thesis will right now manage recordings of roads or surfaces that have been shot in any case, if this venture needs to utilize this break recognition for independent driving continuously. It should make a Smartphone application or it will require little datasets of an equipment gadget if this venture doesn't discover an adequately enormous dataset for crack location there are not many calculations accessible for identifying cracks in road pictures or recordings Low-level picture preparing strategies are not impervious to the adjustment conversely and can't recognize breaks when lighting is low or breaks are antiques that very quickly.

2. Related work

Road crack detection is the method of detecting defects in roads caused by a variety of factors such as heavy rainfall, weather changes road maintenance, natural disasters, and so on, In paper [1]published year **2018** Author et.ai Hyoungskwaniwas proposed topic is "Encoderi–decodery network for pixel-level road crack detection in blacky-box images" He used technique Deep learning CNN, encoder and decoder network and proposed It is the first time a deep completely convolutional encoder–decoder network has been used to detect cracks in black box images but it drawback its accuracy is not better, In paper [2] published year **2019** Author et.ai Yuxiangg Sun He was proposed topic is "Accuratel Lane Detection with Atrous Convolution and Spatial Pyramid Pooling for Autonomous Driving "He was using technique Deep learning Spatial CNN/ANN and it was finding in his research was He did excellent work on deep learning-based semantic segmentation and lane detection using semantic segmentation technologies, both of which are applicable to this article. but it has some drawback about its accuracy, In paper [3] published year **2017** Authors et.ai Caos Vuu Dung He was proposed topic "Autonomous concrete crack detection using deep fully convolutional neural network" and used technology CNN, FCNN and He discovered A vision-based approach for his topic and using FCN was suggested, In paper [4] **2020** Author et.ai Dharneeskhar Dhakshana

was proposed topic was "Deep learning based detection of potholes in indian road using yolo" he used technology YOLO (You can Look once)And he was finding Potholes detection implemented using one of the CNN family unique and Yolo it was one of the best model his accuracy is too fast but it's some disadvantage is that it was more costly and it model is very lenthy it was required experts, In paper [5] published year **2017** Author et.ai Allen Zhang wang was developedhim topic was "Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces" Using a Deep-Learning Network" used technology CNN, 3D Asphalt and he did experiment using 200 testing 3D images revealed that Crack Net can achieve high Precision (90.13 percent), Recall (87.63 percent), and Measure (88.86 percent) all at the same time, but it had the drawback of not being able to achieve more accuracy because it used old technology, In paper [6] published year **2020** Author et.ai Shim seungbocho was developed topic "Light weight semantic segmentation for road surface" used technology ResNet50,DetNet59 and Dense Net and he was proposed in his model Proposed a new neural network algorithm for road surface damage object recognition, In paper [7] published year **2020** Author et.ai Rovetta It established his topic "Audio surveillance Deep learning auto encoder decoder based sample recognition" He used technology MLP deep neural network he was finding State audio event, In paper [8] published year **2019** Author et.ai Colagrande tanalli his topic "Gpr research on damaged road pavement built in cut and fill section" he was used GPR (ground penetrating radar) he was proposed The evaluation of the attenuation of the radar signal detected from the pavement. Here is some drawback that accuracy is little bit poor, In paper [9] published year **2019** Author et.ai Nickola slavkovic his topic is "Risk prediction algorithm based on image texture extraction using mobile vehicle road scanning system as support for autonomous driving" and it used texture classification tool ad it was proposed Using a mobile vehicle, a risk prediction algorithm based on image texture extraction. In paper [10] **2016** Author et.ai Tom B.J topic name was "A review on automated pavement distress detection methods" He used tool SVM and It proposed The scope includes the degree to which distresses and external factors need to be captured For instance. In paper [11]   published year 2018 Author et.ai Mark David Jenkins established his topic "A Deep Convolutional Neural Network for Semantic Pixel-Wise Segmentation using CNN method and it proposed A Semantic Pixel-Wise Segmentation of Road and Pavement Surface Cracks Using a Deep Convolutional Neural Network,In paper [12] Author et.ai Fan Yang his topic name "Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection" using tool FCNN and VGG and he proposed The proposed method is based on the famous Caffe library and an open FCN implementation. The conv1-conv5 of pre-trained VGG is the bottom-up component,

In paper [13] publish year **2020** Author et.ai Qipei Mei his topic name "Densely connected deep neural network considering connectivity" using tool CNN, In paper [14] **2014** Author et.ai IH. Bello-Salau published his topic "Road Defect Detection", In paper [15] **2016** Author et.ai Emanuel Aldea introduced his topic name "Robust crack detection for unmanned aerial vehicles inspection" and using tool CNN, In paper [16], 2015 Author et.ai Shivprakash and his topic was "A robust approach for automatic detection and segmentation" it used technique mathematical morphology, In paper [17] published year **2018** Author et.ai Lin Li 1 topic was "Lane Marking Detection and Reconstruction" using SVM (super vector machine) Technology, In paper [18] published Year 2019 Author et.ai Fred Daneshgaran and his topic was "Use of Deep Learning for Automatic Detection of Cracks in Tunnels" using tool CNN

3. Finding of Survey

This work on Road crack Detection using so many techniques like Ground Penetrating Radars (GPR Internet of Things (IOT), Vehicular Ad-hoc Network Technology (VANET) , GPS systems, it was discussed in literature survey. Following a study of the approaches, it was determined that image pre-processing devices are the most powerful and gainful In this paper CNN AND FCNN deep learning technique is utilized for Road crack detection for autonomous driving

4. Road crack methods are discussed

The proposed work is to inspect crack detection methods in which images are taken from a camera while a fixed GPS monitors, the device will take input images and the images will be converted from RGB (Red-Green-Blue) to grayscale, so the system is very flexible, reliable and cost-effective., Recent improvement in object detection is based on CNN where Faster Region-based CNN (Faster-RCNN) is the basis of object detection methods. In This paper we present image pre-processing technique for road crack detection method focusing on faster-region convolution neural network(Faster-CNN)method, that represent the work of my road crack detection model, Firstly I will take image normal and potholes images and then image pre-processing of those images because in data there is more noisy data, irrelevant data when I preprocess the images all the noisy data will remove all the images come in same size than I will load model which I create than classify the image either there is crack or not and result will be show with better accuracy concludes.
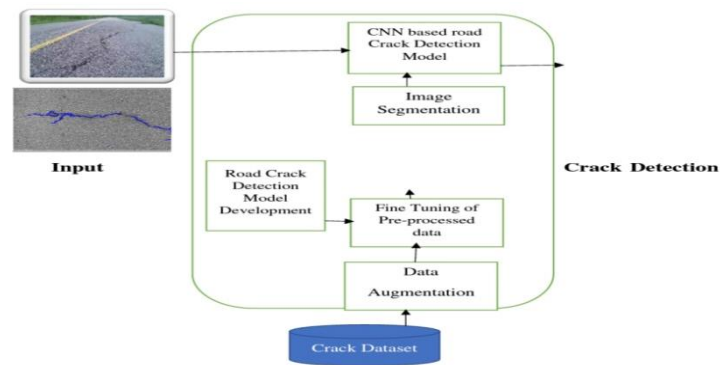
Figure 6 *Architecture diagram*

Image Preprocessing

The above architectural diagram consist of 6 phases that are described below for my Research implementation [1] **Data Collection:** First pictures from open source picture datasets will be gathered to prepare the (CNN) "Such image datasets consist of different kinds of images, such as camera images, laser images, etc".[2] **Data Augmentation:** In demand to accomplish great execution results, fix will be created for each example picture utilizing some testing procedures, for example, an irregular point pivot between 0o to 360o, low cover level between two positive patches, and so on "a picture with break will show a Real (1) in the picture marking measure, and correspondingly a picture without break will show a False (0)". [3]**Fine Tuning:** To train a CNN model without any preparation, the current data sets are not enormous enough. The recently prepared models through exchange learning "this project will therefore fine-tune such as ResNet-50, ResNet-101, VGG-16, VGG 19 SegNet or ImageNet its all pre-trained model".[4] **Feature Map Generation:** Which will be pooled utilizing a delicate max layer of CNN to produce the grouping results for these tweaked models will create a component map. [5] **Classification:** This task will utilize a managed learning-based grouping procedure to order the pictures as pictures with Crack and pictures without Crack into two classes "Autonomous Driving Cars In order to apply the proposed approach to the real-time videos of in order to locate road cracks for the test process, this project will use this view to collect frames from the videos"[6]**Crack Segmentation:** We will attempt to fragment the picture into districts utilizing usable picture division procedures and discover the territory that has broken. The most recent strategies for picture division incorporate locale based division, division of edge location, grouping based division, division dependent on feebly managed Convolutionary Neural Network learning, pressure based techniques and pressure strategies.

5. Conclusion

we conclude that Road crack detection for autonomous driving system we are taking th benefit of CNN Faster-RCNN technique is that it works image data and is also applicable to identify accurately it plays an important role in an industrial city to solve the Road Crack for autonomous driving problem we are using excellent technology which is very effectively and low cost and working on a large amount of dataset and reduce accident because of potholes After brief study issues of Road crack and existing solutions with a particular focus on deep learning we enhance this concept in future to apply in vehicle maintenance

References

[1] Seongdeok Bang, Somin Park, Hongjo Kim and Hyoungkwan Kim, "Encoder–decoder network for pixel-level road crack detection in black-box images," National Research Foundation of Korea; Ministry of Science and ICT, Grant/Award Number: 2018R1A2B2008600; Ministry of Education, Grant/Award Number: 2018 R1A6A1A08025348, doi: org/10.1111/mice.12440.

[2] Yuxiang Sun, Lujia Wang, Yongquan Chen and Ming Liu, "Accurate Lane Detection with Atrous Convolution and Spatial Pyramid Pooling for Autonomous Driving," Proceeding of the IEEE International Conference on Robotics and Biomimetics Dali, China, December 2019 doi: 10.1109/ROBIO49542.2019.8961705.

[3] Cao Vu Dung and Le Duc Anh, "Autonomous concrete crack detection using deep fully convolutional neural network," Received 6 July 2018; Received in revised form 28 October 2018 Accepted 29 November 2018 0926-5805/ © 2018 Elsevier B.V. All rights reserved doi: 10.1016/j.autcon.2018.11.028.

[4] Dharneeshkar J, Soban Dhakshana, V, Aniruthan SA, Karthika R and Latha Parameswaran," Deep learning based detection of potholes in indian road using YOLO," Proceedings of the Fifth International Conference on Inventive Computation Technologies (ICICT-2020) IEEE Xplore Part Number:CFP20F70 doi: 10.1109/ICICT48043.2020.9112424.

[5] Allen Zhang, Kelvin C. P. Wang, Baoxian Li, Enhui Yang, Xianxing Dai, Yi Peng, Yue Fei, Yang Liu, Joshua Q.Li and Cheng Chen, "Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces," Engineering (2017) 1–15 Computer-Aided Civil and Infrastructure doi: 10.1111/mice.12297.

[6] Seungbo Shim and Gye-Chun Cho, "Lightweight semantic segmentation for road surface damage recognition," Received May 14, 2020 accepted May 26, 2020 date of publication May 29, 2020, date of current June 2020 doi: 10.1109/ACCESS.2020.2998427.

[7] Rovetta, Stefano masulli, Francesco and Mnasri zaed, "Audio surveillance of mnasri zaed, using Deep learning auto encoder decoder based sample recognition," The 15th IEEE International Conference on Electronics, Circuits and Systems (ICECS 2008) doi: 10.1112/ICECS345.4332.

[8] S.Colagrande, D.Ranalli and M.Tallini, "GPR research on damaged road pavement built in cut and fill section," AIIT 2nd International Congress on Transport Infrastructure and Systems in a changing world (TIS ROMA 2019), 23rd-24th September 2019, Rome, Italy doi: 10.1016/j.trpro.2020.02.059.

[9] Nickola slavkovic, Milan Bjelica, "Risk prediction algorithm based on image texture extraction using mobile vehicle road scanning system as support for autonomous driving," proceeding the J. of Electronic Imaging, (2019) doi: 10.1117/1.JEI.28.3.033034.

[10] Tom B.J. Coenen and Amir Golroo, "A review on automated pavement distress detection methods," published in 05/09/2017 doi: 10.1080/23311916.2017.1374822.

[11] Mark David Jenkins, Thomas Arthur Carr, Maria Insa Iglesias, Tom Buggy and Gordon Morison, "Deep Convolutional Neural Network for Semantic Pixel-Wise Segmentation of Road and Pavement Surface Cracks where use tool encoder and decoder network" 26th European Signal Processing Conference 2018

doi: 10.23919/EUSIPCO.2018.8553280.

[12] Fan Yang , Lei Zhang, Sijia Yu, Danil Prokhorov, Xue Mei and Haibin Ling, "Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection," Yang, Hang, F., L., YU,S and Ling, H. (2019) Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection. IEEE Transactions on Intelligent system doi: 10.1109_TITS.2019.2910595.

[13] Qipei Mei, Mustafa Gul and Md RiasatAzim, "Densely connected deep neural network considering connectivity of pixels for automatic crack detection using tool CNN," Mei, Q., Gül, M., & Azim, M. R. autcon.2019.103018 doi: 10.1016/j.autcon.2019.103018.

[14] H. Bello-Salau, 2A. M. Aibinu, 3E. N. Onwuka, 4J. J. Dukiya and sA. J. Onumanyi , "Road Defect Detection" A Survey and it used Vehicular Adhoc Network Technology (V ANET," A survey. 2014 11th International Conference on Electronics, Computer and Computation (ICECCO) doi:10.1109/icecco.2014.6997556.

[15] Emanuel Aldea and Sylvie Le Hégarat-Mascle, "crack detection for unmanned aerial vehicles inspection in an a-contrario decision framework," Aldea, E., & Le Hégarat-Mascle, S. (2015) Robust crack detection for unmanned aerial vehicles inspection in ana-contrariodecision framework. Journal of Electronic Imaging doi:10.1117/1.jei.24.6.061119.

[16] Shivprakash Iyer and Sunil K. Sinha, "A robust approach for automatic detection and segmentation of cracks in underground pipeline images," Iyer, S., & Sinha, S. K. (2005) A robust approach for automatic detection and segmentation of cracks in underground pipeline images Image and Vision Computing, 23(10), 921–933. doi:10.1016/j.imavis.2005.05.017 doi: 10.1016/j.imavis.2005.05.017.

[17] Lin Li 1, Wenting Luo 1 ID and Kelvin C. P. Wang, "Lane Marking Detection and Reconstruction with Line-Scan Imaging Data," Li, L., Luo, W., & Wang, K. C. P. (2018) 18(5), 1635. doi:10.3390/s18051635.

[18] Fred Daneshgaran1 , Luca Zacheo2 , Francesco Di Stasio2 , and Marina Mondin, "Use of Deep Learning for Automatic Detection of Cracks in Tunnels: Prototype-2 Developed in the 2017–2018 Time Period" using tool

CNN," Daneshgaran, F., Zacheo, L., Stasio, F. D., & Mondin, M. (2019) Transportation Research Record: Journal of the Transportation Research Board, 036119811984565. doi:10.1177/0361198119845656.

CMDSA-031
SEEK – SEAT

Rahmatunnisa Khatoon

Muffakham Jah College of Engineering and Technology

160417733067@mjcollege.ac.in

Abstract: *An attempt is being made to design a project which gives the information about different colleges based on the rank of the student. Seek-Seat helps the students to find not only the right college but also the courses suitable for their bright future. This project provides user with all the information required about the colleges. In this project, when a student intends to find a college based on their rank, it provides the data based on branch provided in required course and based on college. When the user seeks information about the branch it gives the list of colleges with the starting rank and cut off ranks for the required branch where as in the college option it displays the courses provided by the college with the preceding years cut off ranks as every new beginning comes from some other beginning's end. This project also provides an option for the students who wants a management seat in the college. Apart from this it also provides an option for different courses of bachelor's degree too in various colleges who want to opt the graduation courses instead of engineering seat.*

In total it is a project which helps people in avoiding unnecessary stress which is caused in selecting courses.

Keywords: *Object Oriented Programming, Counseling, Ranks*

1. Introduction

Generally, we see students getting panicked and stressed while choosing a college for graduation or engineering which plays a crucial role in their career path. They seem confused about selecting the branch in a college or college in their preferred branch based on their score or rank as they have no idea about college intake or cut off ranks

As of now there are college predictors available but most of them require registration or payment too. They do not give students a choice of selecting branch in their preferred college. They can just find college based on preferred branch. There is no information provided for students who cannot seek college in engineering for other graduation courses.

Choosing right college is a crucial step in student's career path. Many students get college education to better their chances of getting a good job. But as most of the colleges are expensive, it is important for students and their parents to see a good return on their investment.

Even though there are websites available like career360, siksha.com which predicts the college based on your preferred college still there is no option for students to find a branch if

they want to go to a particular college and there is no information provided for non-engineering students who aspire courses like B.Com, B.Sc etc. And all these sites implement the program using complex languages

In order to deal with the problem at hand, we have built this project of "Seek seat" which provides all the information about the required college and branch. It gives complete idea about the college intake and cut off ranks .Moreover the students can get information about other under graduation courses if they cannot find an engineering college of their choice or who want to pursue graduation course. Apart from these advantages, this project solves the problem in very simple way.

2. Materials and Methods

Object Oriented Programming Concepts

Object-oriented programming – As the name suggests uses objects in programming. Object-oriented programming aims to implement real-world entities like inheritance, hiding, polymorphism, etc in programming. The main aim of OOP is to bind together the data and the functions that operate on them so that no other part of the code can access this data except that function.
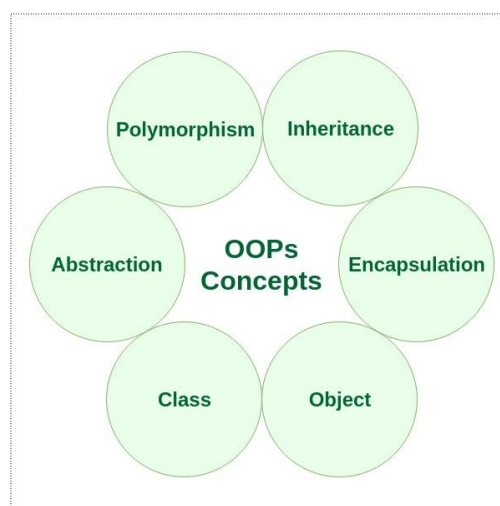
Figure 7: Characteristics of an Object Oriented Programming language

Class: The building block of C++ that leads to Object-Oriented programming is a Class. It is a user-defined data type, which holds its own data members and member functions, which can be accessed and used by creating an instance of that class. A class is like a blueprint for an object.

For Example: Consider the Class of Cars. There may be many cars with different names and brand but all of them will share some common properties like all of them will have 4 wheels, Speed Limit, Mileage range etc. So here, Car is the class and wheels, speed limits, mileage are their properties.

A Class is a user-defined data-type which has data members and member functions.

Data members are the data variables and member functions are the functions used to manipulate these variables and together these data members and member functions define the properties and behavior of the objects in a Class.

In the above example of class Car, the data member will be speed limit, mileage etc and member functions can apply brakes, increase speed etc.

We can say that a **Class in C++** is a blue-print representing a group of objects which shares some common properties and behaviors.

Object: An Object is an identifiable entity with some characteristics and behavior. An Object is an instance of a Class. When a class is defined, no memory is allocated but when it is instantiated (i.e. an object is created) memory is allocated.

Object take up space in memory and have an associated address like a record in pascal or structure or union in C.

When a program is executed the objects interact by sending messages to one another.

Each object contains data and code to manipulate the data. Objects can interact without having to know details of each other's data or code, it is sufficient to know the type of message accepted and type of response returned by the objects.

Encapsulation: In normal terms, Encapsulation is defined as wrapping up of data and information under a single unit. In Object-Oriented Programming, Encapsulation is defined as binding together the data and the functions that manipulate them.

Consider a real-life example of encapsulation, in a company, there are different sections like the accounts section, finance section, sales section etc. The finance section handles all the financial transactions and keeps records of all the data related to finance. Similarly, the sales section handles all the sales-related activities and keeps records of all the sales. Now there may arise a situation when for some reason an official from the finance section needs all the data about sales in a particular month. In this case, he is not allowed to directly access the

206

data of the sales section. He will first have to contact some other officer in the sales section and then request him to give the particular data. This is what encapsulation is. Here the data of the sales section and the employees that can manipulate them are wrapped under a single name "sales section".
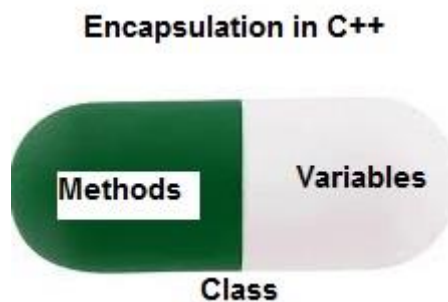
Figure 8: Encapsulation in C++

Encapsulation also leads to **data abstraction or hiding**. As using encapsulation also hides the data. In the above example, the data of any of the section like sales, finance or accounts are hidden from any other section.

Abstraction: Data abstraction is one of the most essential and important features of object-oriented programming in C++. Abstraction means displaying only essential information and hiding the details. Data abstraction refers to providing only essential information about the data to the outside world, hiding the background details or implementation.

Consider a real-life example of a man driving a car. The man only knows that pressing the accelerators will increase the speed of the car or applying brakes will stop the car but he does not know about how on pressing accelerator the speed is actually increasing, he does not know about the inner mechanism of the car or the implementation of accelerator, brakes etc in the car. This is what abstraction is.

- **Abstraction using Classes**: We can implement Abstraction in C++ using classes. The class helps us to group data members and member functions using available access specifiers. A Class can decide which data member will be visible to the outside world and which is not.

- **Abstraction in Header files**: One more type of abstraction in C++ can be header files. For example, consider the pow() method present in math.h header file. Whenever we need to calculate the power of a number, we simply call the function pow() present in the

math.h header file and pass the numbers as arguments without knowing the underlying algorithm according to which the function is actually calculating the power of numbers.

Polymorphism: The word polymorphism means having many forms. In simple words, we can define polymorphism as the ability of a message to be displayed in more than one form. A person at the same time can have different characteristic. Like a man at the same time is a father, a husband, an employee. So the same person possess different behavior in different situations. This is called polymorphism.

An operation may exhibit different behaviors in different instances. The behavior depends upon the types of data used in the operation.

C++ supports operator overloading and function overloading.

- **Operator Overloading**: The process of making an operator to exhibit different behaviours in different instances is known as operator overloading.
- **Function Overloading**: Function overloading is using a single function name to perform different types of tasks. Polymorphism is extensively used in implementing inheritance.
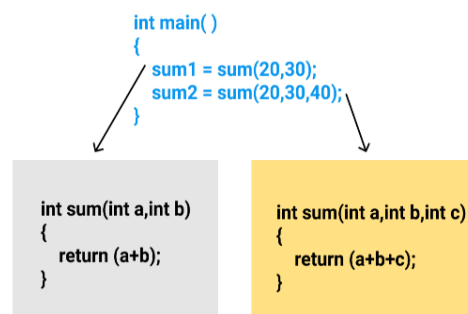
Figure 9: Example of Function Overloading

Example: Suppose we have to write a function to add some integers, sometimes there are 2 integers, sometimes there are 3 integers. We can write the Addition Method with the same name having different parameters, the concerned method will be called according to parameters.

Inheritance: The capability of a class to derive properties and characteristics from another class is called Inheritance. Inheritance is one of the most important features of Object-Oriented Programming.

- **Sub Class**: The class that inherits properties from another class is called Sub class or Derived Class.

- **Super Class**:The class whose properties are inherited by sub class is called Base Class or Super class.

- **Reusability**: Inheritance supports the concept of "reusability", i.e. when we want to create a new class and there is already a class that includes some of the code that we want, we can derive our new class from the existing class. By doing this, we are reusing the fields and methods of the existing class.

Example: Dog, Cat, Cow can be Derived Class of Animal Base Class.
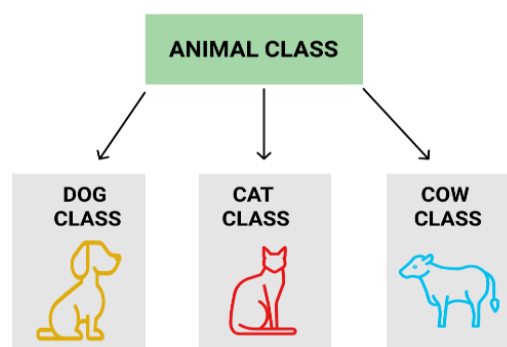
Figure 10: *Describes Inheritance*

Files

Database File System (DBFS) creates a standard file system interface on top of files and directories that are stored in database tables.

Database File System (DBFS) creates a standard file system interface using a server and clients.

DBFS is similar to NFS in that it provides a shared network file system that looks like a local file system and has both a server component and a client component.

At the core of DBFS is the DBFS Content API, a PL/SQL interface in the Oracle Database. It connects to the DBFS Content SPI, a programmatic interface which allows for the support of different types of storage.

At the programming level, the client calls the DBFS Content API to perform a specific function, such as delete a file. The DBFS Content API *delete file* function then calls the DBFS Content SPI to perform that function.

In a user-defined DBFS, the user must implement a delete function based on the specifications in the DBFS Content SPI, along with other functions in the specification.
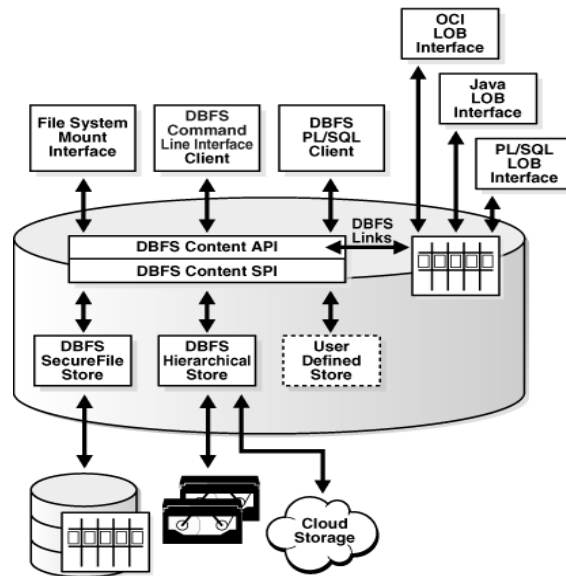
Figure 11: *Architecture of DBFS*

3. Results and Discussion
MODULES:

The project basically has four modules. They are:

Admin Module:

The admin can change the information about the colleges provided in the files.

User Module:

The user should enter the rank to seek the required information about the colleges.

Branch wise Module:

In this user can surf different colleges according to the required branch.

College wise Module:

In this user can search the branches according to the required college by giving the rank as an input.

Other options:

The user can find all the required information about the different courses for which he/she can apply with the main stream subjects.

The innovative Seek seat project is designed for aspirants, which works on the previous year data and trends. With the help of this project and probable rank, the qualified aspirants can check list of colleges they can get admission into. The best part about this project is that it is being made available for everyone and doesn't require any registration or payment. Student can easily get required information in less time.

In Figure 6, the screenshots of the project are displayed where the users can find the college according to the rank or see if the users can be admitted in the dream college or see for the required information regarding other courses available to take as a main stream.
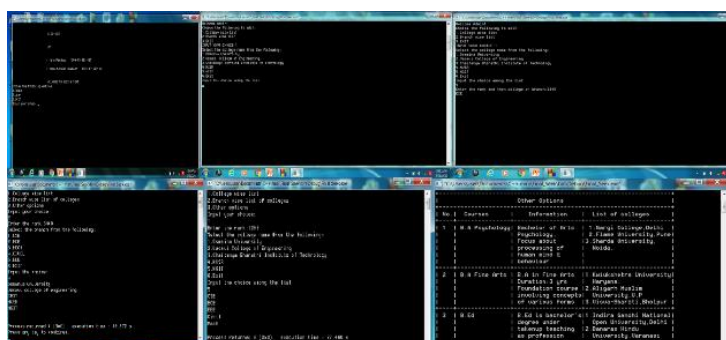
Figure 12: Screenshots of project outputs

4. Conclusion

The Seek Seat project is built using C++ which completely meets the objective of the system which has been developed. It is very helpful to find the information about different colleges based on the rank of the student. Seek-Seat helps the students to find not only the right college but also the courses suitable for their bright future. It overcomes many limitations provided by the existing system and is easy to use due to its user friendly interface.

5. References

1. https://www.geeksforgeeks.org/object-oriented-programming-in-cpp/
2. https://docs.oracle.com/database/121/ADLOB/adlob_fs.htm
3. https://engineering.careers360.com/articles/ts-eamcet-cutoff
4. https://www.collegedekho.com/exam/ts-eamcet/cutoff

6. Copyright

7. Disclosure Policy

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

SCHOOL – BOT

Afroz Samee

Muffakham Jah College of Engineering and Technology

160417733063@mjcollege.ac.in

Abstract: *An attempt is being made to design a project, it is important for the parent to be familiar with the status of their child how they are performing in the academics. Most of the times parents face trouble dealing with their children. To make this hassle free, there is an option available for the parents. Where the options are available accordingly and the selected option gives the student information. The student information like report card, attendance, marks, time table, events which are monitored by the parents using School-Bot. Our project makes use parents to access student information in a smarter way. The data of the students and management are given by admin who controls and updates the formation. Admin has the power to create, update or delete any record of the system.*

Keywords: *Assessment, Object Oriented Programming (OOPs)*

1. Introduction

The main objective of our project School-Bot is to reach out to the users (parents), admin and in turn help the parents. This project helps the admin store all the data of the school in his files. It allows him to keep a check on the data and thereby achieves its objective of supplying a faster service to user.

The problem that is being faced as of now it is important for the parent to be familiar with the status of their child how they are performing in the academics. Most of the times parents face trouble dealing with their children. As of now there are school management software available in the market but most of them either have copyrights of the organizations or they are very costly another drawback is that this software can be installed only in our personal computers or laptops which does not help the user everywhere.

In order to deal with the problem at hand, we have suggested the use of a Code Block's code. This system is free of cost and can be used anywhere and it has almost all the options of the software available in the market. It can help the majority of parents i.e., user by performing search operation, and help keep check on the school updates hence improving students, teachers and parents' relationship by a great deal.

2. Materials and Methods:

To build this system we have used OOPs concepts with C++ Programming. And for data to store we have used files. This uses objects in programming. Object-oriented programming aims to execute real-world entities such as inheritance, hiding, polymorphism, etc. The main focus of OOP is to bind the data and functions together.

3. Results and Discussion

In this system we built the school-bot since the quality of time to spent on specific thing is reduced by using this. we are making a setup or a system that can do all the functions of the software like storing the student, teacher data in an offline database, keep a check on the availability of requests, complains, give advices regarding the school changes and it also has an additional feedback option to complain or comment option which is provides. in this model the parents have authority and help in guidance to children and their requirement, which help the user and admin to get fast and efficient service.

Below are the results of this model.

Figure 13: Screenshots of Project outputs

4. Conclusions

The school data management resource is developed using code blocks and c++ which completely meets the objective of the system which has been developed. It is very helpful to store the information at one place and keep a check on it which helps the parents i.e., users a lot. It overcomes many limitations provided by the existing system and is to use due to its user-friendly interface.

5. References

1. https://www.geeksforgeeks.org/object-oriented-programming-in-cpp/
2. https://docs.oracle.com/database/121/ADLOB/adlob_fs.htm
3. https://engineering.careers360.com/articles/ts-eamcet-cutoff

6. Copyright

7. Disclosure Policy

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Glimpses of the event

Below are some of the beautiful moments from the Event

Media Coverage

The event has been covered by media, and was live telecasted on YouTube, can be accessed from the links below. Some of the newspaper clippings can also be found in this section

Day 1: Inaugural session: https://youtu.be/CW3d6ujCMjI

Day 1: Paper presentations: https://youtu.be/4GfZYwn_YZc

Day 2: Technical talks session: https://youtu.be/mhodxeXxu2Y

Day 2: Paper presentations: https://youtu.be/MTwiCIZma2w

نئی تعلیمی پالیسی میں سائنس و ٹکنالوجی کے فروغ پر خصوصی توجہ

اردو یونیورسٹی میں ڈیٹا سائنس پر آن لائن قومی کانفرنس کا انعقاد

حیدرآباد۔25مئی (پریس نوٹ) قومی تعلیمی پالیسی 2020 میں بین شعبہ جاتی تعلیم، مالیہ کی فراہمی اور تعلیمی اداروں کی تشکیح کے ساتھ سائنس و ٹکنالوجی کے فروغ پر خصوصی توجہ دی گئی ہے۔ ان خیالات کا اظہار پروفیسر ایس ایم رحمت اللہ، وائس چانسلر انچارج نے کل مولانا آزاد نیشنل اردو یونیورسٹی میں شعبہ کمپیوٹر سائنس و آئی ٹی کے زیر اہتمام دو روزہ آن لائن قومی کانفرنس کے افتتاحی اجلاس میں صدارتی خطاب کے دوران کیا۔ اس کانفرنس کا موضوع "کمپیوٹیشنل میتھڈس، ڈیٹا سائنس اینڈ اپلیکیشنز" تھا۔ پروفیسر رحمت اللہ نے تعلیمی پالیسی میں سائنس اور ٹکنالوجی کے فروغ سے متعلق نکات کا تفصیلی جائزہ لیا۔ شعبہ کمپیوٹر سائنس و آئی ٹی، ریگولر اور فاصلاتی دونوں طرز کی تعلیم میں نئی پالیسی پر عمل آوری میں یونیورسٹی کیلئے معاون ثابت ہوگا۔ اس موقع پر خطاب کرتے ہوئے مہمان خصوصی پروفیسر عبدالقیوم انصاری، شعبہ الیکٹریکل انجینئرنگ، جامعہ ملیہ اسلامیہ، نئی دہلی نے تدریس اور سیکھنے کے عمل میں ہونے والی تبدیلیوں اور آئی ٹی کے وسیع تحقیقی اپلی کیشن پر کمپیوٹیشنل ماڈلز کے حوالے سے اظہار خیال کیا۔ انہوں نے طلبہ کی صلاحیتوں کو پہچان کران کی حوصلہ افزائی کی ضرورت پر زور دیا اور اخلاص اور یقین کے ساتھ کام کرنے کی تلقین کی۔ ڈاکٹر سید امتیاز حسن، صدر شعبہ سی آئی ٹی نے شرکاء کا خیر مقدم کیا۔ قرأت کلام پاک سے جلسہ کا آغاز ہوا۔ کانفرنس کے آج دوسرے دن پروفیسر ڈاکٹر خالد رضا، شعبہ کمپیوٹر سائنس، جامعہ ملیہ اسلامیہ؛ ڈاکٹر محمد زکوان، مانو پالی ٹیکنیک حیدرآباد، و دیگر نے مقالے پیش کیے۔ پروفیسر سلمان عبدالمعز، اسکول آف کمپیوٹر اینڈ انفارمیشن سائنسس، یونیورسٹی آف حیدرآباد اختتامی اجلاس کے مہمان اعزازی تھے۔ ڈاکٹر پردیپ کمار نے کانفرنس کی رپورٹ پیش کی۔ کانفرنس میں 40 تحقیقی مقالے پیش کیے گئے۔

టెక్నాలజీపై ఎన్ఈపీలో ప్రత్యేక దృష్టి

● 'మను' ఇన్చార్జి వీసీ ప్రొ. రహమతుల్లా

రాయదుర్గం: జాతీయ విద్యావిధానంలో ఇంటర్ డిసిప్లినరీ విధానంతో సైన్స్ అండ్ టెక్నాలజీని ప్రోత్సహించడంపై ప్రత్యేక దృష్టి పెట్టారని మౌలానా ఆజాద్ జాతీయ ఉర్దూ విశ్వవిద్యాలయం ఇన్చార్జి వైస్చాన్స్లర్ ప్రొ. ఎస్ఎం రహమతుల్లా పేర్కొన్నారు. గచ్చిబౌలిలోని ఉర్దూ విశ్వవిద్యాలయంలోని స్కూల్ ఆఫ్ టెక్నాలజీ, కంప్యూటర్ సైన్స్ అండ్ ఇన్ఫర్మేషన్ టెక్నాలజీ విభాగం ఆధ్వర్యంలో గణన పద్ధతులు, డేటాసైన్స్పై జాతీయ స్థాయి ఆన్లైన్ సమావేశాన్ని నిర్వహించారు.

◆ ఈ సందర్భంగా రహమతుల్లా మాట్లాడు తూ... రెగ్యులర్తోపాటు డిస్టెన్స్ మోడ్

కోర్సులకు ఎన్ఈపీ అదేస్తాయిలో ప్రాధా న్యత ఇచ్చారన్నారు. న్యూడిల్లీలోని జేఎంఐ ప్రొ. అబ్దుల్ ఖుయ్యూమ్ అన్సారి మాట్లాడు తూ.. బోధన, అభ్యాస ప్రక్రియలో పరి జ్ఞానాన్ని, గణన నమూనాలో ఐటీ విస్తృత పరిశోధన అనువర్తనాలను వివరించారు. విద్యార్థులు తమ సామర్థ్యాన్ని గ్రహించాల న్నారు. కార్యక్రమంలో స్కూల్ ఆఫ్ టెక్నాలజీ డీన్ ప్రొ. అబ్దుల్వాహిద్, జేఫాత్ అకాడమిక్ కాలేజ్ అధ్యాపకులు డాక్టర్ మాలిక్ యూసఫ్, సదస్సు కన్వీనర్ డాక్టర్ ప్రదీప్కుమార్, సీని యర్ అసిస్టెంట్ ప్రొఫెసర్ డాక్టర్ ఖాలిదర జా, డాక్టర్ మహమూద్ జక్వాన్, ప్రొఫెసర్ సల్మాన్ అబ్దుల్ మోయిజ్ ప్రసంగించారు.

States »

South

Posted at: May 25 2021 5:45PM

🔵🔵🟢➕

NEP lays special focus on Technology: Prof Rahmatullah

Hyderabad, May 25 (UNI) National Education Policy (NEP) lays special focus on promotion of Science and Technology with an interdisciplinary approach.

The School of Computer Sciences & Information Technology at MANUU will play an important role in implementing the NEP for both regular and Distance mode courses.

Prof S M Rahmatullah, Vice Chancellor Incharge, expressed these views on Monday while delivering the presidential address in the inaugural session of two day National conference "Computational methods, Data Science and Applications" organized by the School of Technology, Department of Computer Science and Information Technology, Maulana Azad National Urdu University (MANUU).

The inaugural was held virtually through Zoom & MANUU YouTube channel.

The Chief Guest, Prof Abdul Quaiyum Ansari, Faculty of Engineering, department of Electrical Engineering, Jamia Millia Islamia (JMI), New Delhi briefly shared the evolution in the teaching