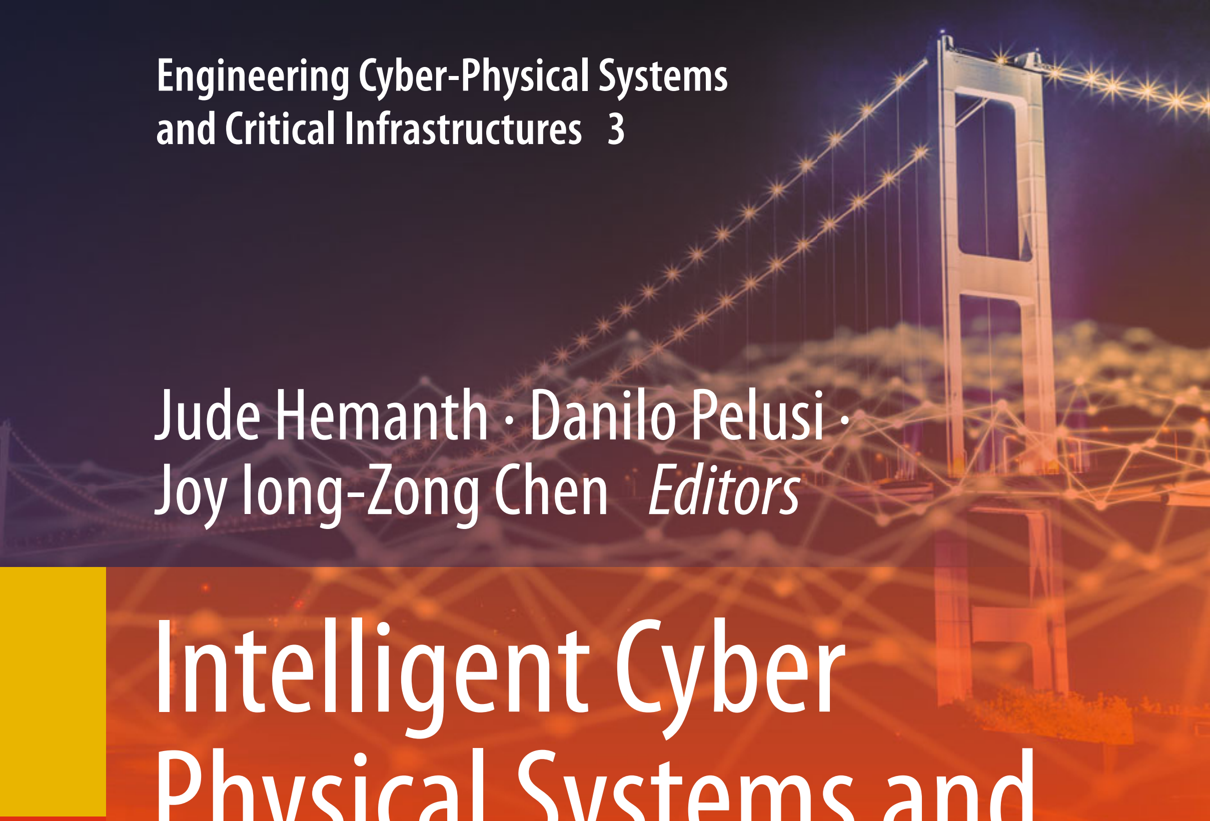


Engineering Cyber-Physical Systems and Critical Infrastructures 3

Jude Hemanth · Danilo Pelusi · Joy Iong-Zong Chen *Editors*

Intelligent Cyber Physical Systems and Internet of Things

ICoICI 2022

Springer

Engineering Cyber-Physical Systems and Critical Infrastructures

Volume 3

Series Editor

Fatos Xhafa, Departament de Ciències de la Computació, Technical University of Catalonia, Barcelona, Spain

The aim of this book series is to present state of the art studies, research and best engineering practices, real-world applications and real-world case studies for the risks, security, and reliability of critical infrastructure systems and Cyber-Physical Systems. Volumes of this book series will cover modelling, analysis, frameworks, digital twin simulations of risks, failures and vulnerabilities of cyber critical infrastructures as well as will provide ICT approaches to ensure protection and avoid disruption of vital fields such as economy, utility supplies networks, telecommunications, transports, etc. in the everyday life of citizens. The intertwine of cyber and real nature of critical infrastructures will be analyzed and challenges of risks, security, and reliability of critical infrastructure systems will be revealed. Computational intelligence provided by sensing and processing through the whole spectrum of Cloud-to-thing continuum technologies will be the basis for real-time detection of risks, threats, anomalies, etc. in cyber critical infrastructures and will prompt for human and automated protection actions. Finally, studies and recommendations to policy makers, managers, local and governmental administrations and global international organizations will be sought.

Jude Hemanth · Danilo Pelusi ·
Joy Iong-Zong Chen
Editors

Intelligent Cyber Physical Systems and Internet of Things

ICoICI 2022

Springer

Editors
Jude Hemanth
Department of Electronics
and Communication Engineering
Karunya Institute of Technology
and Sciences
Coimbatore, Tamil Nadu, India

Danilo Pelusi
Faculty of Communication Sciences
University of Teramo
Teramo, Italy

Joy Iong-Zong Chen
Department of Electrical Engineering
Da-Yeh University
Dacun, Changhua, Taiwan

ISSN 2731-5002 ISSN 2731-5010 (electronic)
Engineering Cyber-Physical Systems and Critical Infrastructures
ISBN 978-3-031-18496-3 ISBN 978-3-031-18497-0 (eBook)
https://doi.org/10.1007/978-3-031-18497-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

We are honored to dedicate the proceedings of ICoICI 2022 to all the participants and editors of ICoICI 2022.

Preface

It is with deep satisfaction that I write this preface to the proceedings of the ICoICI 2022 held in JCT College of Engineering and Technology, Coimbatore, Tamil Nadu, India, during August 11–12, 2022.

This conference proceedings volume contains the written versions of most of the contributions presented during the conference of ICoICI 2022. The conference provided a setting for discussing recent developments in a wide variety of topics including Cyber-Physical Systems, Data Communication, Computer Networking, Communicational Technologies, Cryptography, Big Data, Cloud Computing, IoT, and Healthcare Informatics. The conference has been a good opportunity for participants coming from various destinations to present and discuss topics in their respective research areas.

ICoICI 2022 conference tends to collect the latest research results and applications on Intelligent Cyber Physical Systems and Internet of Things. It includes a selection of 65 papers from 302 papers submitted to the conference from universities and industries all over the world. All of accepted papers were subjected to strict peer-reviewing by 2–4 expert referees. The papers have been selected for this volume because of quality and the relevance to the conference.

ICoICI 2022 would like to express our sincere appreciation to all authors for their contributions to this book. We would like to extend our thanks to all the referees for their constructive comments on all papers, especially, we would like to thank to committee members for their hard working. Finally, we would like to thank the Springer publications for producing this volume.

Guest Editors—ICoICI 2022

Coimbatore, India
Teramo, Italy
Dacun, Taiwan

Prof. Dr. Jude Hemanth
Dr. Danilo Pelusi
Prof. Dr. Joy Iong-Zong Chen

Contents

Term Frequency Tokenization for Fake News Detection

Pallavi Suresh, Abhishek Shettigar, M. Karunavathi, Ajith, and M. G. Ramanath Kini

Abstract In today's world, when the internet is pervasive, everyone gets news from a variety of online sources. As the use of social media platforms has grown, news has travelled quickly among thousands of people in a very less duration. The propagation has been far reaching for the fake news generation in repercussions, from altering election outcomes in support of specific politicians, creating prejudiced viewpoints. Furthermore, spammers use appealing news headlines to make cash through click-bait adverts. In today's world knowingly or unknowingly fake news spreads around the world through internet. This has a great impact on the people who blindly believe whatever the internet provides. Hence, fake news identification has become a new study subject that is attracting a lot of attention. However, due to a lack of resources, such as datasets and processing and analysis procedures, it encounters several difficulties. This research uses a non-probabilistic machine learning models of computational prototypes to address this problem. Furthermore, the comparison of Term Frequency-Inverse Document Frequency (TF-IDF) is done, for the purpose of determining the best vectorizer used for detecting fake news. In order to raise the accuracy, stop words of English are used. To predict bogus news, a Support Vector Machine (SVM) classifier is deployed. According to the simulation data, the SVM and the TF-IDF produce results with high accuracy.

Keywords Fake news detection · TF-IDF · SVM · Machine learning

P. Suresh · A. Shettigar (✉) · M. Karunavathi · Ajith · M. G. Ramanath Kini
Department of Electronics and Communication Engineering, Mangalore Institute of Technology and Engineering, Moodabidri, Mangalore, Karnataka, India
e-mail: abhishekshettigar51@gmail.com

M. G. Ramanath Kini
e-mail: ramanath@mite.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_1

1 Introduction

As more people spend their time communicating in social media platforms through online, they are consuming more news and seeking out news through social media instead of referring to the other regional news organization. Many people believe false news influenced the 2016 presidential election campaign in some countries. Therefore, the word 'fake' news has entered the common language as a result of this election [1].

In today's world fake news has become a major problem in case of family, friends, and coworkers, and the core reason for this is social media. People disseminates a large area of news on specific topics, for example, the issues about the government, famous properties, etc. through internet. Bunch of fake rumors about the Covid-19 has spread through the internet between 2020 of January and 2020 of May. Using the modification of text, during the pandemic, people had spread false information. The goal of such kind of people is to spread false news to further incite political agenda and religious conflict. The fake news and rumors which comes from the internet cause problems in every region, including the health, agriculture, the private sector, school and colleges, industries, banks, and so on. Fake news is created to deceive the public in a developed and verified false manner [2, 3]. Micro-blog, Instagram, blogs, and Facebook are examples of social media platforms that can speed up the information among users which accelerate the distribution to every corner of the world. For the transformation of the information and for the social interaction, the social media has been used as the primary platform due to its user- friendliness, low cost, and rapid rate. Fake news can take the visible shape of images, texts, audio, or video [1]. A fakester manipulates information in order to mislead the public. The spread of the data via web-based networking media is done from the clients without prior check.

A proper methodology for the detection of fake news employs the following techniques:

- **Text preprocessing**: To remove the stop words along with special characters, steaming and analyzing the text.
- **Text encoding**: By the help of TF-IDF (Term Frequency- Inverse Document Frequency)
- **Characteristic extraction**: For the accurate identification of fake information, including source from news and the author, the date, and with emotion conveyed through the text to be the news features.
- **Support Vector Machine**: An improvised machine learning algorithm which can classify new data.

2 Literature Review

To select the best individual model, researchers have gone through several models about machine learning from various datasets. Paper [4] used SVM for the purpose of classifier and TF-IDF bunch with texts along with n-grams for technique of feature extraction. In addition, they executed a dataset which contains both the true and fake news, to be trained for the upcoming model. Based on the test which are done through different experiments, the TF-IDF algorithm outshone among other various intelligent classification algorithms. In paper [5], to train the twenty-three datasets from the supervised artificial intelligence algorithms, three datasets were used. Except recall, the algorithm of decision tree surpassed various smart performance matrices using classification algorithms, according to their results obtained. In article [6], with the help of set that include the feature about the article structure encapsulation, readability, and title-to-body similarity, they chose six machine learning algorithms for the experiment such as, AB (AdaBoost), DT (Decision Tree), KNN (K-Nearest Neighbour), RF (Random Forest), SVM (Support Vector Machine), and XGB (XGBoost). The confusion matrix model with other evaluation metrics applied, quantify a classifier's performance. Given the experiment's structure, the best results were produced from the support vector machine classifier.

Some researchers have developed extraordinary neural networks for detecting fake news. Article [7] developed a system with the methodology in order to point out fake news found in web articles by the help of machine learning, and natural language processing. For the generation of feature vector in the work proposed, various methods which features the count vector TF-IDF, with the collection of words were applied. To identify news as fake or as real, seven different machine learning classification algorithms were trained, especially in comparison in terms of accuracy, F1 Score, recall, and precision, before they chose the one which is best for building a system which have to predict a news as fake or real. Machine learning algorithms were used on different datasets from different sources in [8]. They also included each model's analysis results. The difficult task of detecting fake news can be simplified by using the right models and tools.

Some researchers have attempted to combat the wide spread of fake news on different social media sources. A feature of new set to predict quality from the current situations and characteristics of false information through detection of fake news was proposed in [9]. It shows intriguing findings about utility, also the significance about detecting false news features. Lastly, discussion about how to use fake news detection has been practiced, highlighting challenges as well as opportunities. Among those who contributed to this work is [10], that considered automating shaped news differentiating proof in Twitter datasets using a model for understanding mass produced news messages through post on the Twitter. Following that, they independently performed a similarity check within the five Machine Learning algorithms, such as the Support Vector Machine, Naive Bayes Method, Logistic Regression, and Recurrent Neural Network models, for the demonstration of the dataset to yield better result. The results revealed that SVM with Naive Bayes classifiers outperformed other calculations. The theory-driven model was applied in [2] for the detection of the fake news from extracting false information and also clicks bait characteristics of news article text and using models of machine learning which includes SVM, Random Forest, XGBoost, Naive Bayes, and Logistic Regression.

Some of the researchers have also made use of other news features to improve their results. A model was estimated in [11] to distinguish the fake news from news articles intuitively. It has proposed an unseen feature consisting of the set of classifiers for the purpose of machine learning. A dataset used in the experiment is a collection of both datasets containing balanced amounts of both true and fake news about politics. The linguistic/stylometric features, and bunch containing texts TF, and a BOW TF-IDF vector were extracted from the dataset's text fields, and then various machine learning models were applied, which include methods of bagging and boosting, for the purpose of obtaining the greater probability of accuracy. Purevdagva et al. [12] used an automated framework to detect phony political speech. It extracted the features of statements of political speech and their metadata using various classification methods, location, with the speech subject, writer's credibility, writer's profile, and also information of speech context. Machine learning model should be trained containing automatic selection, using features with parameterization. For a good amount of detection accuracy on the "Liar" dataset a SVM with trained model was achieved. The results which are evaluated show that the framework has been effective in detecting phony political speeches.

While the other researchers have concluded deep learning framework by themselves and received excellent results on their datasets. The precision accuracy, F1-score, and recall, evaluation measures, and also a proper version of McNemar's test was performed in [13], to check if the model performance have different output. Then, on the ISOT and KDnugget datasets, a novel stacking model of novel received accuracy for training and testing with 99.94% and 96.05%, respectively. Furthermore, when compared to baseline methods, the proposed method performed well. As a result, they strongly recommend it for detecting fake news. Paper [14] evaluated and compared various approaches to mitigate this issue, along with approaching several traditional machine learning approaches, like Naive Bayes, also other famous deep learning algorithms like RNN and hybrid CNN. The comparison was made not only between traditional and deep learning algorithms, but also between traditional and non-traditional methods. The groundwork for deciding on machine learning or deep learning procedures used to solve problems that bring out a balance between accuracy and light weightlessness. Manzoor et al. [15] conducted a study on other machine learning approaches for detecting fake and made out news [16]. Limitations of these kinds of approaches and improvisation through the implementation of deep learning are also discussed.

3 Methodology

The flow chart of the methodology is shown in Fig. 1.

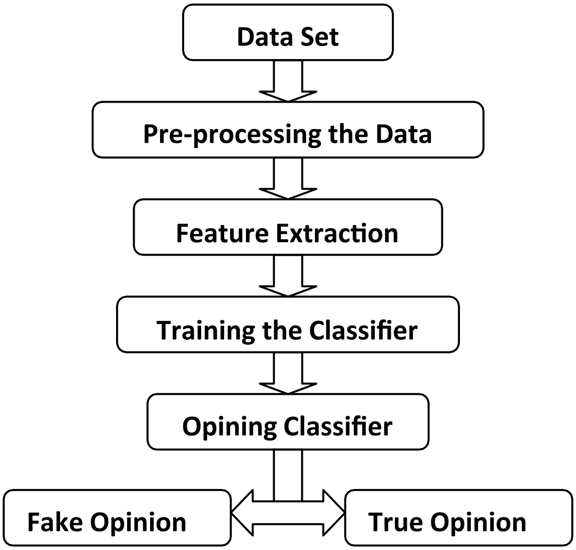


Fig. 1 Process of machine learning classification

A. Dataset

The dataset for the project has been selected from the site Kaggle, which included the data from several websites which included political news, government news, left news, middle east news, and US news, which has around 1 post recorded for 30 days. The main features consist of the title, text about the news body, subject about the news, authorized date, and the label, which included the training dataset of 18,525 posts and around 8980 for testing.

B. Pre-processing

Before the data are fed through machine learning algorithms, the text must be preprocessed using methodologies such as converting all letters from the document to the lowercase, removing stop words, tokenization, sentence fragmentation, and punctuation, and removing accent marks. These operations can greatly assist in choosing suitable terms for improving the performance of the model.

Two datasets are derived through real-world news articles. Therefore, cleaning up data by removing the stop words is performed. To finish the sentence structure, stop words are commonly applied in the English sentences, but these have little significance in terms of showing an individual's attitudes. As a result, they are removed from all experiments to prevent generating large unwanted noise. Further it is tokenized with TF-IDF by cleaning the data containing text.

- Term Frequency-Inverse Document Frequency (TF-IDF)

TF is a prevalent method used for tokenization which calculates in order to validate the proposed model by counting the set containing text available in the document. Every file is represented using a sequence containing the word that counts, which is done using TF technique. Later, for all vectors, the sum of elements will be unique, converting word counts into accuracy. Consider ‘D’ as a corpus, and a document is represented as ‘d’. Assume ‘w’ as word in document and $\mathrm{n_w(d)}$ will be the count of w in the data. As a result, dimensions of ‘d’ is denoted as:

$$|\mathrm{d}| = \sum_{\mathrm{wed}} \mathrm{n_w(d)}$$

The normalized TF used in the word ‘w’ in source ‘d’ is as follows:

$$TF(w)_d = \frac{n_w}{|d|}$$

In the machine learning experiments, TF-IDF is applied to convert the data in the form of vectors. The TF-IDF is a type of weighing metrics that frequently comes in hand for classification of text problems. This also has to gain a score to each term in a document, indicating the importance of the term. The significance of a term in this method grows in proportion to its frequency containing dataset. Let ‘D’ denote a corpus, that is, a collection of articles from the news. In this, ‘d’ represents an article made up of the words ‘w’. Using the equation below, mathematical calculation of IDF is done.

$$IDF(w)_D = \left(1 + \log \frac{|D|}{(\{|d| : D|w \in d|\})}\right)$$

The term TF-IDF used in the word (w) in relation for the document (d) and corpus (D) is known by the following.

$$\mathrm{TF - IDF(w)_{d,D} = TF(w)_d \times IDF(w)_D}$$

C. Classification: support vector machine (SVM)

A SVM Classifier is a guided technique machine learning which helps to tackle different classification along with problem related to regression. This algorithm is mainly used for solving difficulty on categorization. Every data point in n-dimension is plotted as a point (where ‘n’ is considered as the number that the features contain), considering that value in every feature have its amount with particular coordinate present in SVM algorithm. Later, it should be accomplished for the classification to distinguish the data by locating the hyper-plane as shown in Fig. 2.

4 Results and Discussion

In this work, first all letters from the document are converted to the lowercase, stop words are removed, sentence fragmentation, and punctuation, and removing

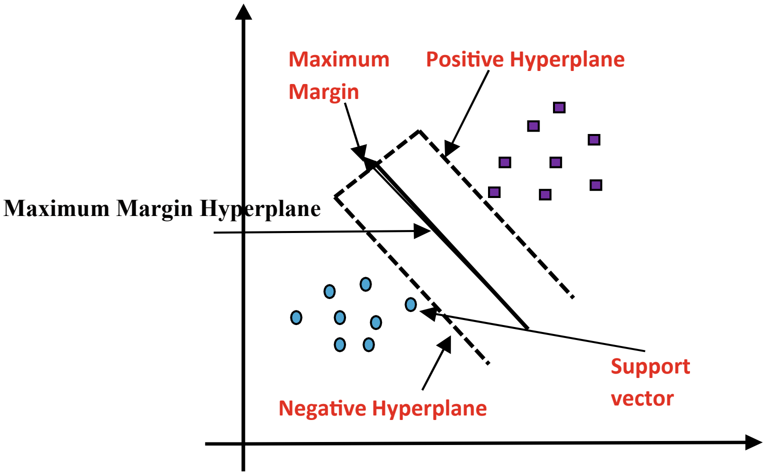


Fig. 2 Support vector machine classification

accent marks are performed. Then, TF-IDF method is utilized for tokenization to obtain the text representation. After that, text representation future is trained by a machine learning models i.e., SVM. The evolution of accuracy percentage belonging on features is by testing on the training data. The dataset is explored using Google Colab, a free cloud service backed by Google. In this study, 80% of the datasets are utilized to train a classifier, while the other 20% of the datasets are used to test the classification model.

The confusion matrix data collected at a classifier are summarized in Table 1.

1. False Positives is represented by FP, True Positives is represented by TP, False Negatives is represented by FN and True Negatives is represented by TN, which are the four metrics in the confusion matrix that can be used to evaluate a classifier's performance. When the data are examined, the SVM classifier holds a minimal error rate of 28 articles for false negatives and 20 articles for false positives. Furthermore, the SVM classifier provides the most accurate classification of false at 4619 articles as TN and real at 4313 articles as TP.

The equations for the contents in the Table 2 are:

$$\text{Precision}: \frac{TP}{(TP+FP)} = \frac{4313}{(4313+20)} \times 100 = 99.53\%$$

Table 1 Confusion matrix results by classifier predicted label

True label	Predicted label	
	TN	FP
	4619	20
	FN	TP
	28	4313

Table 2 Performance results of precision, accuracy, recall and F-measure

Classifier	Precision	Accuracy	Recall	F-Measure
SVM	99.3%	99.4%	99.4%	99.4%

$$\begin{aligned}\text{Accuracy}: \frac{(TP+TN)}{(TP+TN+FP+FN)} &= \frac{(4313+4619)}{(4313+4619+20+2)} \times 100\\ &= 99.46\%\end{aligned}$$

$$\text{Recall}: \frac{TP}{(TP+FN)} = \frac{4313}{(4313+28)} \times 100 = 99.35\%$$

$$\text{F - Measure}: \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} = \frac{(2 \times 99.53 \times 99.35)}{(99.53 + 99.35)} = 99.43$$

The confusion matrix gives metrics that can also be utilized to calculate the achievement of a project. Accuracy, precision, recall, and F-measure are examples of such measurements. The test dataset (which includes 20% of all articles) is used. The results of the classifier and the selected hyper parameter configurations are reported in Table 2. The results are achieved by utilizing features that characterize the content and structure of each document. Simple, traditional techniques of expressing documents (TF-IDF) do not keep the meanings and relationships of words. When it comes to accuracy, SVM came out with a score of 99.4%. When it comes to precision, SVM came out on top with 99.3%. In terms of recall, SVM came out on top with a score of 99.4%. SVM had the best result for F-Measure, with a score of 99.4%.

5 Conclusion

This paper proposes a term frequency extraction for support vector machine-based system for the purpose of detecting fake news, with the goal aimed to determine the optimal techniques along with features for detecting fake news. Thus, it is initiated from researching the area related fake news, it's effect and ways of detecting it. Then it is devised and a system is built that extracts a set of characteristics that may be used to detect false news from dataset of news that has been pre-processed using feature extraction (TF-IDF algorithms), training the classifier, and using opinion classifier (support vector machine) to classify the fresh data. The findings of this proposed approach are summarized as follows:

- Text, author, source, date, and sentiment are greatest features for the detection of fake news.
- This procedure provided successfully a noticeable rate.

- Even though analysis from the text's sentiment is tempting, it has been more influential in the case regarding opinion mining.
- With large datasets and large texts, the TF-IDF performs better compared to other algorithms.
- The support vector machine (SVM) seemed to be the better method for detecting fake news, as this provided higher noticeable rate which is allowed for the classification of every piece of information with a degree of confidence.

References

1. Zhang X, Ghorbani AA (2020) An overview of online fake news: characterization, detection, and discussion. Inf Process Manage 57(2)
2. Zhou X, Jain A, Phoha VV, Zafarani R (2020) Fake news early detection: a theory-driven model. Res Pract 1(2):1–25. Article no.: 12
3. Poddar K, Umadevi KS (2019) Comparison of various machine learning models for accurate detection of fake news. In: 2019 Innovations in power and advanced computing technologies (i-PACT), vol 1. IEEE
4. Baarir NF, Djeffal A (2021) Fake News detection Using Machine Learning. In: 2020 2nd International workshop on human-centric smart environments for health and well-being (IHSH), pp 125–130
5. Ozbay FA, Alatas B (2020) Fake news detection within online social media using supervised artificial intelligence algorithms. Phys A Stat Mech Appl 540:123174. ISSN: 0378-4371
6. Ngada O, Haskins B (2020) Fake news detection using content-based features and machine learning. In: 2020 IEEE Asia-Pacific conference on computer science and data engineering (CSDE), pp 1–6
7. Smitha N, Bharath R (2020) Performance comparison of machine learning classifiers for fake news detection. In: 2020 Second international conference on inventive research in computing applications (ICIRCA), pp 696–700
8. Mandical RR, Mamatha N, Shivakumar N, Monica R, Krishna AN (2020) Identification of fake news using machine learning. In: 2020 IEEE international conference on electronics, computing and communication technologies (CONECCT), pp 1–6
9. Reis JCS, Correia A, Murai F, Veloso A, Benevenuto F (2019) Supervised learning for fake news detection. IEEE Intell Syst 34(2):76–8
10. Bharath G, Manikanta KJ, Prakash GB, Sumathi R, Chinnasamy P (2021) Detecting fake news using machine learning algorithms. In: 2021 International conference on computer communication and informatics (ICCCI), pp 1–5
11. Jain MK, Gopalani D, Meena YK, Kumar R (2020) Machine Learning based Fake News Detection using linguistic features and word vector features. In: 2020 IEEE 7th Uttar Pradesh section international conference on electrical, electronics and computer engineering (UPCON), pp 1–6
12. Purevdagva C, Zhao R, Huang PC, Mahoney W (2020) A machine-learning based framework for detection of fake political speech. In: 2020 IEEE 14th international conference on big data science and engineering (Big Data SE), pp 80–87
13. Jiang T, Li JP, Haq AU, Saboor A, Ali A (2021) A Novel stacking approach for accurate detection of fake news. IEEE Access 9:22626–22639
14. Han W, Mehta V (2019) Fake news detection in social networks using machine learning and deep learning: performance evaluation. In: 2019 IEEE international conference on industrial internet (ICII), pp 375–380

15. Manzoor SI, Singla J, Nikita (2019) Fake news detection using machine learning approaches: a systematic review. In: 2019 3rd International conference on trends in electronics and informatics (ICOEI), pp 230–234
16. Zhou X, Zafarani R (2018) Fake news: a survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315 2

Aquaculture Monitoring System Using Internet of Things

G. V. R. Kameshwar Rao, T. J. Dhivya Shrilaa, I. Akash, and G. Gugapriya

Abstract Aquaculture comprises an important part of agriculture which has been the pillar in the primary sector of India. Many factors such as temperature, natural calamities and artificial chemicals when monitored irregularly can severely impact the survival of the species. There is more time and energy invested by various scientists across the world to monitor these species in an efficient manner. This is where the idea of an utopian system is initiated to monitor various parameters such as temperature, turbidity, pH and much more through a union of various sensors connected to a microcontroller. The principles of Internet of things and machine learning assist the system to estimate the probability of survival of the species by using the data collected from the sensors. The data which is gathered from the sensors is predicted using the services of IBM and is visualized through a web application that helps the user to interpret data in a coherent manner.

Keywords Aquaculture monitoring · Internet of things · Parameter prediction · Machine learning · Chatbot

1 Introduction

Aquaculture harvesting was something that was introduced more than 2000 years ago. Fishes and other species are cultivated in small ponds and lakes which contribute to a major share of the agricultural sector. Over the years, India has developed to become a major exporter of shrimps and various other species which has improved the gross domestic product of India [1]. The harvesting of shrimps is very delicate [2] because of the nature of the shrimps not being able to handle the minimal differences with regards to some natural parameters such as temperature, pH, turbidity as well as being unable to withstand the intake of artificial chemicals in minimal quantities. These parameters have been very difficult to monitor daily which becomes a major problem when harvesting these aquatic species [3].

G. V. R. Kameshwar Rao (✉) · T. J. Dhivya Shrilaa · I. Akash · G. Gugapriya
School of Electronics Engineering, Vellore Institute of Technology, Chennai, India
e-mail: kamesh.gvr.2000@gmail.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_2

The conventional way of monitoring these species was through a doctor who is specialized in the domain of aquatic species. They would have their kit to test out the quality of water as well as the species and can easily find out about their chances of survival when the species is affected by diseases [4]. The flaw with this approach was the monitoring of these parameters daily. Certain species are very vulnerable and can easily die out from the excess of any parameter that is being monitored [5].

The Internet of things has created a massive boost with regards to the approach to solving real-world problems [6]. Users have been able to access their gadgets and monitor various parameters through the medium of the internet. The data which is gathered by various sensors such as a temperature sensor, pH sensor, and much more are transmitted to the IBM Cloud Services through the principles of the internet of things which eases the extraction and visualization of data [7].

With the internet of things, the system that we propose can monitor parameters essential to predict the survival of the species. The parameters that the system is going to monitor apart from temperature and pH would be the values of turbidity and total solvents that are dissolved in the environment. These sensors would record the data that would be beneficial to be used for the process of data telemetry and interpretation by making the best use of the internet of things.

With the system being able to interpret the data using the concepts of parameter forecasting, the visualization of the parameters viewed through an application makes the system user-friendly and also assists the user by implementing the services of the chatbot as well along with the application [8]. The culmination of all these objectives would ensure a utopian system that can monitor the parameters required for the survival of aquatic species in an efficient and user-friendly model.

2 Related Works

A system has been proposed where in addition to conventional water quality monitoring, research on inter-dependency of the parameters has been considered and is used to reduce the number of sensors involved by rejecting the sensor used for a predictable parameter. In this case, the dissolved oxygen has been neglected as it is expected to be in the normal range if temperature, pH, and conductivity are normal. The results were promising and paved the way for further advancement in soft sensing other parameters. The overall cost of the system is high due to the use of a cell phone to capture the photo of the water to measure the cleanliness and the use of Raspberry Pi for the computation of the photo to detect the color of the water and to store the database [9].

In 2021, the authors [10] discussed other techniques used for soft measurement of water quality parameters using regression and deep learning network. The IPSO (Improved Particle Swarm Optimization)-Least Squares Support Vector Regression (LSSVR) method used to perform a nonlinear prediction model of dissolved oxygen when compared to the traditional Least Squares Support Vector Regression (LSSVR) model, provides better results in terms of reduced root mean square error and mean absolute error values. The Long short-term memory (LSTM) deep learning network performed by the authors shows results up to 98.56% and 98.97% short-term prediction accuracy of pH and water temperature respectively [11].

Android apps have been developed to facilitate the remote monitoring of the water quality parameters where the authors implement a three-layered architecture system consisting of the presentation layer mobile application, data access layer—MySQL database manager, and the devices layer—sensors, actuators, etc. This system uses a general packet radio service module to send the data to the cloud storage and thus is suitable for implementation in remote locations without access to the internet. The mobile application also displays the parameter values in a graph for daily reports. The authors compared the obtained parameter values with the manual tested values and found the variation to be negligible. The overall performance of the system in terms of size, weight, and percentage of survival of the shrimp was observed to be more effective than the manual testing systems [12].

In 2019, the authors have focused on the error percentage of the sensors between laboratory tests and sensor data acquired and forecast of the water quality parameters using an automatic exponential smoothening model using historical values of the parameter thus giving an insight to the farmers to help them in better decision making. As the model used is a time series forecasting method and the parameters are not entirely seasonal with respect to time, the forecast trend obtained is unreliable [13].

Other methods have also been used for the forecast of water quality such as the artificial neural networks (ANNs) for fish farm management. The model forecasts the future data using the historical data and the results obtained are used to perform necessary actions to prevent future catastrophes. The authors construct a wireless sensing network with the use of a general packet radio service combined with virtual private network functionality to collect data on a country-wide scale. The acquired data can be viewed in real-time through the internet [14].

3 Methodology

The proposed method consists of 5 subsystems that make up the overall monitoring system. The subsystems are data extraction, data telemetry and visualization, parameter prediction, chatbot, and actuator.

3.1 Data Extraction

To assess the quality of the water to maintain a suitable thriving condition for the aquaculture farm, various parameters are required to be monitored. This model aims to measure the following parameters-pH, total dissolved solids (TDS), temperature of water, turbidity and water level. The parameters are chosen based on the dependency of various hard-to-measure water quality parameters on the crucial and easily measurable parameters. The dissolved oxygen has been neglected since we can assume its normalcy if the pH, temperature, and conductivity are in the normal range. Sensors used to measure these parameters are connected to the microcontroller directly.

3.2 Data Telemetry and Visualization

The extracted data is sent over the internet to a cloud platform where it is stored in a database [15]. These values are also sent to an android app with a customer login-based system. The data is visualized for easy interpretation in the form of graph charts for each parameter and is displayed in the dashboard.

3.3 Parameter Prediction

To perform soft sensing of a parameter to effectively neglect the use of its sensor, a machine learning algorithm is used to predict the parameter based on historical data of dependency on other parameters. Various prediction models are used on the training data and applied to testing data [16]. The model which provides the highest accuracy is chosen and is automated to run on the cloud platform using real-time data to predict the parameter which in this case, pH using temperature, turbidity, and water level. The output value is sent to the dashboard in the app.

3.4 Chatbot

The user-friendly chatbot service has been provided mainly for solving the queries of the customers. The chatbot has a corresponding response for every action (query) of the user, for example, if the user states that the temperature is low and dissolved oxygen levels are high after heavy rains, then the chatbot suggests that shrimp may be subjected to White Spot Syndrome Virus (WSSV). The chatbot suggests the possible causes of the disease, fluctuation in any vital parameters, resolves doubts regarding the app and the system and helps contact support in case of emergency.

3.5 Water Quality Maintenance

Actuators generally act as a response of the system where the data received by sensors determine their response accordingly. In this system the water pump acts as an actuator for maintaining the cleanliness of the water, if the value measured by the turbidity sensor is beyond the normal range, then the water should be replaced immediately with clean water to avoid possible stress to the shrimps. The proposed system performs the automation of this process by replacing the water using pumps which can be controlled by the microcontroller connected via a relay module. The microcontroller is designed to automatically switch on the inlet pump to input clean water and the outlet pump to discharge the unclean water. This process can also be controlled and monitored manually through the app efficiently. The manual controlling has been enabled by the message queuing telemetry transport (MQTT) broker where the app acts as publisher for that particular instance and the NodeMCU being subscriber has subscribed to the topic. According to the payload received the relay is controlled by NodeMCU which in turn controls the pump, thus enabling both manual and automatic controlling of the water pump.

4 Design and Implementation

The design of the system (shown in Fig. 1) can be categorized into two parts-the hardware and software. The hardware category consists of sensors that are integrated with the microcontroller and the actuator to respond according to the condition. The software category consists of different IDEs used to program the hardware and develop the machine learning services, chatbot, and android app. A detailed explanation has been discussed in the following sections.

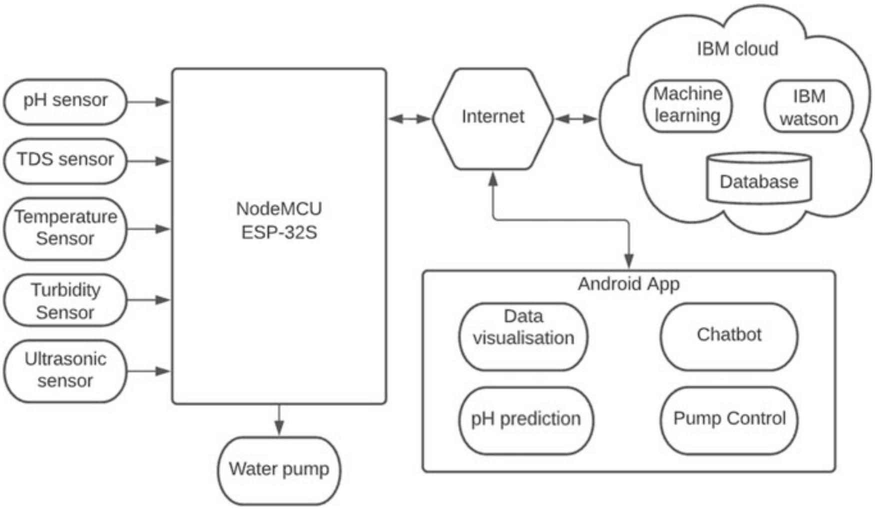


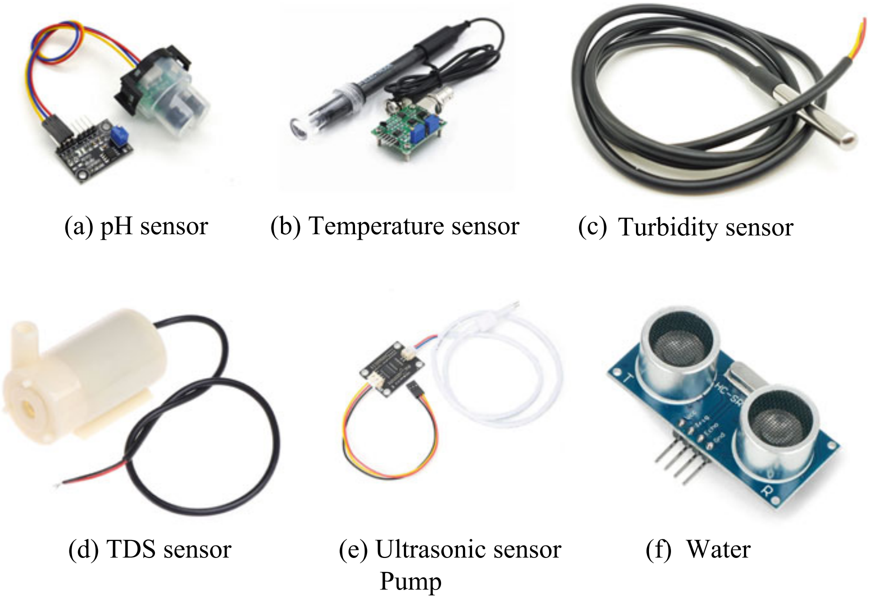
Fig. 1 Block diagram of the system

4.1 Hardware Details

Sensors. The Techtonics pH sensor (shown in Fig. 2a) is used to measure the pH level of the water in this work with a measuring range of 0–14 pH with an accuracy of ± 0.1pH. The operating temperature ranges from 0 to 60 °C. The pH electrode is connected via a Bayonet Neil-Concelman (BNC) connector where the signal is amplified by the microcontroller. The suitable pH range for shrimps is 7.9–8.2 pH. It is one of the most important parameters in terms of shrimp monitoring.

The DS18B20 temperature sensor (shown in Fig. 2b) is water-resistant with operating temperatures ranging from − 55 to 125 °C and accuracy of ± 0.5 °C. It can be easily controlled via One wire (i.e., single data line required for communication) which makes it one of the most reliable and cheap sensors for usage. The viable range of temperature for shrimps is 27–29 °C which benefits in faster growth and higher reproduction rates.

The turbidity sensor (shown in Fig. 2c) measures the value by detecting the amount of light received by the electrode by passing through the water. The operating temperatures range from 5 to 90 °C with a response time of less than 500 ms. The unit for turbidity is NTU which stands for "Nephelometric Turbidity Units". The optimal turbidity value for shrimp is 21.42 ± 4.43 NTU.



(a) pH sensor (b) Temperature sensor (c) Turbidity sensor

(d) TDS sensor (e) Ultrasonic sensor Pump (f) Water

Fig. 2 Sensors and pump

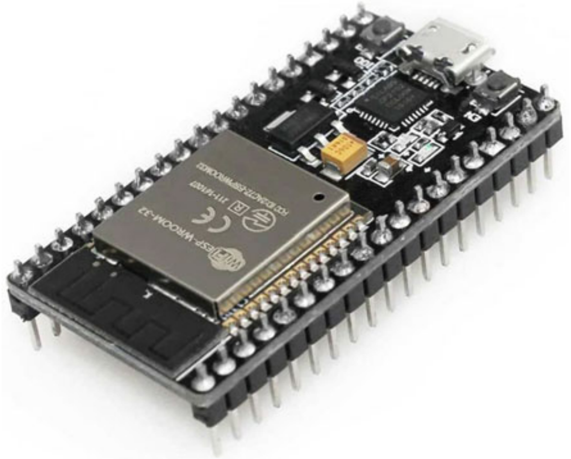
The CentIot total dissolved solvents (TDS) sensor (shown in Fig. 2d) measures the hardness of the water (i.e. salts, minerals, metals, etc.) Turbidity is generally expressed in parts per million (ppm). Its measuring range varies from 0 to 1000 parts per million (ppm) with an accuracy of ± 10% F.S. The optimal total dissolved solvents (TDS) value for shrimp is 200–400 parts per million (ppm).

The HC-SR04 Ultrasonic Sensor (shown in Fig. 2e) measures the distance or depth of the object from the sensor using the principles of ultrasonic waves. This sensor can measure up 2–400 cm and can trigger an input pulse for every 10 uS which helps to monitor the species continuously. It can sustain under cold temperatures underwater while utilizing around 20 mA thus being an efficient sensor to monitor these aquatic species.

The submersible direct current (DC) water pump (shown in Fig. 2f) pumps the water outward via the plastic probe attached to the pump. The operating range of the pump is 3–6 V and can be controlled by a relay attached to it. The pump has been used in this work to remove the water if the turbidity values surpassed the permissible level of the shrimps.

Microcontroller. The NodeMCU development board (WROOM 32) has 38 general purpose input–output (GPIO) pins and consists of an ESP32-S microcontroller (shown in Fig. 3) which supports compatibility with wireless fidelity (Wi-Fi) and low powered Bluetooth (BLE). It has 520 KB of static random-access memory (SRAM), 448 KB of read-only memory (ROM), supports an analog-to-digital converter (ADC), and is programmable via Arduino integrated development environment. This development board has been chosen for the application as it has provided excellent compatibility with wireless fidelity (Wi-Fi) and supports analog-to-digital converter (ADC) conversion.

Fig. 3 Node MCU-ESP 32S



4.2 Software Details

Arduino IDE. The open-source software is an integrated development environment (IDE) that has been used to program the NodeMCU development board and the code is uploaded via a universal serial board connector. The major advantages of the software are compatible with various versions of development boards, community support, and ease to compile/build the program.

Jupyter Notebook. The Jupyter notebook is an integrated development environment (IDE) to develop, train and test the prediction models used in parameter prediction. In this work, we aim to predict the pH of the water using two parameters-turbidity and temperature. We perform the comparison of the following multiple regression models-Random Forest regressor, Gradient Boosting regressor, Ridge regressor, Lasso regressor, and Elastic Net regressor. The model that provides the highest R-squared score and least mean average error (MAE) score is chosen as the final model.

IBM Cloud Services. IBM cloud, also famously known as Bluemix, is one of the major cloud platforms with various services such as Watson, Internet of Things (Internet of things), database management, analytics, storage, and many more. IBM Cloud is chosen as the cloud service platform for this work to build a chatbot using the Watson Assistant service, to deploy and monitor machine learning models for parameter prediction, and to visualize the parameters in the dashboard.

IBM Watson Assistant. It is an advanced processor which performs analytics on the data received and answers the questions accordingly. In this work, we have used it as a user-friendly chatbot to resolve issues with customers. Watson has corresponding action and response to the vast data stored which suggests the best possible solutions for the concerned queries. An information survey is done to gain knowledge on possible information that could be included in the chatbot. The information can be categorized into two divisions-System-related queries and general queries. System-related queries consist of the queries that the users could raise for the successful performance of the system. Example: Hardware system not connected to the internet. General queries consist of the queries that the users could raise related to aquaculture farming. Example: disease detection. This chatbot will be connected to the app through an application program interface (API) so that it can be accessed by the users.

Parameter Prediction. The IBM Machine learning service consists of various tools which can be used for deep learning, building and deploying various models, an interactive programming environment, and many more tools which make it user-friendly to build and test various models with vast datasets. The final model is uploaded to the IBM machine learning service through an application program interface to be deployed on the server. The model is then deployed using the real-time data acquired from the sensors and the predicted value of pH is then compared with the acquired sensor value.

IBM Cloud Dashboard. The IBM cloud dashboard is generally used to visualize the data which gives great insights about the data and helps in preventing losses. The dashboard is customizable and can be shared with other users too. The data is received in the form of JavaScript object notation (JSON) via the publish-subscribe method. The dashboard displays the live data in the form of a graph and historical data of the parameter values which is sent to the app so that the user can access it from anywhere with internet access.

Android Studio. The Android Studio is an integrated development environment (IDE) used for the sole purpose of android application development. The design of the mobile app has been accomplished by coding in Java. The android app consists of a login page in which the users can create a new account or login into their accounts. After logging in, the dashboard page will be displayed which will consist of the real-time sensor data values and pH values predicted using the model. A navigation panel on the top right corner can be used to access other pages-an aquaculture database page used to access the historical data of all the parameters and a water quality maintenance page to monitor or manually control the water inlet and outlet pumps in the system. All the pages will have the chatbot icon in the bottom right corner.

4.3 Overview of Implementation

The system (shown in Fig. 4) is powered by a 5 V battery connected to the microcontroller. The sensors are calibrated and connected to the microcontroller which is programmed to acquire the water quality parameter values with the interval of one minute. The microcontroller should be constantly connected to wireless fidelity (Wi-Fi) in order to send the data over to the IBM Watson Internet of things platform which uses the publish-subscribe method. The parameter values are displayed in the IBM dashboard along with the pH value automatically predicted by the machine learning model. This data is sent to the app which also requires constant internet access. The automization of the pump control is done which checks the conditions required to perform the action. The control of the pump can also be done manually through the app using the message queuing telemetry transport (MQTT) protocol. The pump will be run till the conditions are satisfied and the status will be updated in real-time in the app.

4.4 Implementation of pH Prediction

Data Collection. Regressor-based machine learning models used for prediction require a dataset consisting of the dependent parameter-pH and the other independent parameters turbidity, temperature, and water level. The datasets required for this purpose are acquired from Kaggle and GitHub repositories [17, 18].

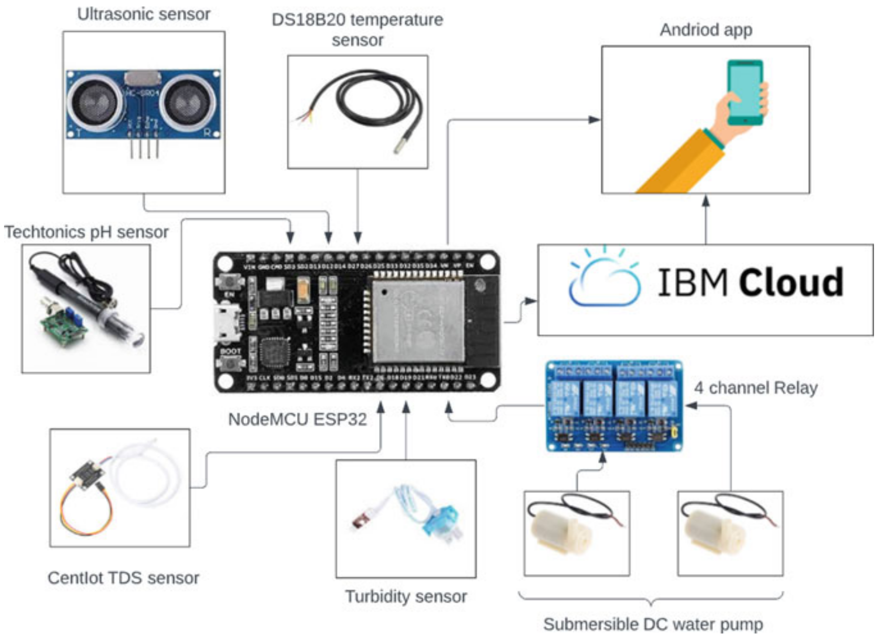
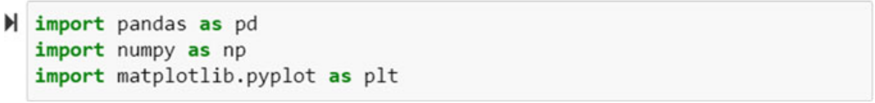


Fig. 4 Block diagram

Data Preparation. The dataset is analyzed for outliers which are removed in Excel sheets. Then the dataset is uploaded to the Jupyter server hosted locally. A Jupyter notebook (shown in Fig. 5) is created on the server and the required modules such as numpy and pandas are imported for visualization.

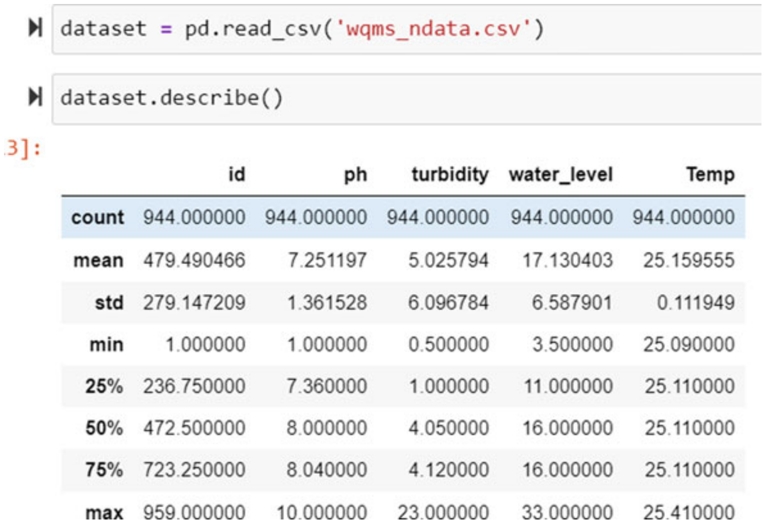
The dataset is imported to a variable and a description of the dataset is obtained for analysis. It can be inferred (shown in Fig. 6) that there are 944 entries in this dataset and the minimum and maximum values of the parameters are in the threshold range.

Visualization of the parameters in the dataset is performed using Matplotlib.pyplot and it can be inferred from the output (shown in Fig. 7) that the parameters are correlated to each other. Therefore, a regression model can be used to predict a dependent parameter using the other independent parameter due to the level of correlation between them.



```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Fig. 5 Required modules are imported



```
dataset = pd.read_csv('wqms_ndata.csv')
```

```
dataset.describe()
```

3]:

	id	ph	turbidity	water_level	Temp
count	944.000000	944.000000	944.000000	944.000000	944.000000
mean	479.490466	7.251197	5.025794	17.130403	25.159555
std	279.147209	1.361528	6.096784	6.587901	0.111949
min	1.000000	1.000000	0.500000	3.500000	25.090000
25%	236.750000	7.360000	1.000000	11.000000	25.110000
50%	472.500000	8.000000	4.050000	16.000000	25.110000
75%	723.250000	8.040000	4.120000	16.000000	25.110000
max	959.000000	10.000000	23.000000	33.000000	25.410000

Fig. 6 Importing data

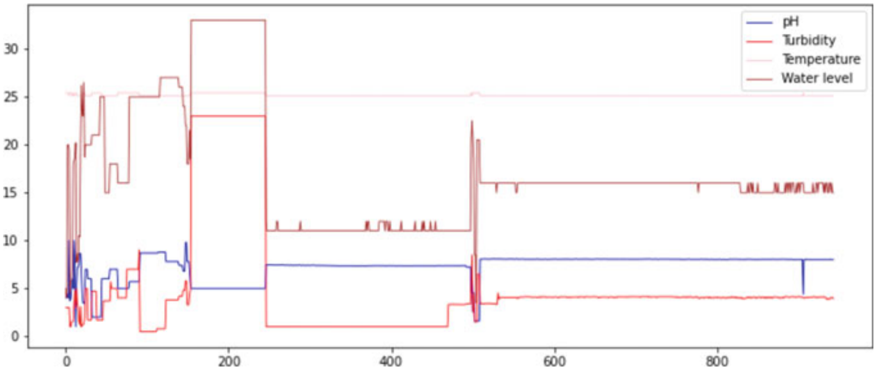
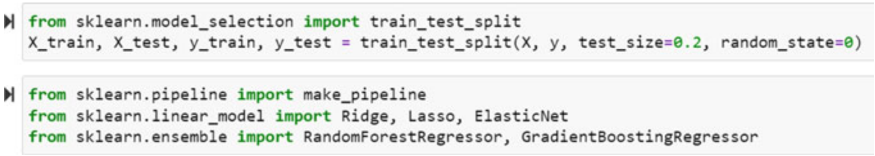


Fig. 7 Visualization of the dataset

Choosing and Training the Model. To choose the model that would give the best accuracy for the dataset, five regressor prediction models are trained and tested on the dataset to analyze the accuracy. The models are-Random Forest regressor, Gradient boosting regressor, Ridge regressor, Lasso regressor, and Elastic Net regressor. The dataset is split into training (80% of the dataset) and testing (20% of the dataset) data (shown in Fig. 8). The required modules of the regressors are imported.

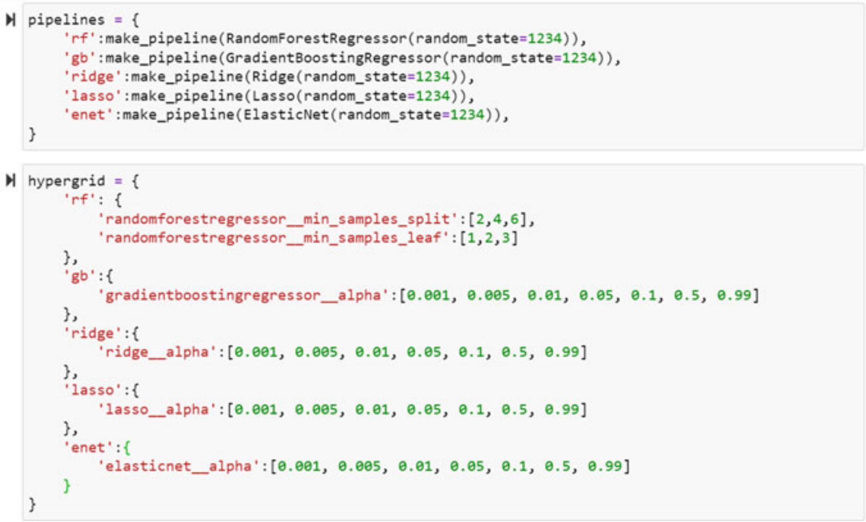
The regressor models are pipelines (shown in Fig. 9) to automate the machine learning process and the weights are set for each model. The weights are tuned after a number of trial and error and feedback after prediction.



```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
```

```
from sklearn.pipeline import make_pipeline
from sklearn.linear_model import Ridge, Lasso, ElasticNet
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor
```

Fig. 8 Splitting the dataset into training and testing data



```
pipelines = {
    'rf':make_pipeline(RandomForestRegressor(random_state=1234)),
    'gb':make_pipeline(GradientBoostingRegressor(random_state=1234)),
    'ridge':make_pipeline(Ridge(random_state=1234)),
    'lasso':make_pipeline(Lasso(random_state=1234)),
    'enet':make_pipeline(ElasticNet(random_state=1234)),
}
```

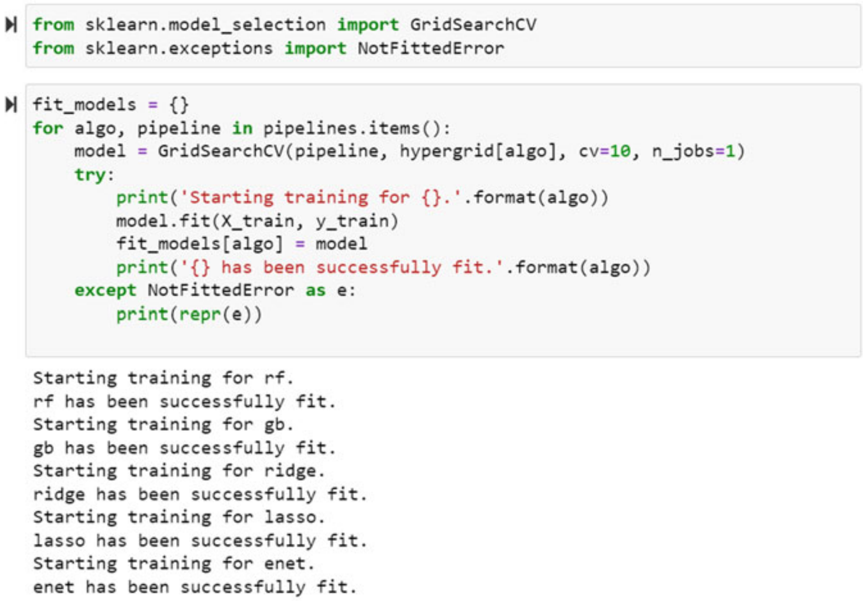
```
hypergrid = {
    'rf': {
        'randomforestregressor__min_samples_split':[2,4,6],
        'randomforestregressor__min_samples_leaf':[1,2,3]
    },
    'gb':{
        'gradientboostingregressor__alpha':[0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 0.99]
    },
    'ridge':{
        'ridge__alpha':[0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 0.99]
    },
    'lasso':{
        'lasso__alpha':[0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 0.99]
    },
    'enet':{
        'elasticnet__alpha':[0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 0.99]
    }
}
```

Fig. 9 Pipeline making and tuning weights

Required modules are imported and the models are fitted (shown in Fig. 10) for the training data.

Evaluation and Inference. On performing evaluation and comparing the R-squared score (R2) score and mean absolute error results, it is found that the random forest regressor performs well for the dataset used in this work. The highest R-squared (R2) score and the lowest mean absolute error is obtained in the random forest regressor (RFR) model (shown in Fig. 11). We can infer from the above table that the random forest regressor (RFR) model achieves approximately 97.3% accuracy in predicting the pH for the testing data. Therefore, we choose this as the final model to be used for soft sensing the pH parameter from the real-time inputs in the IBM cloud.

Deployment. The model has to be deployed in the IBM cloud in order to request pH prediction and to automize the prediction using the values of independent parameters acquired from the sensor to successfully replace the pH sensor. The deployment is performed using the application program interface (API) client of the IBM Watson Machine Learning service and the required modules are imported. The application

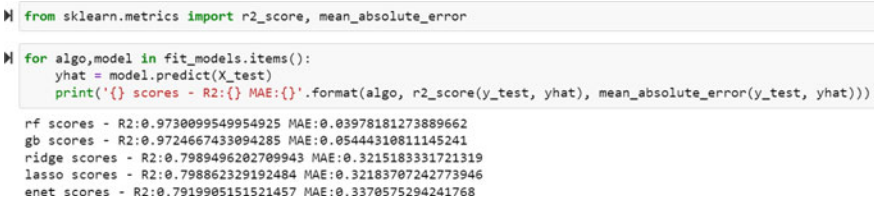


```
from sklearn.model_selection import GridSearchCV
from sklearn.exceptions import NotFittedError
```

```
fit_models = {}
for algo, pipeline in pipelines.items():
    model = GridSearchCV(pipeline, hypergrid[algo], cv=10, n_jobs=1)
    try:
        print('Starting training for {}.'.format(algo))
        model.fit(X_train, y_train)
        fit_models[algo] = model
        print('{} has been successfully fit.'.format(algo))
    except NotFittedError as e:
        print(repr(e))
```

```
Starting training for rf.
rf has been successfully fit.
Starting training for gb.
gb has been successfully fit.
Starting training for ridge.
ridge has been successfully fit.
Starting training for lasso.
lasso has been successfully fit.
Starting training for enet.
enet has been successfully fit.
```

Fig. 10 The regressor models are fit for the training data



```
from sklearn.metrics import r2_score, mean_absolute_error
```

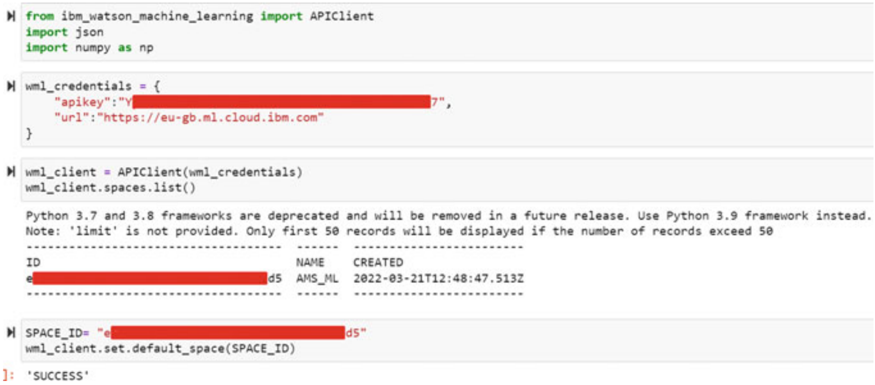
```
for algo,model in fit_models.items():
    yhat = model.predict(X_test)
    print('{} scores - R2:{} MAE:{}'.format(algo, r2_score(y_test, yhat), mean_absolute_error(y_test, yhat)))
```

```
rf scores - R2:0.9730099549954925 MAE:0.03978181273889662
gb scores - R2:0.9724667433094285 MAE:0.05444310811145241
ridge scores - R2:0.7989496202709943 MAE:0.3215183331721319
lasso scores - R2:0.798862329192484 MAE:0.32183707242773946
enet scores - R2:0.7919905151521457 MAE:0.3370575294241768
```

Fig. 11 Comparison of R2 and mean absolute error scores of pH prediction

program interface key, uniform resource locator (URL) of the server, and the space ID of the service is acquired from the IBM cloud website (shown in Fig. 12).

The meta data of the model necessary to deploy in the IBM cloud is defined. The Random Forest model is stored in the ‘BEST_MODEL’ variable. The ‘MODEL_NAME’ and ‘DEPLOYMENT_NAME’ are provided. The software specification UID is acquired through a function that searches the ID using the python version of the Jupyter notebook which is version 3.8. The model is then stored in the IBM cloud repository along with the training dataset (shown in Fig. 13).



```
from ibm_watson_machine_learning import APIClient
import json
import numpy as np
```

```
wml_credentials = {
    "apikey":"Y[redacted]7",
    "url":"https://eu-gb.ml.cloud.ibm.com"
}
```

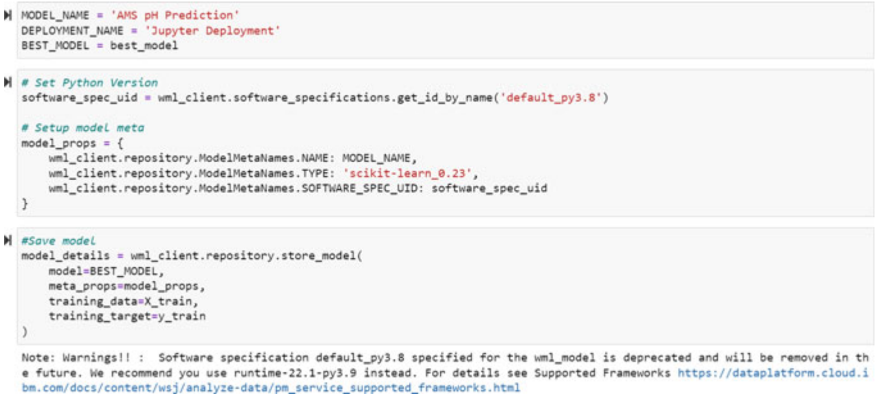
```
wml_client = APIClient(wml_credentials)
wml_client.spaces.list()
```

```
Python 3.7 and 3.8 frameworks are deprecated and will be removed in a future release. Use Python 3.9 framework instead.
Note: 'limit' is not provided. Only first 50 records will be displayed if the number of records exceed 50
------------------------------------  ------  ------------------------
ID                                    NAME    CREATED
e[redacted]d5                         AMS_ML  2022-03-21T12:48:47.513Z
------------------------------------  ------  ------------------------
```

```
SPACE_ID= "e[redacted]d5"
wml_client.set.default_space(SPACE_ID)
```

```
]:  'SUCCESS'
```

Fig. 12 Setting the credentials and space ID



```
MODEL_NAME = 'AMS pH Prediction'
DEPLOYMENT_NAME = 'Jupyter Deployment'
BEST_MODEL = best_model
```

```
# Set Python Version
software_spec_uid = wml_client.software_specifications.get_id_by_name('default_py3.8')

# Setup model meta
model_props = {
    wml_client.repository.ModelMetaNames.NAME: MODEL_NAME,
    wml_client.repository.ModelMetaNames.TYPE: 'scikit-learn_0.23',
    wml_client.repository.ModelMetaNames.SOFTWARE_SPEC_UID: software_spec_uid
}
```

```
#Save model
model_details = wml_client.repository.store_model(
    model=BEST_MODEL,
    meta_props=model_props,
    training_data=X_train,
    training_target=y_train
)
```

```
Note: Warnings!! :  Software specification default_py3.8 specified for the wml_model is deprecated and will be removed in th
e future. We recommend you use runtime-22.1-py3.9 instead. For details see Supported Frameworks https://dataplatform.cloud.i
bm.com/docs/content/wsj/analyze-data/pm_service_supported_frameworks.html
```

Fig. 13 Storing the model in IBM cloud repository

5 Results and Discussion

A system that is capable of extracting data values from various sensors that are imbued with the microcontroller along with the services of IBM would provide the results with which the user can cultivate the species in a productive manner. The data is visualized in the application which makes the user to interpret the values easily.

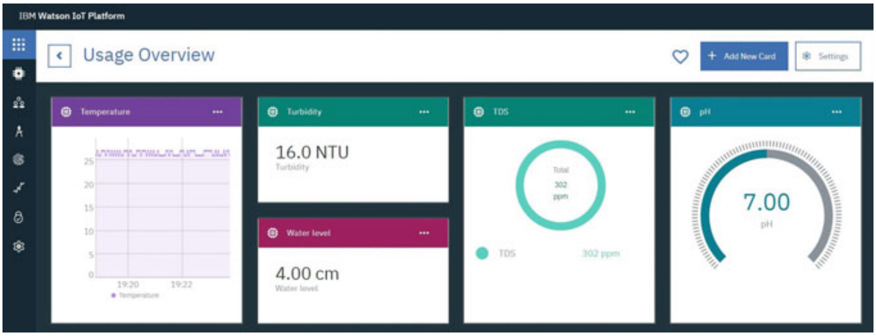


Fig. 14 Data visualization in IBM Watson

5.1 *Sensor Data Visualization in IBM Watson Internet of Things Platform*

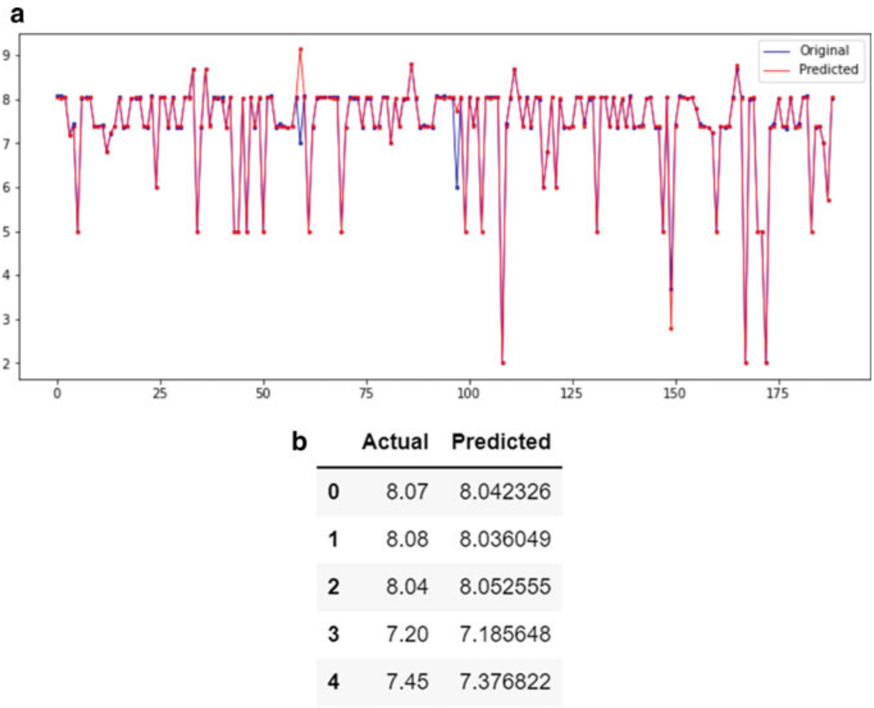
The data which is recorded in their individual sensors is transferred to the IBM Watson Platform by accessing the microcontroller through the medium of Internet. In this IBM Watson Platform (shown in Fig. 14), the data that is transferred from the microcontroller is visualized in their individual units such as temperature being recorded in the units of Celsius and turbidity being recorded in the units of nephelometric turbidity units (NTU). This helps to understand the influence of the parameters and their impact on the aquatic species.

5.2 *Parameter Prediction*

On evaluating the Random Forest model for pH prediction using the testing data, 0.973 R2 score and 0.0397 mean absolute error score is achieved.

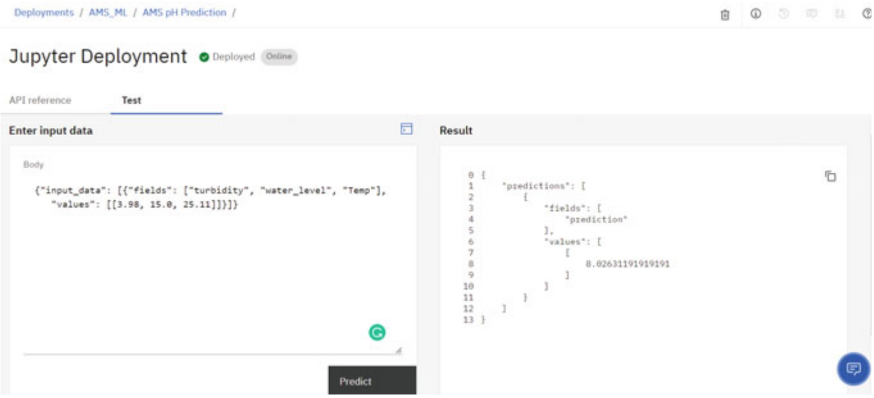
From the above graph and tabulation (shown in Figs. 15a, b), it can be seen that the difference between the original and predicted values of the pH are negligible and therefore we can safely implement the system without using the pH sensor by soft sensing the parameter using temperature, turbidity, and water level parameters. This model is uploaded to the IBM cloud using the application program interface (API) client and deployed (shown in Fig. 16).

After uploading the model in IBM, the test tab can be used to test the model by giving the input fields in JavaScript object notation (JSON) format and the results return the predicted value of the pH. This value is displayed in the IBM dashboard and also sent to the android app.



	Actual	Predicted
0	8.07	8.042326
1	8.08	8.036049
2	8.04	8.052555
3	7.20	7.185648
4	7.45	7.376822

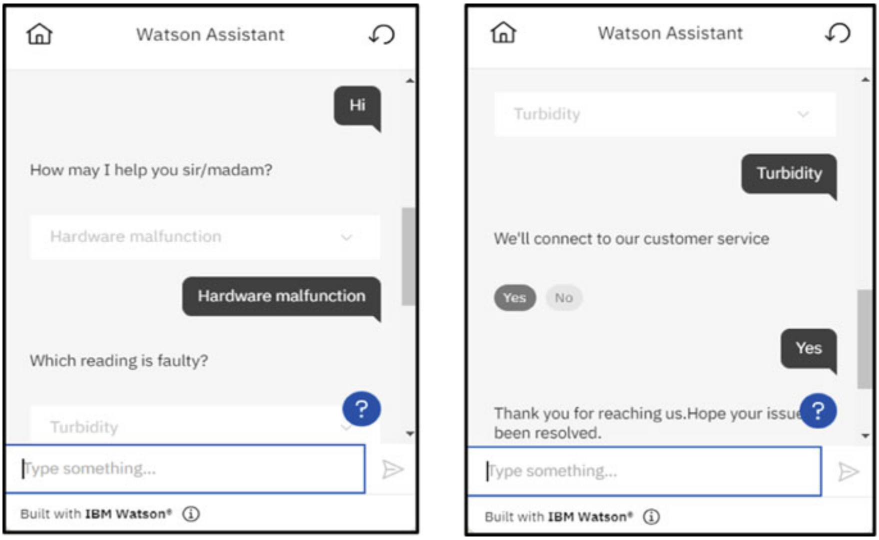
Fig. 15 **a** Original pH values versus predicted pH values, **b** examples of original pH values versus predicted pH values



```
{"input_data": [{"fields": ["turbidity", "water_level", "Temp"],
   "values": [[3.98, 15.0, 25.11]]}]}
```

```
{
    "predictions": [
        {
            "fields": [
                "prediction"
            ],
            "values": [
                [
                    8.02631191919191
                ]
            ]
        }
    ]
}
```

Fig. 16 Deployment in IBM cloud



Hi

How may I help you sir/madam?

Hardware malfunction

Hardware malfunction

Which reading is faulty?

Turbidity

Turbidity

Turbidity

We'll connect to our customer service

Yes No

Yes

Thank you for reaching us.Hope your issu
been resolved.

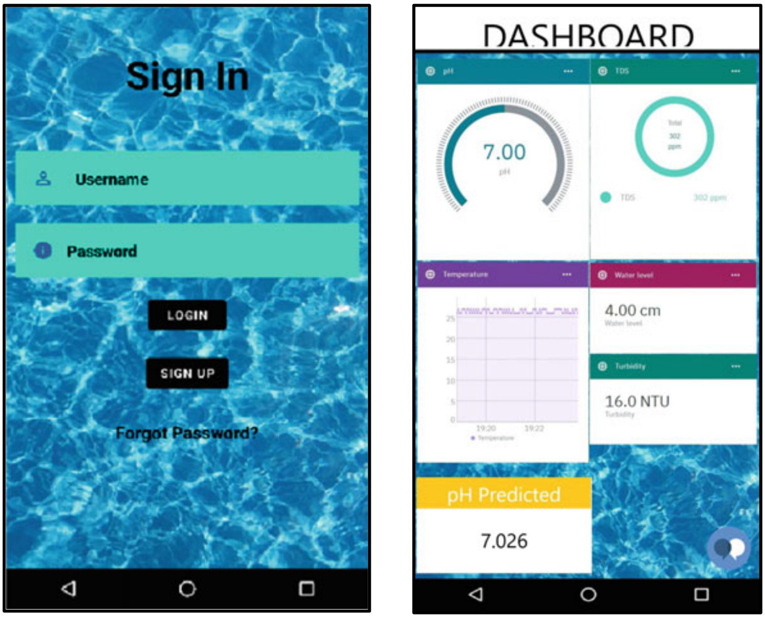
Fig. 17 Chatbot interface

5.3 Chatbot

The Chatbot service is essential for the user to understand the nuances about the operations of the system. A user might encounter some trouble with regards to the system either in the form of hardware malfunction, data visualization, connectivity issues, or more problems. For resolving these issues as well as to learn about a particular component of a function of the system, the chatbot would be beneficial in terms of providing the solution (shown in Fig. 17). It is built in with standard commands with regards to the configuration of the system as well as to provide information with regards to the main components of the system. The chatbot follows a user-friendly approach which makes the user interact with the chatbot service with ease. It also helps in connecting the user with the customer support team when the issue is beyond the scope of standard queries.

5.4 Android App

The login page (shown in Fig. 18a) is the page where the users sign up or login into their accounts. Upon successful login, the app displays the dashboard (shown in Fig. 18b) with the real-time updates of all the parameters monitored in their system. The chatbot icon is displayed in the bottom right corner of all the pages. The navigation panel allows the users to navigate between the database and the maintenance pages.



a) Login screen

b) Dashboard

Fig. 18 Android app

6 Conclusion and Future Scope

This work designs and implements a monitoring system that not only monitors and displays the results to the farmers but also is a step towards efficient soft sensing through its prediction model, therefore, reducing the costs involved in the hardware and its maintenance. The system also provides customer support and information regarding the important aspects of aquaculture thereby enabling the farmers to make well-informed decisions. The app lets the farmer access the state of their aquaculture farm from anywhere with the use of the internet and solves the dependence on manual testing.

Aquaculture monitoring, though a well-researched field, provides ample opportunities for research and new inventions. More actuators can be included in the system and automated thereby fully leveraging the latest technologies. The scope of research in the area of soft sensing is immense and can be improvised to predict all the important parameters.

References

1. Fisheries statistics. https://dof.gov.in/sites/default/files/2021-08/Handbook_on_Fishery_2020.pdf
2. Nisar U, Zhang H, Navghan M et al. (2021) Comparative analysis of profitability and resource use efficiency between *Penaeus monodon* and *Litopenaeus vannamei* in India. PLOSONE J
3. Dixon W, Chiswell B (1996) Review of aquatic monitoring program design. Pergamon J 30(9)
4. Nash CE (2011) The history of aquaculture
5. Quach A, Murray F, Morrison-Saunders A (2017) The vulnerability of shrimp farming income to climate change events. Int J Clim Change Strat Manage. Ca Mau, Vietnam
6. Kumar S, Tiwari P, Zymbler M (2019) Internet of Things is a revolutionary approach for future technology enhancement. J Big Data
7. Kochut A, Deng Y, Head M, Munson J, Sailer A et.al. (2011) Evolution of the IBM Cloud: enabling an enterprise cloud services ecosystem. IBM J Res Dev
8. Tratar LF, Mojškerc B, Toman A (2016) Demand forecasting with four-parameter exponential smoothing. Int J Prod Econ Part A 181
9. Saha S, Hasan Rajib R, Kabir S (2018) IoT based automated fish farm aquaculture monitoring system. In: International conference on innovations in science, engineering and technology (ICISET)
10. Wang C, Li Z, Wang T, Xu X et al. (2021) Intelligent fish farm—the future of aquaculture. Aquac Int
11. Hu Z, Zhang Y, Zhao Y et al. (2019) A water quality prediction method based on the deep LSTM network considering correlation in smart mariculture. Sens J
12. Gómez-Chabla R, Real-Avilés K, Delgado-Vera C, Chávez C, Vera-Lucio N (2018) Monitoring system for shrimp farming. In: CITI 2018. Communications in computer and information science, vol 883. Springer, Cham
13. Cordova-Rozas M, Aucapuri-Lecarnaque J, Shiguihara-Juárez P (2019) a cloud monitoring system for aquaculture using IoT. In: 2019 IEEE sciences and humanities international research conference (SHIRCON)
14. Li D, Liu S (2013) Remote monitoring of water quality for intensive fish culture. In: Mukhopadhyay S, Mason A (eds) Smart sensors for real-time water quality monitoring. Smart sensors, measurement and instrumentation, vol 4. Springer, Berlin, Heidelberg
15. Christina G (Sept2021) A review on microstrip patch antenna performance improvement techniques on various applications. J Trends Comput Sci Smart Technol 03(03):175–189
16. Hamdan YB (2021) Construction of statistical SVM based recognition model for handwritten character recognition. J Inf Technol 3(02):92–107
17. Dataset for assessing the various parameters of aquaculture monitoring. https://github.com/pkErbynn/IoT-WQMS
18. Dataset for aquaculture monitoring. https://www.kaggle.com/alexisaucapuri/fishfarm-monitoring

A Comprehensive Study and Implementation of Memory Malware Analysis with Its Application for the Case Study of CRIDEX

Digvijay Singh and Rajesh Yadav

Abstract The advancement in technology and the rising demand for the internet have witnessed an increased cyber threat which can be highlighted as a major challenge. Cyber criminals make use of malware programmes to fulfil the malicious purpose. Malware is software coded and designed to bring harm to the target machine by various means. This paper focuses on describing memory analysis techniques under malware analysis with a comparative analysis of different tools. The present paper highlights the application to the case study of CRIDEX malware for a better understanding through a practical implementation of appropriate tools. With its initial appearance as CRIDEX, the bank stealing trojan has evolved in the past decade and has been witnessed to spread malware infection through its new variants as discovered in the past year for which it has been selected as an appropriate candidate to be analysed for understanding its basic working.

Keywords Malware · Memory analysis · Artifacts · Memory dump · Cridex

1 Introduction

Malware is an intrusive software purposefully designed by cybercriminals with intentions of a potential threat. Malwares can steal, corrupt and destroy data. Each malware has a special characteristic which makes it distinct from another malware. Memory analysis is a powerful technique of Malware analysis, it deals with capturing physical memory of the infected state of a machine and later performing a complete analysis of that memory. This technique has been proven efficient for performing malware analysis and is also broadly categorized under memory forensics [1]. It came into existence in recent years.

D. Singh · R. Yadav (✉)
Department of Computer Science, BML Munjal University, Gurgaon, Haryana, India
e-mail: rajesh.yadav@bmu.edu.in

D. Singh
e-mail: digvijay.singh.18csc@bmu.edu.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_3

Memory analysis fulfils the purpose of performing a digital investigation thus giving useful artifacts. An analyst uses this technique by analysing a dumped memory of an infected state of the target machine. Once the malware is executed, it makes modifications to the system and these are obtained via artifacts similar to the ones being used for investigation [2]. This research describes an approach which can be applied by an analyst to perform memory analysis on different malware samples. We will be showing a comparison of different tools to differentiate the features and identify the appropriate tools capable of giving faster and accurate results. We have included Cridex malware to analyse the infected state of the memory dump captured post its execution.

The remainder of this paper is organised as follows: Sect. 2 talks about the background work, Sect. 3 states the Literature survey. In Sect. 4 we will discuss the Methodology and Sect. 5 will show the implementation of the case study followed by a comparative analysis of the different tools. Section 6 highlights the results obtained by the analysis and Sect. 7 concludes the paper.

2 Background Work

Memory analysis technique is popular for providing a comprehensive view of the infected system. Being efficient and delivering accurate results, it is encouraged by both analysts and forensic investigators. It is divided in two phases—Memory Acquisition and Memory Dump Analysis. Memory acquisition refers to the process of acquiring or capturing a physical memory state of an infected system by using appropriate tools [3] followed by the second phase of analysing that memory dump using suitable features provided by the tool.

Memory Analysis

Memory is an essential component used for storing information. The Central Processing Unit works synchronously with the memory unit for the storage operations. When we talk about Malware Memory Analysis, we focus on the RAM [4]. It is responsible for allowing the data to read or write independent of the physical memory. For memory management, we typically have a stack and heap memory. Stack works on the principle of push and pop operation, making it grow eventually. For Dynamic memory allocation, a heap memory is used.

- Memory Acquisition—This stage deals with the dumping of an infected memory state. Once a malware is executed and infects the system, it gets loaded in the RAM. It is essential to dump this memory using a suitable tool for completing further analysis.
- Memory Dump Analysis—This stage deals with analyzing a dumped memory state. An analyst uses the acquired state for performing operations to extract information like process information, network related information and much more [5]. For understanding this stage any memory dump can be used. There are various

Table 1 Different types of malware and their impacts

Types of malware	Impact
Virus	Infects application by exploiting vulnerabilities
Trojan	Performs malicious activities in the background
Spyware	Monitors user activities while aiming for sensitive information
Bot	Automates infection spreading tactics
Ransomware	Demands a ransom against any sort of attack
Wiper	Usually wipes off the data and memory
Worm	Spread the infection over a network and flood the network servers
Rootkit	Usually conceals the presence and is accessible remotely

databases containing memory dump of an infected state caused by malware. Once the memory state is acquired the tools are used for gathering as much information to analyze the impact caused by the malware. Different types of malware and their impacts are shown in Table 1.

Cridex Malware

To strengthen our application-based study we have included a Cridex malware sample in our case study. It is banking malware capable of stealing the banking-credentials and other vital information [6]. It is categorized as a Trojan which can gain complete access to the infected system. It has a capability of spreading via copying itself in different drives.

3 Related Work

The research papers based on Malware Analysis focuses broadly on different techniques of analyzing malware. Most common techniques involve Static and Dynamic analysis; however, some researchers have highlighted the importance of Hybrid and Memory analysis.

Seo et al. [7] has described different ways of gathering memory data. The authors have included dumping techniques and highlighted the difficulties while dumping a pure memory and also discussed the challenges of dumping a physical memory of RAM more than 4 GB. Manson et al. [8] has shown a comparison between three different tools- Autopsy, FTK imager and Encase. The comparative analysis by the authors mentions that the image import was done fastest by Encase but the user experience was reported not smooth for the analysts. Apart from Encase, the other two tools were much easier to be used and FTK imager is best suited for windows. The authors have highlighted its intuitive interface. Okolica et al. [9] has introduced the CMAT tool. It is capable of parsing memory dump to identify all processes. It works on the principle of extracting data from a live memory by compiling the responses into a suitable form for further analysis. Carvajal et al. [10] have done a comparison of six tools which are used for analyzing the memory. The parameters for comparison are based on the interface experience, artifacts, processing time, reports. The study tells that the GUI based tools left more artifacts than the CLI based. The authors of [11] have done a survey on tools and techniques for memory acquisition and analysis which specifically deal with windows-based systems. The authors of [12] have illustrated the prevalent concepts of creating memory snapshot and has also shown methods of analyzing the memory data. Kim et al. [13] have developed a techniue which can be used for classification of malwares. The importance of malware classification is highlighted in order to bring attention for identification of malware and their families. It helps to understand different patterns related to a malware infection. The author of [14] has proposed an approach to identify malicious android applications by providing a complete mechanism. Current techniques were compared using different datasets. The algorithm can be considered as a complete in terms of solving real challenges linked to android categorization. The authors of [15] have introduced memory aware cloud scheduling for a faster scheduling. The idea is to improve overall security and maintaining an error free process. It was tested through simulations with total value of 50 applications. Ori et al. [16] have shown a survey to overview different dynamic analysis methods. The authors have also focused on classification and behavioural analysis. The paper discuses about Memory forensic technique which brings attention towards this approach of analysing the malware.

4 Memory Malware Analysis Approach

Memory Analysis is about capturing a malware infected state of memory by acquiring RAW memory and further analyzing the memory dump to perform an assessment of different parameters. An analyst retrieves vital information and associates it with the artifact which could be left unnoticed if not analyzed.

Memory analysis is performed mainly for three reasons

- An analyst can retrieve real time data from the physical memory.
- The encrypted data is decrypted once it is acquired in the RAW memory.
- Information of higher significance can be obtained by this technique, which can be difficult while performing static and dynamic analysis [17].

Memory Acquisition

Acquisition is the technique of acquiring/Dumping a physical state of memory. When a system is infected by malware, there are equal chances of the system going to a dead or alive state. The execution of malware makes the necessary changes as coded by the threat actor. If the system is alive then different tools capable of capturing the memory can be used for making a memory dump and if the system is dead then a forensic disc controller can be used to collect the data and the remaining process remains the same. To ensure that there are no changes in the Memory Dump, a hash value can be calculated and verified as required [18]. Some of the common tools which can be used to acquire the physical memory state are- FTK imager, Autopsy Forensic Browser, Encase, Nuix, Hibernation Recon and Forensic explorer.

Memory Dump analysis

Once an analyst acquires a memory dump the next step is to analyze it. Memory Dump analysis is the technique of analyzing an infected memory dump to extract vital information that will be used to analyze a malware. This technique focusses on analysing the artefacts from the memory dump [19] as in Fig. 1. Depending on the size of the memory dump, appropriate tools analyze the memory file and the processing time is directly proportional to the size of that particular dump. It is also observed that a larger dump is capable of producing more artifacts than a smaller one. This phase can be applied on any available memory dump. An analyst can examine memory dump either by acquiring through capturing or simply use an existing memory dump available on the internet for the desired malware sample. Potential information which can be extracted through this technique generally includes: suspicious processes, DDL and handles, network related information, injected code, Dump processes and drivers, kernel memory etc. Some of the common tools which can be used to analyze a memory dump are- Volatility, Autopsy Forensic Browser, Caine, Rekall, Cado Live and Redline.

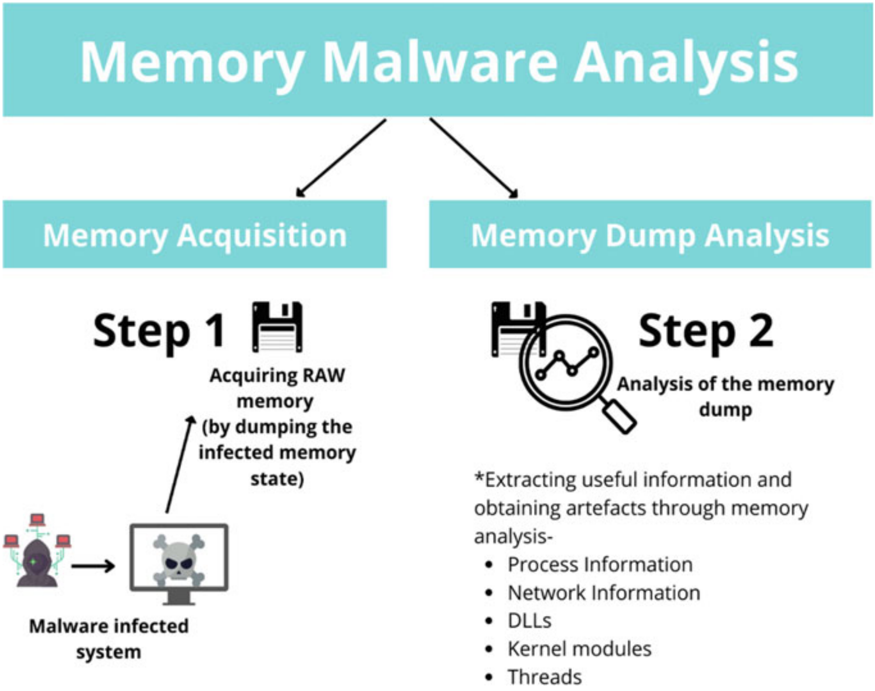


Fig. 1 Memory malware analysis process

5 Available Tools

FTK imager: FTK Imager is one of the best open-source tools used for creation of disk images. It provides an interactive user interface with advanced capabilities. An analyst uses this tool to acquire images without making changes in the system. The tool is capable of recovering the files which are deleted from the Recycle Bin but the changes are not overwritten on the drive. It allows the user to mount an image to view the contents of the acquired image as on the original drive.

Autopsy Forensic Browser: Autopsy Forensic Browser is a GUI based application which is capable of analysing hard drives and smartphones. It offers file type identification, registry analysis, data extraction from Android and is used for military investigations. It is extensively used for mobile device and digital media forensics.

EnCase: EnCase tool is considered as a Global standard in digital investigation. It efficiently conducts the needs of investigators and analysts by using a repeatable and defensible process. It supports a broad range of file systems. It offers automation in the analysis of artifacts which proves to be time saving unlike other memory analysis tools. The wide range of features offered by the tool are not completely free.

Forensic Explorer: Forensic Explorer (FEX) is a tool which is mainly used for preservation and analysis of memory. It gives a suitable experience to its users including law enforcement, government and military agencies. It offers a simple and interactive GUI experience providing multiple functionalities at a time. The tool is capable of processing a larger volume of data and provides detailed report covering different aspects like- File and Registry Analysis, Boot Virtualization etc.

Volatility: Volatility is an open-source framework for memory analysis. Currently it is one of the most popular tools supporting Windows, Mac and Linux. The python-based framework is best suited to analyse the RAM in 32/64-bit systems. It allows us to analyse different parameters including- date and time of the image captured, Network related information, Memory address, Process related information, libraries loaded and kernel modules. Volatility is recommended for memory analysis because of its faster and efficient algorithm of analysing RAM dumps. Since it is compatible for different operating systems, it has a larger repository storing profiles for different Operating Systems.

Memoryze: Memoryze is a memory analysis tool. It was formerly known as MANDIANT Free Agent. This tool is not only limited towards acquisition of physical memory rather it can perform advanced memory analysis while the system is in a live state. It can image a complete address of a process to a disk including dlls and exe’s. Also, it can verify the digital signature of dlls and exe’s.

6 CRIDEX Case Study

Memory Acquisition.

This step deals with capturing memory of the system. The idea is to make a memory dump of the malware infected. FTK imager was used to acquire RAW. memory. User Interface of FTK imager tool is shown in Fig. 2.

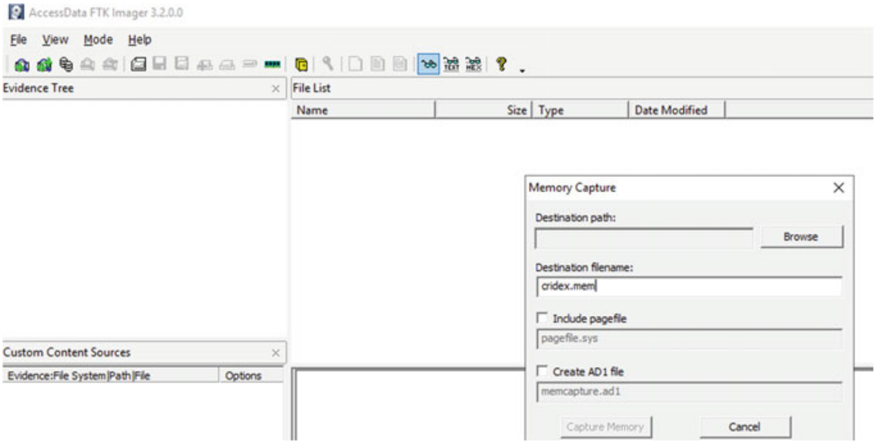
Memory Analysis

This step deals with analysis of the memory dump as acquired previously. The idea is to gain information by monitoring memory changes. It reveals process related information which helps an analyst to examine suspicious activities and produce maximum artifacts. We have used Volatility framework to fulfil the purpose of analysing memory. A view of command line interface is shown in Fig. 3.

To begin with the analysis, we provided the cridex.mem file and applied various operations supported by volatility framework. Firstly, we used "volatility.exe -f cridex.mem imageinfo" to determine the profile based on kDBG search. By using this command, we identified that the tool hints on targeted OS as WindowsXP as shown in Fig. 4.

The next step was to list an overview of the running processes using "volatility.exe -f cridex.mem pslist". More than 15 running processes were identified as shown in Fig. 5.

Out of the 17 processes listed in the above image 13 of them are executed by Windows OS, however 4 processes reader_s1.exe, alg.exe and wuaclt.exe (listed twice) are not part of windows services hence suspicious.


AccessData FTK Imager 3.2.0.0

Memory Capture

Destination path:

Destination filename: cridex.mem

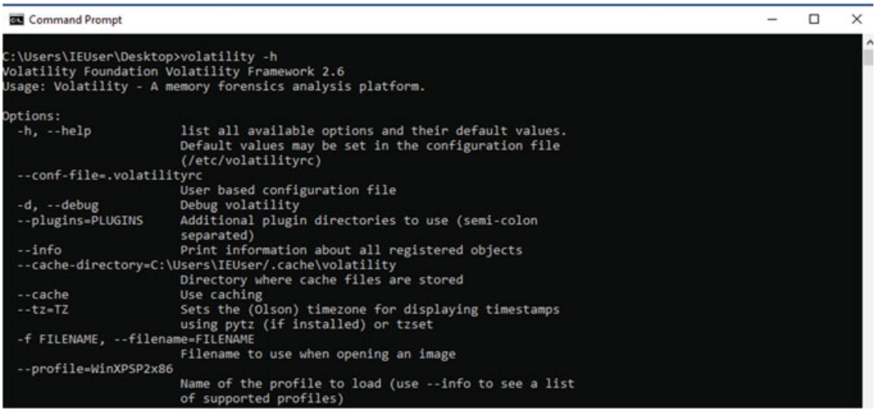
□ Include pagefile

pagefile.sys

□ Create AD1 file

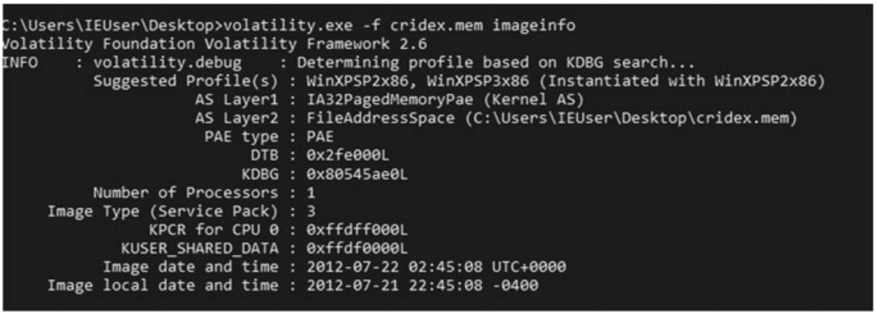
memcapture.ad1

Fig. 2 User Interface of FTK imager tool



```
:\Users\IEUser\Desktop>volatility -h
olatility Foundation Volatility Framework 2.6
sage: Volatility - A memory forensics analysis platform.

ptions:
  -h, --help            list all available options and their default values.
                        Default values may be set in the configuration file
                        (/etc/volatilityrc)
  --conf-file=.volatilityrc
                        User based configuration file
  -d, --debug           Debug volatility
  --plugins=PLUGINS     Additional plugin directories to use (semi-colon
                        separated)
  --info                Print information about all registered objects
  --cache-directory=C:\Users\IEUser/.cache\volatility
                        Directory where cache files are stored
  --cache               Use caching
  --tz=TZ               Sets the (Olson) timezone for displaying timestamps
                        using pytz (if installed) or tzset
  -f FILENAME, --filename=FILENAME
                        Filename to use when opening an image
  --profile=WinXPSP2x86
                        Name of the profile to load (use --info to see a list
                        of supported profiles)
```

Fig. 3 User ınterface of volatility framework



```
:\Users\IEUser\Desktop>volatility.exe -f cridex.mem imageinfo
olatility Foundation Volatility Framework 2.6
NFO    : volatility.debug    : Determining profile based on KDBG search...
          Suggested Profile(s) : WinXPSP2x86, WinXPSP3x86 (Instantiated with WinXPSP2x86)
                     AS Layer1 : IA32PagedMemoryPae (Kernel AS)
                     AS Layer2 : FileAddressSpace (C:\Users\IEUser\Desktop\cridex.mem)
                      PAE type : PAE
                           DTB : 0x2fe000L
                          KDBG : 0x80545ae0L
          Number of Processors : 1
     Image Type (Service Pack) : 3
                KPCR for CPU 0 : 0xffdff000L
             KUSER_SHARED_DATA : 0xffdf0000L
           Image date and time : 2012-07-22 02:45:08 UTC+0000
     Image local date and time : 2012-07-21 22:45:08 -0400
```

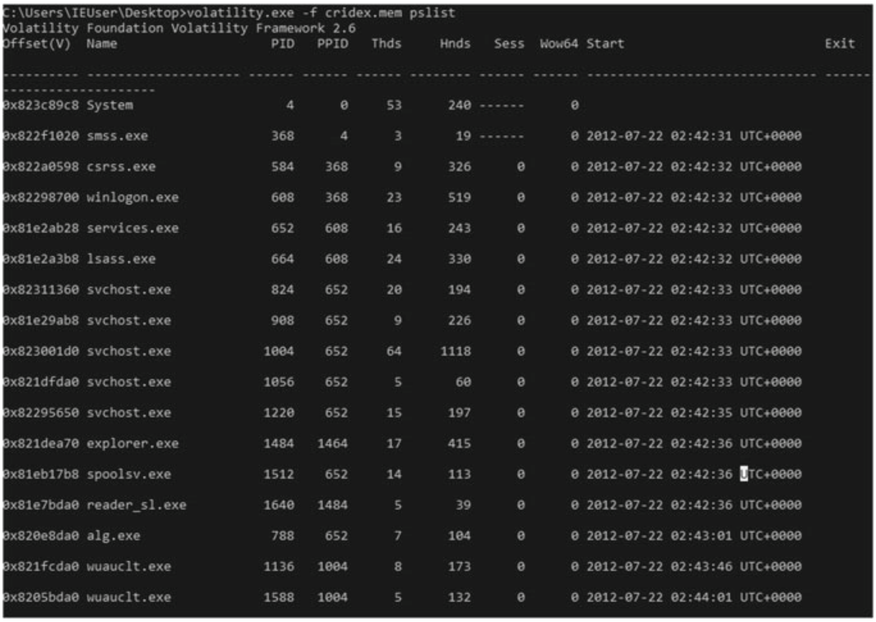
Fig. 4 Profiles associated with the acquired memory dump

Further to listing running processes we displayed the the processes as a parent and child tree using "volatility.exe -f cridex.mem pstree". Similar to pslist this command also highlights the suspicious processes but alg.exe and wuauclt.exe are indicated as a part of winlogon.exe which makes them non suspicious however reader_s1.exe is shown as standalone as shown in Fig. 6.

Process reader_s1.exe has Pid 1640 and PPid 1484 which means it is directly running under explorer.exe. The focus was towards identifying this suspicious process.

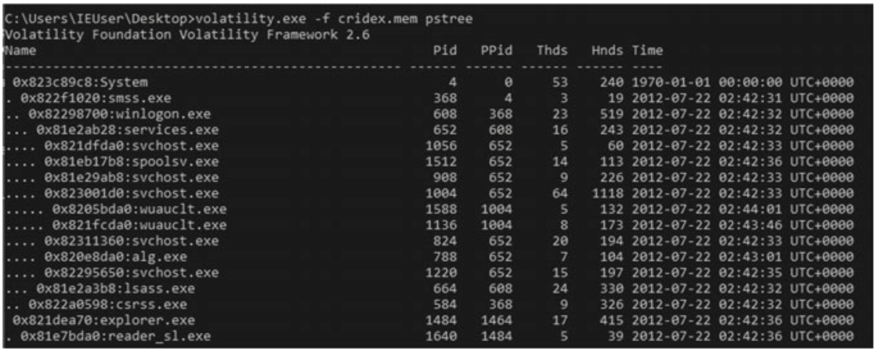
We used "volatility.exe -f cridex.mem connscan" to get the network related information and we were able to identify the local and remote address linked to Pid 1484 as in Fig. 7.

By using "volatility.exe -f cridex.mem sockets" we identified that the port number of PID:1484 is 1038 which provides the service for Message tracking Query Protocol as shown below in Fig. 8.



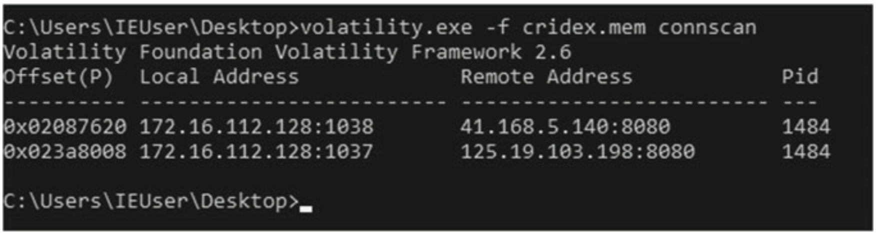
```
C:\Users\IEUser\Desktop>volatility.exe -f cridex.mem pslist
Volatility Foundation Volatility Framework 2.6
Offset(V)  Name                    PID   PPID   Thds     Hnds   Sess  Wow64 Start                          Exit

---------- -------------------- ------ ------ ------ -------- ------ ------ ------------------------------ ------
--------------------
0x823c89c8 System                    4      0     53      240 ------      0

0x822f1020 smss.exe                368      4      3       19 ------      0 2012-07-22 02:42:31 UTC+0000

0x822a0598 csrss.exe               584    368      9      326      0      0 2012-07-22 02:42:32 UTC+0000

0x82298700 winlogon.exe            608    368     23      519      0      0 2012-07-22 02:42:32 UTC+0000

0x81e2ab28 services.exe            652    608     16      243      0      0 2012-07-22 02:42:32 UTC+0000

0x81e2a3b8 lsass.exe               664    608     24      330      0      0 2012-07-22 02:42:32 UTC+0000

0x82311360 svchost.exe             824    652     20      194      0      0 2012-07-22 02:42:33 UTC+0000

0x81e29ab8 svchost.exe             908    652      9      226      0      0 2012-07-22 02:42:33 UTC+0000

0x823001d0 svchost.exe            1004    652     64     1118      0      0 2012-07-22 02:42:33 UTC+0000

0x821dfda0 svchost.exe            1056    652      5       60      0      0 2012-07-22 02:42:33 UTC+0000

0x82295650 svchost.exe            1220    652     15      197      0      0 2012-07-22 02:42:35 UTC+0000

0x821dea70 explorer.exe           1484   1464     17      415      0      0 2012-07-22 02:42:36 UTC+0000

0x81eb17b8 spoolsv.exe            1512    652     14      113      0      0 2012-07-22 02:42:36 UTC+0000

0x81e7bda0 reader_sl.exe          1640   1484      5       39      0      0 2012-07-22 02:42:36 UTC+0000

0x820e8da0 alg.exe                 788    652      7      104      0      0 2012-07-22 02:43:01 UTC+0000

0x821fcda0 wuauclt.exe            1136   1004      8      173      0      0 2012-07-22 02:43:46 UTC+0000

0x8205bda0 wuauclt.exe            1588   1004      5      132      0      0 2012-07-22 02:44:01 UTC+0000
```

Fig. 5 Process list generated by volatility framework



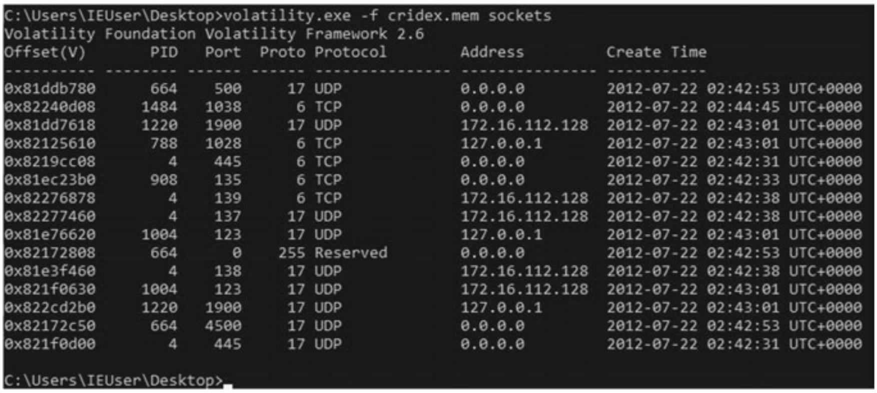
```
C:\Users\IEUser\Desktop>volatility.exe -f cridex.mem pstree
Volatility Foundation Volatility Framework 2.6
Name                                                  Pid   PPid   Thds   Hnds Time
-------------------------------------------------- ------ ------ ------ ------ ----
 0x823c89c8:System                                      4      0     53    240 1970-01-01 00:00:00 UTC+0000
. 0x822f1020:smss.exe                                 368      4      3     19 2012-07-22 02:42:31 UTC+0000
.. 0x82298700:winlogon.exe                            608    368     23    519 2012-07-22 02:42:32 UTC+0000
... 0x81e2ab28:services.exe                           652    608     16    243 2012-07-22 02:42:32 UTC+0000
.... 0x821dfda0:svchost.exe                          1056    652      5     60 2012-07-22 02:42:33 UTC+0000
.... 0x81eb17b8:spoolsv.exe                          1512    652     14    113 2012-07-22 02:42:36 UTC+0000
.... 0x81e29ab8:svchost.exe                           908    652      9    226 2012-07-22 02:42:33 UTC+0000
.... 0x823001d0:svchost.exe                          1004    652     64   1118 2012-07-22 02:42:33 UTC+0000
..... 0x8205bda0:wuauclt.exe                         1588   1004      5    132 2012-07-22 02:44:01 UTC+0000
..... 0x821fcda0:wuauclt.exe                         1136   1004      8    173 2012-07-22 02:43:46 UTC+0000
.... 0x82311360:svchost.exe                           824    652     20    194 2012-07-22 02:42:33 UTC+0000
.... 0x820e8da0:alg.exe                               788    652      7    104 2012-07-22 02:43:01 UTC+0000
.... 0x82295650:svchost.exe                          1220    652     15    197 2012-07-22 02:42:35 UTC+0000
... 0x81e2a3b8:lsass.exe                              664    608     24    330 2012-07-22 02:42:32 UTC+0000
.. 0x822a0598:csrss.exe                               584    368      9    326 2012-07-22 02:42:32 UTC+0000
 0x821dea70:explorer.exe                             1484   1464     17    415 2012-07-22 02:42:36 UTC+0000
. 0x81e7bda0:reader_sl.exe                           1640   1484      5     39 2012-07-22 02:42:36 UTC+0000
```

Fig. 6 Prcoess tree indicating parent child relation between processes



```
C:\Users\IEUser\Desktop>volatility.exe -f cridex.mem connscan
Volatility Foundation Volatility Framework 2.6
Offset(P)  Local Address             Remote Address            Pid
---------- ------------------------- ------------------------- ---
0x02087620 172.16.112.128:1038       41.168.5.140:8080         1484
0x023a8008 172.16.112.128:1037       125.19.103.198:8080       1484

C:\Users\IEUser\Desktop>_
```

Fig. 7 Network information fetched through the memory



```
C:\Users\IEUser\Desktop>volatility.exe -f cridex.mem sockets
Volatility Foundation Volatility Framework 2.6
Offset(V)       PID   Port  Proto Protocol        Address         Create Time
---------- -------- ------ ------ --------------- --------------- -----------
0x81ddb780      664    500     17 UDP             0.0.0.0         2012-07-22 02:42:53 UTC+0000
0x82240d08     1484   1038      6 TCP             0.0.0.0         2012-07-22 02:44:45 UTC+0000
0x81dd7618     1220   1900     17 UDP             172.16.112.128  2012-07-22 02:43:01 UTC+0000
0x82125610      788   1028      6 TCP             127.0.0.1       2012-07-22 02:43:01 UTC+0000
0x8219cc08        4    445      6 TCP             0.0.0.0         2012-07-22 02:42:31 UTC+0000
0x81ec23b0      908    135      6 TCP             0.0.0.0         2012-07-22 02:42:33 UTC+0000
0x82276878        4    139      6 TCP             172.16.112.128  2012-07-22 02:42:38 UTC+0000
0x82277460        4    137     17 UDP             172.16.112.128  2012-07-22 02:42:38 UTC+0000
0x81e76620     1004    123     17 UDP             127.0.0.1       2012-07-22 02:43:01 UTC+0000
0x82172808      664      0    255 Reserved        0.0.0.0         2012-07-22 02:42:53 UTC+0000
0x81e3f460        4    138     17 UDP             172.16.112.128  2012-07-22 02:42:38 UTC+0000
0x821f0630     1004    123     17 UDP             172.16.112.128  2012-07-22 02:43:01 UTC+0000
0x822cd2b0     1220   1900     17 UDP             127.0.0.1       2012-07-22 02:43:01 UTC+0000
0x82172c50      664   4500     17 UDP             0.0.0.0         2012-07-22 02:42:53 UTC+0000
0x821f0d00        4    445     17 UDP             0.0.0.0         2012-07-22 02:42:31 UTC+0000

C:\Users\IEUser\Desktop>
```

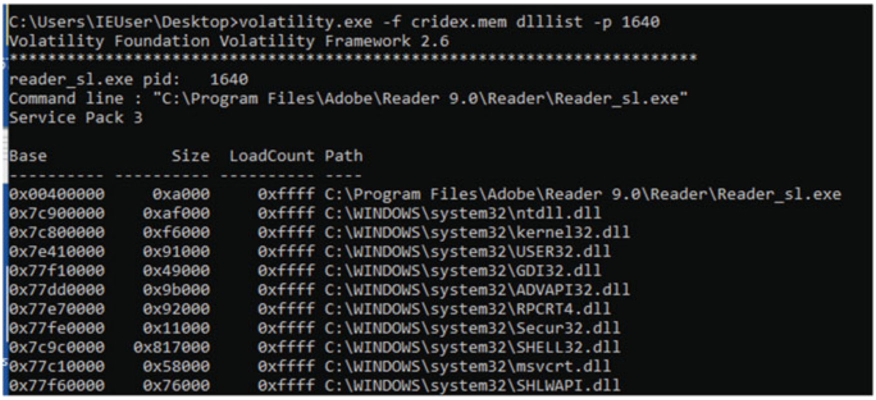
Fig. 8 Scoket information displaying port information for respective PID's

Since Pid 1640 was identified to use PPid 1484 which was marked suspicious in the aforementioned steps, then we used "volatility.exe -f cridex.mem dlllist -p 1640" to get a list of DLLs. The information extracted through this command shows that the process uses kernel 32 and network dll as shown below in Fig. 9.

With the strengthening of our doubt against process with Pid 1640 we used "volatility.exe -f cridex.mem procdump -p 1640 –dum-dir." to dump the file in the existing directory as shown below im Fig. 10.

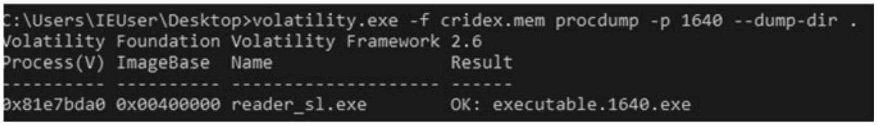
As soon as the command was executed, executable.1640.exe file appeared on desktop as shown in Fig. 11.

To analyse the nature of the suspicious file we uploaded the sample on virus total engine and identified that the file was a trojan as shown in Fig. 12.



```
C:\Users\IEUser\Desktop>volatility.exe -f cridex.mem dlllist -p 1640
Volatility Foundation Volatility Framework 2.6
************************************************************************
reader_sl.exe pid:   1640
Command line : "C:\Program Files\Adobe\Reader 9.0\Reader\Reader_sl.exe"
Service Pack 3

Base             Size  LoadCount Path
---------- ---------- ---------- ----
0x00400000     0xa000     0xffff C:\Program Files\Adobe\Reader 9.0\Reader\Reader_sl.exe
0x7c900000    0xaf000     0xffff C:\WINDOWS\system32\ntdll.dll
0x7c800000    0xf6000     0xffff C:\WINDOWS\system32\kernel32.dll
0x7e410000    0x91000     0xffff C:\WINDOWS\system32\USER32.dll
0x77f10000    0x49000     0xffff C:\WINDOWS\system32\GDI32.dll
0x77dd0000    0x9b000     0xffff C:\WINDOWS\system32\ADVAPI32.dll
0x77e70000    0x92000     0xffff C:\WINDOWS\system32\RPCRT4.dll
0x77fe0000    0x11000     0xffff C:\WINDOWS\system32\Secur32.dll
0x7c9c0000   0x817000     0xffff C:\WINDOWS\system32\SHELL32.dll
0x77c10000    0x58000     0xffff C:\WINDOWS\system32\msvcrt.dll
0x77f60000    0x76000     0xffff C:\WINDOWS\system32\SHLWAPI.dll
```

Fig. 9 List of DLLs obtained through volatility



```
:\Users\IEUser\Desktop>volatility.exe -f cridex.mem procdump -p 1640 --dump-dir .
olatility Foundation Volatility Framework 2.6
rocess(V) ImageBase  Name                 Result
---------- ---------- -------------------- ------
x81e7bda0 0x00400000 reader_sl.exe        OK: executable.1640.exe
```

Fig. 10 Creation of a dump file for process 1640

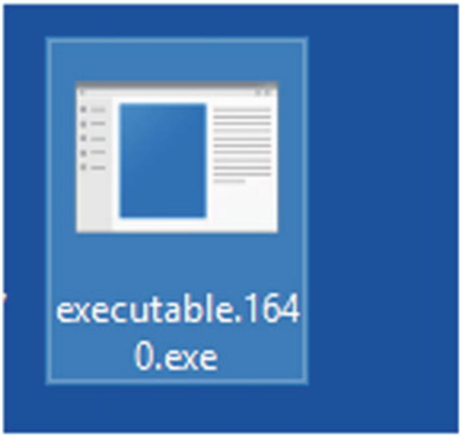
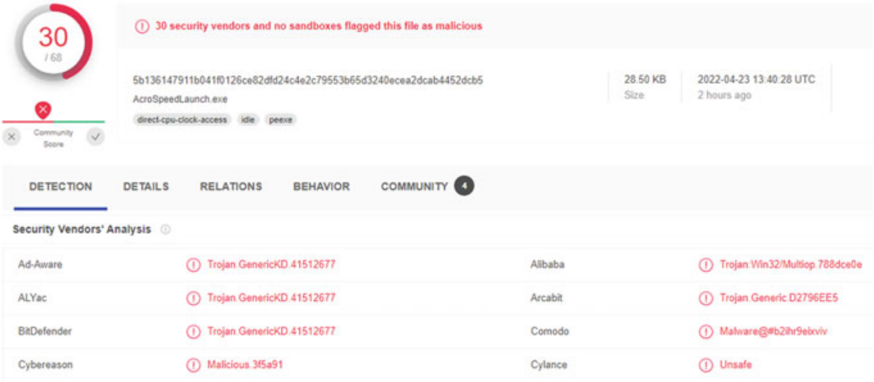


Fig. 11 An exe file being generated on desktop after dumping the process



30 security vendors and no sandboxes flagged this file as malicious

5b136147911b041f0126ce82dfd24c4e2c79553b65d3240ecea2dcab4452dcb5
AcroSpeedLaunch.exe
direct-cpu-clock-access idle peexe

28.50 KB
Size

2022-04-23 13:40:28 UTC
2 hours ago

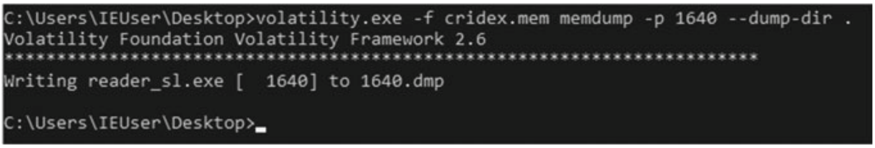
Security Vendors' Analysis

Ad-Aware	Trojan.GenericKD.41512677	Alibaba	Trojan:Win32/Multiop.788dce0e
ALYac	Trojan.GenericKD.41512677	Arcabit	Trojan.Generic.D2796EE5
BitDefender	Trojan.GenericKD.41512677	Comodo	Malware@#b2ihr9ebxviv
Cybereason	Malicious.3f5a91	Cylance	Unsafe

Fig. 12 API based detection using virus total

The previous step confirms our doubt of the suspicious process with Pid:1640. Now we used memdump to create a.dmp file to further extract relevant information as shown in Fig. 13.

Once the ".dmp" file was generated, we used strings command line utility to extract the strings from the file to identify more information. The trojan consisted of several strings containing website addresses of different banks. Similar bank related



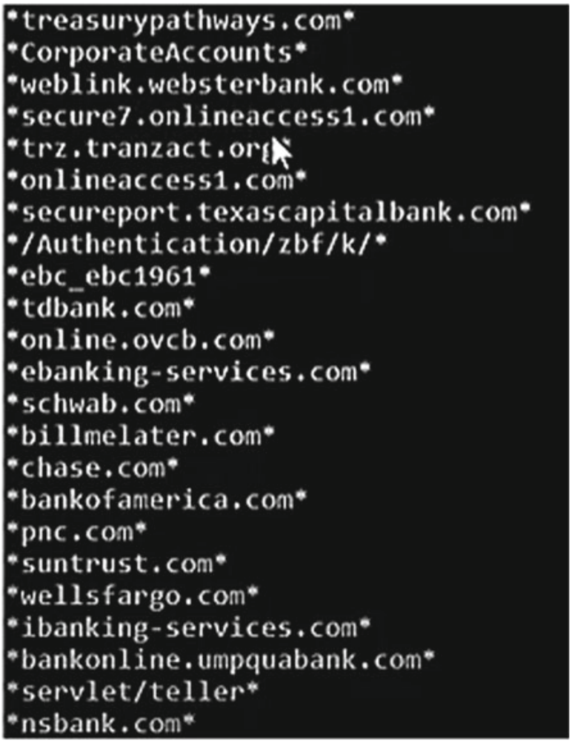
```
C:\Users\IEUser\Desktop>volatility.exe -f cridex.mem memdump -p 1640 --dump-dir .
Volatility Foundation Volatility Framework 2.6
************************************************************************
Writing reader_sl.exe [  1640] to 1640.dmp

C:\Users\IEUser\Desktop>_
```

Fig. 13 Creation of.dmp file

information was revealed through other strings. A view of extracted strings have been shown in Fig. 14.

The experimental findings indicate that the process with PID 1484 is suspicious, the process uses port 1038. Port number 1038 provides a Message tracking Query protocol. On dumping the file application, an exe file was visible on desktop. Executable 1640.exe was used as an input on virus total. The tool Classified the file as a trojan which was further examined by doing string analysis. On performing string analysis, it was revealed that the trojan was focussed towards stealing banking credentials and other confidential information. The strings consisted a big list of bank websites.

Fig. 14 Banking information present in the form of strings



```
*treasurypathways.com*
*CorporateAccounts*
*weblink.websterbank.com*
*secure7.onlineaccess1.com*
*trz.tranzact.org
*onlineaccess1.com*
*secureport.texascapitalbank.com*
*/Authentication/zbf/k/*
*ebc_ebc1961*
*tdbank.com*
*online.ovcb.com*
*ebanking-services.com*
*schwab.com*
*billmelater.com*
*chase.com*
*bankofamerica.com*
*pnc.com*
*suntrust.com*
*wellsfargo.com*
*ibanking-services.com*
*bankonline.umpquabank.com*
*servlet/teller*
*nsbank.com*
```

7 Conclusion

Memory Analysis technique is an emerging technique widely used in Malware Analysis. It helps to acquire and analyse an infected memory dump. Memory Analysis and Forensic are similar techniques dealing with analysis of a memory file. Our work has described the memory analysis technique by highlighting its importance in Malware Analysis. We have identified and surveyed different tools capable of performing memory acquisition and memory dump analysis.

Memory Malware Analysis follows a 2-step approach- fetching a RAW memory, analysing the memory dump. Our work has included an analysis of different tools capable of performing Memory Malware Analysis. For a better understanding, we have included CRIDEX malware as a part of our case study. Our experimental analysis was focussed towards analysing the memory dump of Cridex infected state and our experimental findings tells that the malware is a trojan. We identified that Cridex was responsible for execution of a bank stealing trojan. The file executable-1640.exe was categorised as a trojan on Virus Total engine and the.dmp file was used to extract strings denoting several banking websites and other legal information related to bank accounts.

References

1. Rathnayaka C, Jamdagni A (2017) An efficient approach for advanced malware analysis using memory forensic technique. [online] IEEE Xplore. Available at: https://ieeexplore.ieee.org/abstract/document/8029568/. [Accessed 19 Jun 2020]
2. Sihwail R, Omar K, Zainol Ariffin K, Al Afghani S (2019) Malware detection approach based on artifacts in memory image and dynamic analysis. Appl Sci 9(18):3680
3. Stüttgen J, Cohen M (2013) Anti-forensic resilient memory acquisition. Digit Investig 10:S105–S115
4. Ravindra Sali V, Khanuja HK (2018) RAM forensics: the analysis and extraction of malicious processes from memory ımage using GUI based memory forensic toolkit. In: 2018 Fourth ınternational conference on computing communication control and automation (ICCUBEA)
5. Guangqi L, Lianhai W, Shuhui Z, Shujiang X, Lei Z (2014). Memory dump and forensic analysis based on virtual machine. [online] IEEE Xplore. Available at: https://ieeexplore.ieee.org/document/6885969?arnumber=6885969. [Accessed 25 Apr 2022].
6. Webopedia (2015) What is cridex? [online] Available at: https://www.webopedia.com/definitions/cridex-malware. [Accessed 25 Apr. 2022].
7. Seo J, Lee S, Shon T (2013) A study on memory dump analysis based on digital forensic tools. Peer-to-Peer Netw Appl 8(4):694–703
8. Manson D, Carlin A, Ramos S, Gyger A, Kaufman M, Treichelt J (2007) Is the open way a better way? Digital forensics using open source tools. In: 2007 40th Annual Hawaii ınternational conference on system sciences (HICSS'07).
9. Okolica J, Peterson G (2010) A compiled memory analysis tool. Adv. Digit Forensics VI:195–204
10. Carvajal L, Varol C, Lei Chen (2013) Tools for collecting volatile data: a survey study. [online] IEEE Xplore. Available at: https://ieeexplore.ieee.org/abstract/document/6557293. [Accessed 16 May 2020].

11. Dolan-Gavitt B (2008) Forensic analysis of the Windows registry in memory. Digit Investig 5:S26–S32
12. Vömel S, Freiling FC (2011) A survey of main memory acquisition and analysis techniques for the windows operating system. Digit Investig 8(1):3–22
13. Kim M, Kim D, Hwang C, Cho S, Han S, Park M (2021) Machine-learning-based android malware family classification using built-in and custom permissions. Appl Sci 11:10244. https://doi.org/10.3390/app112110244
14. Vivekanandam B (2021) Design an adaptive hybrid approach for genetic algorithm to detect effective malware detection in android division. J Ubiquitous Comput Commun Technol 3(2):135–149. https://doi.org/10.36548/jucct.2021.2.006
15. Haoxiang W, Smys S (2020) Secure and optimized cloud-based cyber-physical systems with memory-aware scheduling scheme. J Trends Comput Sci Smart Technol 2:141–147. https://doi.org/10.36548/jtcsst.2020.3.003
16. Or-Meir O, Nissim N, Elovici Y, Rokach L (2019) Dynamic malware analysis in the modern era—a state of the art survey. ACM Comput Surv 52(5):48. Article 88 (Sept 2020). https://doi.org/10.1145/3329786
17. Armstrong B (n.d.) About dump encryption [online] docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/windows-server/virtualization/hyper-v/manage/about-dump-encryption [Accessed 25 Apr 2022]
18. Faiz MN, Prabowo WA (2018) Comparison of Acquisition Software for Digital Forensics Purposes. Kinetik: Game Technol Inf Syst Comput Netw Comput Electron Control [online] 37–44. Available at: http://repository.ittelkom-pwt.ac.id/5808/1/WAP%20Paper%20KINETIK%20Comparison%20of%20Acquisition%20Software.pdf [Accessed 25 Apr 2021]
19. Dayalamurthy D (2013) Forensic memory dump analysis and recovery of the artefacts of using tor bundle browser—the need. [online] Available at: https://core.ac.uk/download/pdf/41535535.pdf [Accessed 25 Apr 2022]

IoT Based Anti Poaching of Trees and Protection of Forest

E. V. Kameswararao, M. Jaya Shankar, T. V. Sai Lokesh, and E. Terence

Abstract Theft of the world's most valuable trees, such as sandalwood, lumber, teak, rosewood, and pinewood, presents a huge danger to forest resources, causes substantial economic harm, and has a terrible impact on the ecosystem across the globe. These trees are very pricey and scarce around the globe. These trees are employed in both medicinal and cosmetics research. To stop such smuggling and conserve the world's forests, various preventative measures must be created. Many incidences of tree cutting occur as a result of the higher amount of money involved in selling such trees. This study presents an anti-poaching system based on IoT and WSN technologies. Three sensors are used in the structural framework: a tilt detector (to detect the tendency of a tree while it is being cut), a fire sensor and a smoke detector (to detect timberland fires), and a sound detector Detection of even the sound of a tree being hacked down may be used to catch illegal loggers, for example WSN technology is commonly employed in remote monitoring applications (where monitoring is difficult).

Keywords Threat · Devasting · Restrict · Preventive measures · Frame work · Unlawful logging

1 Introduction

Timber theft of ecologically and commercially significant tree species in wooded regions—such as Teakwood, coast redwood, Sandalwood, Pine, and Rosewood—has expanded considerably as a result of poaching or smuggling. To address these issues, several players, including the Indian government, have launched a number of initiatives. Anti-poaching observers and/or private/government security guards may

E. V. Kameswararao (✉) · M. Jaya Shankar · T. V. Sai Lokesh · E. Terence
Department of ECE, Hindustan Institute of Technology and Science, Chennai, India
e-mail: lakshmikamesh70325@gmail.com

E. Terence
e-mail: eterence@hindustanuniv.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_4

be recruited, trained, and deployed around woods to help with this. Strict punishment for convicted criminals, as well as unique incentives for smuggling operations (Five Year Plan 2012–2017), were all aimed at putting a stop to the threat. However, the punitive measures have mostly been ineffectual, and poachers have thrived as a result [1–4].

The most effective option is to Build a real-time, wireless sensor network (WSN) and data recording system, a complex and low-cost contemporary technology that will make monitoring more robust, effective, and feasible. The Wireless Sensor Network (WSN) is a novel technology. That is already being utilised in a variety of industrial applications, including monitoring, disaster observation (such as forest fires), living space surveillance, maintenance, security and control, and remote monitoring applications. WSNs are commonly utilised in forest regions for detecting forest fires, detecting tree smuggling, environmental monitoring, and other purposes [5].

Wireless Sensor Networks are simple to set up and maintain, and they save money by eliminating the usage of expensive wires. We can adopt the method that was used to reduce the degree of smuggling in the forest regions with the aid of WSN and other detectors. Poaching isn't only a problem in India; it's also a problem in China, Australia, and certain African countries. In India, sandalwood, rosewood, and pine food tree cost between 12,000 and 13,000 INR per kg, whereas Red Sanders costs INR 10 focus for each ton in the global market. The Indian sandalwood tree has become more interesting as of late, provoking the Indian government to endeavor to restrict sandalwood exports in order to combat the tree's potential hardship. According to the government, the maximum allowable weight gain for an individual is three.8 kg. If the tree is presently under government control, its removal will be banned, regardless of whether it is for personal or sanctuary reasons. The primary purpose of this project is to create a framework that may be used to prevent sandalwood, pine, and teak timber trees from being stolen.

The proposed work is arranged in the following order:

i. Introduction
ii. Related works
iii. Proposed work
iv. Results
v. Conclusion and Future work
vi. References.

2 Related Works

Khandare et al presents the design of a system for detection of vibration and sound for prevention of cutting of trees, detection of temperature and fire for prevention of forest fires. The sensing device can sense the vibration, sound and temperature, fire and then information sent them over zig-bee network to forest office. The fundamental disadvantage of this system is its reliance on batteries for continuous monitoring and data transmission to a central server [6].

Baraddi a technique that may be utilised to prevent smuggling. Three sensors are used in the design system: a sound sensor, a tilt sensor, and a temperature sensor. The Blynk App continually monitors the data collected by these sensors. In terms of sensors, the relay switch turns on the sensors' output devices. A signal is sounded for the slant and sound sensors, and a water siphon is incited for the temperature sensor. The Wi-Fi module saves the created information in Blynk Server. The main drawback of this system is it Can't be communicated to long distance [7].

Narasimman this system is a reliable and low-cost wireless vibration monitoring system. Vibration was measured using a 3-axis digital output MEMS Accelerometer sensor. Vibrations as low as 0.0156 g and as high as 8 g may be detected by the Accelerometer detector, with 1 g equalling 9.81 m/s2. The AT-mega328p microcontroller on the Arduino UNO board is used to connect with the MEMS Accelerometer detector. The primary disadvantage of this technology is that it is susceptible to higher frequency noise [8].

Rohan devised a system that can be used to prevent tree smuggling, thereby forestalling deforestation and keeping up with natural strength. Each tree is furnished with an electronic division, which includes sensors such as the Raspberry Pi, accelerometer sensor, Micro Controller, Fire sensor, Flex Sensor, and GSM/GPRS module. The flex sensor and accelerometer sensor will be used to detect forest tree cutting. GPRS/GSM modules will be used to communicate between the trees and the server. The biggest disadvantage of this method is the price. In comparison to other systems, it is somewhat costly [9].

Mhaske developed a method that can be used to control and restrict tree smuggling, which would eliminate deforestation, reduce wood theft, and maintain environmental stability, all of which would assist with lightening one of the difficulties of a worldwide temperature alteration. A Micro Controller, Flex Detector, Accelerometer Sensor, Temperature Detector, Zigbee and GSM module are just some of the sensors and controllers that are included into each tree. A flex sensor and an accelerometer sensor will be used to detect tree chopping in order to avert it. GSM modules will communicate with the server and the trees. TWO phases comprise the system: 1. The tree unit The main server unit is referred to as B. The primary disadvantage of this technology is that it is susceptible to higher frequency noise [10].

Raghavendra et al the fundamental goal of this method is to reduce and prevent smuggling, conserve precious trees, and minimise wood theft in request to keep a decent eco-framework by diminishing deforestation. The technique use GPS technology to pinpoint the position of the tree where poaching occurs. The system uses a chip (micro controller board) with embedded sensors (flex sensor and fire sensor) that are controlled via IoT. These sensors monitor and control parameters such as tree tilting, burning, and cutting, and they can be accessed via an Android App on an Android phone. The biggest disadvantage of this technique is that it's tough to keep track of the forest [11].

Arunprasath et al author created a methodology that may be used to reduce sneaking. Detecting illegal logging, for example, involves listening for noises generated when cutting down a tree. Other uses for the temperature and tilt sensors include detecting forest fires (to recognise the tendency of tree when its being cut). The web page/app continuously monitors the data generated by these sensors. Sensors' yield devices are activated by a hand-off switch, which is a relay switch. A ringer is established for the inclination and sound sensors, while a water siphon is impelled for the temperature sensor. The Wi-Fi module is utilized to save the information made in the cloud server. The fundamental disadvantage of this system is its reliance on batteries for continuous monitoring and data transmission to the central server [12].

Sudharani et al to limit sneaking and screen trees, author structure can be made using gyro locater (to perceive the inclination of tree when it is being cut), temperature pointer (NTC 10 k thermistor), Wi-Fi Module (esp8266), and GSM Module. The data accumulated by these sensors is ceaselessly noticed using the ThingSpeak cloud stage, which is sent from a microcontroller (Arduino Uno). The data is saved in ThingSpeak Server through the Wi-Fi module. The main disadvantage of this system is that it is inflexible. Long-distance communication is not possible [13].

Hingane et al author designed a system using some sensors like flex sensor, pH Sensor, Fire Sensor, Ultrasonic Sensor, Node MCU(microprocessor) Relay, Water pump, Buzzer to monitoring the forest continuously for avoid the smuggling and save the forest from forest fire when forest is in fire. The main drawback of this system is fetched (cost). It is very expensive as compared to other systems [14].

Kirangond suggested an anti-poaching system that includes an archetype of an IoT model that monitors trees and alerts the base station in [15]. When the tilt values change, the base station and the registered cell phone number get the coordinates of the poached tree. In comparison to the current plans, the suggested model reduces human involvement. GSM module, Raspberry Pi board, and accelerometer sensor GPS module are included in the proposed model, which will aid in tree monitoring. This document presents the proposed model's working theory, which will aid in the improvement of the current system while also acting as a resiliency for the proposed system. Small places, such as private property, and vast areas, such as national reserves, may both benefit from the suggested approach. The system's biggest flaw is that it's far away. When compared to other systems, it is rather pricey [16].

3 Proposed Work

The proposed system will be made up of two modules: one with sensors and controllers that will be installed in the tree, and the other with an Android phone or computer. We created an application that would receive sensor data on a constant basis. This is an IoT-based project in which sensor data is regularly uploaded to the biodots cloud. Tilt sensor and sound sensor is used to determine whether the tree is cut down or not. Similarly, fire/flame sensor and smoke sensor is used to determine whether the forest is on fire or not. The sensed data continuously uploaded in the biodots cloud for every 30 s. The block diagram of the proposed architecture is shown in Fig. 1.

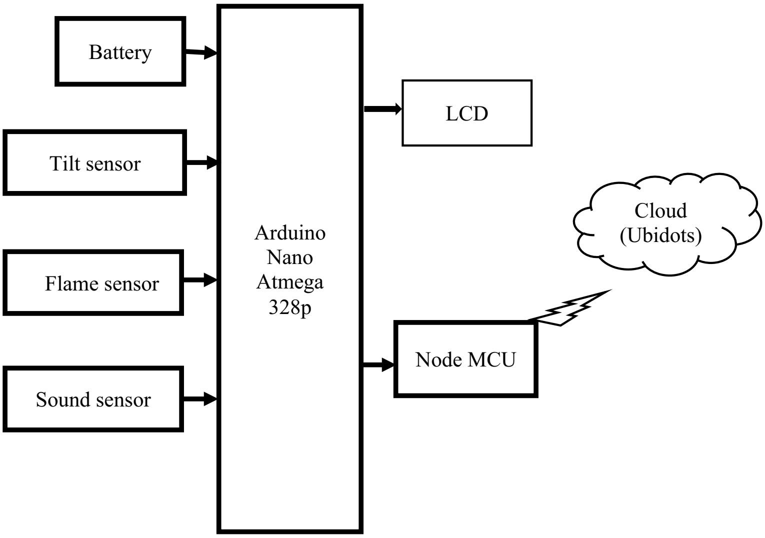


Fig. 1 Block diagram

The major objective of this project is to construct a portable wireless sensor node that may be utilised in a Wireless Sensor Network. An Android phone or computer with sensors and controls will be embedded in a tree as part of a planned two-modular system.

We created an application that would receive sensor data on a constant basis. This is an IoT project in which we regularly upload sensor data to the ubidots cloud. Tilt sensor and sound sensor is used to determine whether the tree is cut down or not similarly Fire/Flame sensor and smoke sensor is used to determine whether the forest is on fire or not. The sensed data continuously uploaded in the ubidots cloud for every 30 s. We can able to know the actual GPS location of the sensors. When a sound is detected, the sound sensor generates an output voltage signal. A microcontroller receives voltage and begins doing the required processing. The process flow diagram is shown in Fig. 2.

Noise levels in decibels (dBs) may be measured by the sound sensor at frequencies between 3 and 6 kHz, which is about where the human ear has sensitivity. It reacts to sound intensity in the same manner that the human ear does. It accurately monitors sound levels across a single range of 55–110 dB to within 3 dB. There are noise levels in the forest owing to trees and animals, so if a tree falls, the noise level is abnormal, which is reported to the forest officer, and there is a distinction between tree fall down due to natural disasters and tree fall down due to unlawful methods.

Table 1 shows the threshold value of tilt and smoke sensor and Table 2 on flame and sound sensor.

A smoke sensorr is a device that senses smoke (air particles of different gases or fire), typically as an indicator of fire. Smoke sensor do not have a listed spacing. They have a recommended spacing of 30 feet between sensors. However, smoke sensor

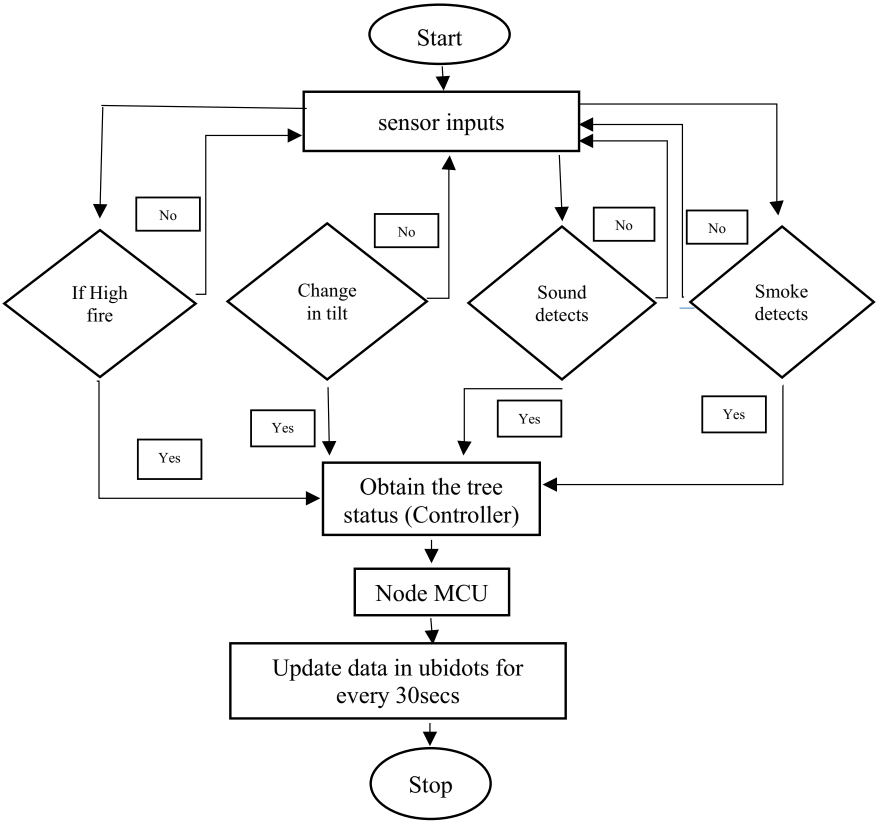


Fig. 2 Flow chart

Table 1 Threshold values of tilt and smoke sensor

Sensor name	Threshold value
Tilt sensor	430(When it is NOT tilted)
Smoke sensor	212

Table 2 Threshold values of flame and sound sensor

Sensor name	High	Low
Flame sensor	1 (Presence of fire)	0
Sound sensor	1 (Presence of sound)	0

can be installed up to 41 feet. A fire locator is a sensor that recognizes and answers the presence of a fire or fire, making fire identification conceivable. They operate at moderate speed with a range of up to 200–250 feet from the flame source.

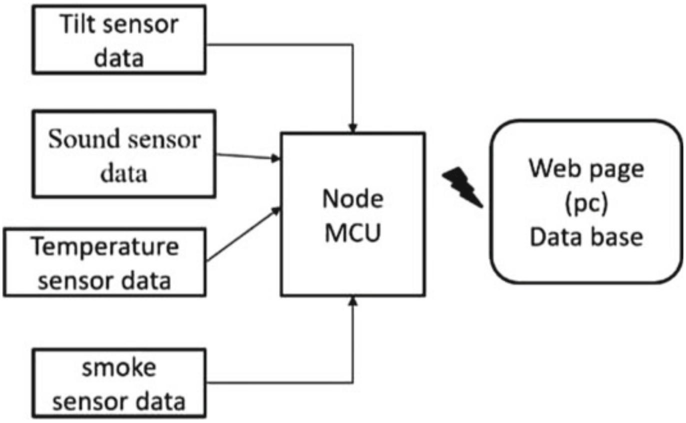


Fig. 3 Node MCU architecture

Initially, when it is NOT tilted position, it shows the fixed threshold value 430 in the LCD as well as in the web pageWhen it is inclined in a particular direction or angle, it shows value of tilting position of a tilt sensor on the LCD as well as in the web page. If we tilt the sensor in the left side direction the values are decreased. Similarly, if we tilt that sensor in right side direction the values are increased. It can detect the values of tilting position of a tree in between 20 and 60°. Node MCU architecture is shown in Fig. 3.

The sensed data, sends to controller. After receiving the data, the controller starts processing and perform the necessary operations in it. The Wi-Fi module receives data as well. The controller sends data to the Wi-Fi module, which subsequently sends it to the cloud through the internet (Ubidots). We can monitor the data in Ubidots cloud.

4 Results

Figures 4 and 5 shows the only either Left—Right side direction or Front—Back side direction. The tilt sensor works from 0 to −70° as well as from 0 to 70°. The remaining angle position −70 to 180 and 70–180° is uncertain region. If the tilt sensor is NOT tilted position (i.e., at 0°), it shows the actual threshold value 430 in the LCD and as well as in the Web page. If the sensor tilted left (i.e., it bends to −20-−70°), it shows the value of decreasing threshold and varies according to bending angle and if it is tilted right (i.e., 20–70°), it shows the value of increasing threshold and varies according to bending value. It is shown in the Web page. If the angle between −20 and −70 and 20–70°, the tree is tilted heavily and we can conclude that it indicates like there is something illegal happening. Similarly in front and back directions also same. The front direction indicates the right-side direction and backside direction indicates the left side direction. The results are getting from right and front directions are same, similarly in left and backside directions are same. Table 3 shows the practical values of tilt sensor.

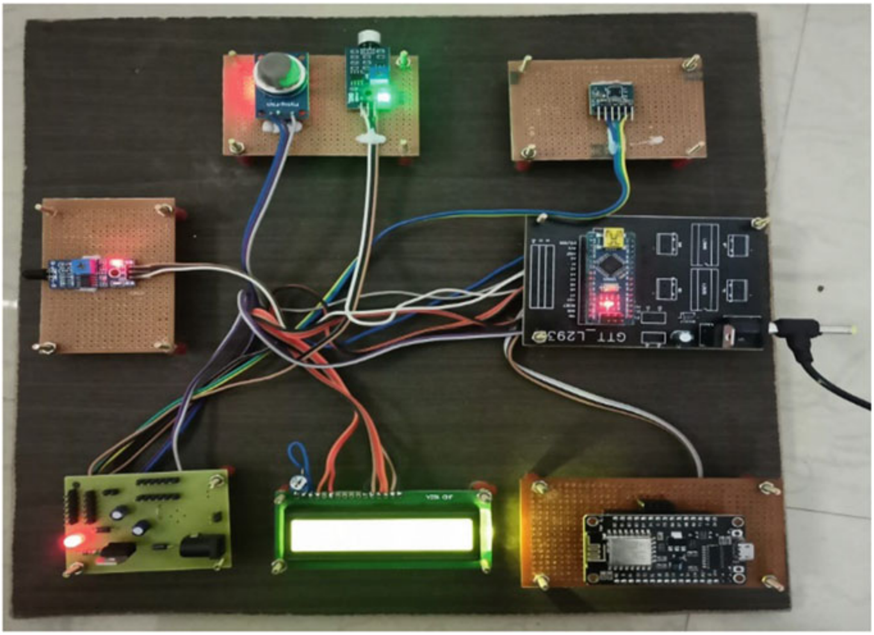


Fig. 4 Result

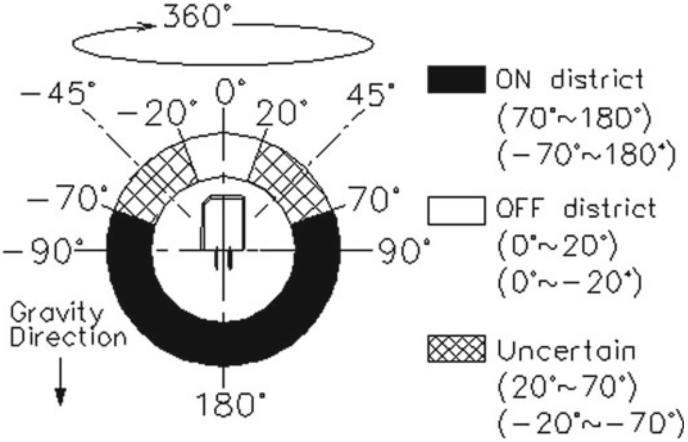


Fig. 5 Working position of tilt sensor

Table 3 Data table of tilt sensor

Direction	Actual angle	Practical value
Left (or) Back	0° (When it is NOT tilted)	430
Right (or) Front	−45°	362
	−60°	315
	0° (When it is NOT tilted)	430
	45°	488
	60°	510

Table 4 Data table of sound sensor

Sensor Name	Practical values	Condition
Sound sensor	0	Low
	1	High
	1	High
	0	Low

If sound sensor detects sound, it shows high value '1'.If there is no sound, it shows the value '0'on the Web page. Table 4 shows the practical values of sound sensor.

If flame sensor detects any fire near by its, it shows the 'High' value '1'. If there is no fire it shows 'Low' value '0'.in the LCD as well as in web page. Table 5 shows the practical values of flame sensor.

The smoke sensor threshold value is 212. If the smoke sensor value shows below threshold value, there is no much amount of smoke and less impact of fire. If the value is reaches above threshold value, there is a high chances of forest fire. Table 6 shows the output values of smoke sensor.

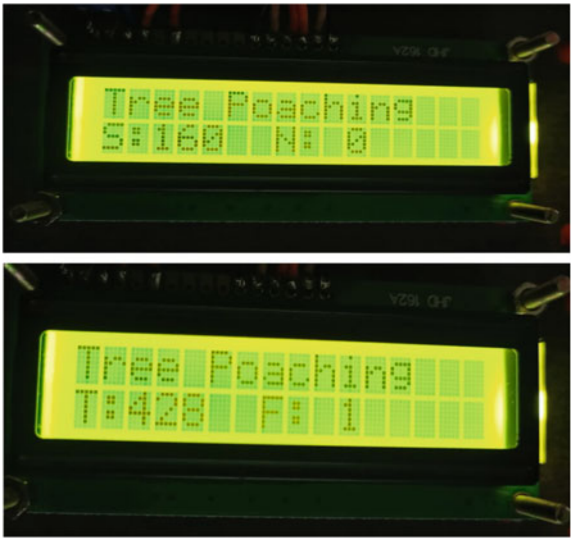
Figure 6 shows the LCD output.

Table 5 Data table of flame sensor

Sensor name	Practical values	Condition
Flame sensor	1	High
	1	High
	1	High
	1	High

Table 6 data table of smoke sensor

Sensor name	Threshold value	Practical value
Smoke sensor	212	216
		228
		237
		236



```
Tree Poaching
S:160  N: 0
```

```
Tree Poaching
T:428  F: 1
```

Fig. 6 LCD (output display)

5 Conclusion and Future Works

5.1 Conclusion

A portable wireless sensor network has been developed for save the forest from forest fire, deforestation and some important trees like Sandalwood, Teakwood, Pine and Rosewood etc. The motive of this design for keeping trespassers away and save the forest from timber mafia. For this reason, Different sensors like Tilt, Sound, Smoke, Fire sensors and Arduino Nano board, LCD, Wi-Fi module is used. To avoid the Trees cutting, Tilt and Sound sensor is used. To save the forest from forest fire, fire/flame and Smoke sensor is used. For remote terminal through wireless media, Wi-Fi module is used. We can monitor the data on LCD as well as in the Ubidots cloud.

5.2 Future Works

This study may be improved in the future by using a multi-node network with microphones, motion detector sensors, data collection systems may be improved by using temperature and human or animal interference sensors as well as sensors to monitor fires.

References

1. Sudha BS (2018) Forest monitoring system using wireless sensor network. Int J Adv Sci Res Eng 4(4), E-ISSN: 2454–8006
2. Dalvi A (2018) Undetected detective to protect the forest trees against poaching using WSN technology, vol 3, no 6
3. Gaikwad S (2015) Design WSN Node for Protection of Forest Trees Against Poaching based on Zigbee. In: 2015 IEEE international conference on electronics, computing and communication technologies (CONECCT)
4. Ghousia Sultana B (2018) IOT based anti-poaching alarm system for trees in forest using wireless sensor network. Int J Adv Res Comput Sci 9(3)
5. Hamza HC (2013) In: Tree theft control system 2013 texas instruments India educators conference.
6. Khandare A, Malve M, Kulkarni A (2014) Zigbee based wireless sensor network (WSN) for protection high cost trees in from fire and poaching. IRJET
7. Baraddi P, Hanchinal N, Jadhav R, Shushma, Banni R (2020) IOT Based Anti-Poaching Alarm System for trees in forest. IJERT. https://www.ijitee.org/wp-content/uploads/papers/v8i6s/F60510486S19.pdf
8. Narasimman V, Anand, Anandha kumar, Krishnan T (2018) Design of a wsn node for forest trees against poaching. IJARSE. http://www.ijarse.com/images/fullpdf/1521184541_DACE2062ijarse.pdf
9. Rohan (2018) Real time forest anti-smuggling monitoring system based on IOT using GPRS and GSM. Int J Res Eng Appl Manage (IJREAM) (ICSGUPSTM 2018). ISSN: 2454-9150
10. Mhaske D (2016) Anti-smuggling system for trees in forest using flex sensor with GSM & Zigbee network. Int J Adv Res Comput Commun Eng 5(4)
11. Raghavendra (2019) IoT based anti-smuggling system for trees in forest. Int J Sci Res Rev 8(6)
12. Arunprasath, Naveenraj MT, Srinivasan R, Jeevabarathi C (2019) IoT based anti-poaching alarm system for trees in forest. (IJITEE) 8(6S), ISSN: 2278-3075
13. Sudharani CH, Shilpa (2019) IoT based anti-poaching of trees. Int J Eng Adv Technol (IJEAT) 8(5), ISSN: 2249-8958
14. M.C.Hingane, Snehal Choudhari, Vandana Datta Ingale, Sonali Awachare, "Anti-poaching Alarm System For Tree in Forest", IJEDR 2019 | Volume 7, Issue 4 | ISSN: 2321–9939.
15. Kirangond, Mudanuri Chaithanya K, Shellagi PR, Sampath Kumar NG, Nivedha S (2020) Anti poaching of trees using Raspberry Pi. IJESC 10(6), ISSN 2321-3361 © 2020
16. Narhari R, Kotkar ME (2014) (ESD AND VLSI) Anti-smuggling system for trees in forest using flex sensor and Zigbee. Int J Adv Res Comput Eng Technol (IJARCET) 3(9)

Artificial Intelligence Based Efficient Activity Recognition with Real Time Implementation for ATM Security

S. Srinivasan, AL. Vallikannu, Adapa Sankar Ganesh, Iragamreddy Raj Kumar, and Beereddy Venu Gopal

Abstract Recognizing human activities plays a substantial role in human-to-human and human-to-computer interactions. Recognizing human activities from video sequences or pictures is a difficult task because of troubles, such as history clutter, partial occlusion, modifications in scale, viewpoint, lights and look. Human action is difficult to classify as a time series. Predicting a person's movements is a part of this. In this paper, the KTH video dataset is used for designing the system. Feature extraction methods like optical flow and spatiotemporal techniques are being utilized to extract the features. Triple stacked autoencoders are used for clusterization to reduce the data dimensions. An efficient BoW vector feature extraction method is used for extracting text data, by which data is obtained for training the model. A deep learning algorithm such as VGG19 is used to determine and classify the activities of a human. The objective of this efficient model is to apply as an ATM surveillance as a camera module fixed in the room to perform constant surveillance. The Police department can have an mobile application through which they can monitor and desist any unwanted human activities happening in the ATM.

Keywords Human activity recognition · ATM security · VGG19 deep learning · Sequence processing · BoW vector feature extraction · KTH video dataset · ATM app

S. Srinivasan (✉) · AL. Vallikannu
Department of ECE, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India
e-mail: ssrinivasan@hindustanuniv.ac.in

AL. Vallikannu
e-mail: vallikannu@hindustanuniv.ac.in

A. S. Ganesh · I. R. Kumar · B. V. Gopal
Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India
e-mail: 18121154@student.hindustanuniv.ac.in

I. R. Kumar
e-mail: 18121166@student.hindustanuniv.ac.in

B. V. Gopal
e-mail: 18121183@student.hindustanuniv.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_5

1 Introduction

Human Activity Recognition is a broad subject of study concerned with identifying the precise motion or movement of a person based on sensor information. Activities are regularly done standard movements which include walking, standing, sleeping, and sitting. They may also be extra targeted on other activities such as the varieties of activities executed in a kitchen or in a workplace.

The sensor data can be remotely recorded, inclusive via video, radar, or other wireless methods. Alternatively, information can be recorded straightforwardly regarding the matter, for example, by carrying custom hardware or smartphones that have accelerometers and gyroscopes. Fitness and health tracking technologies such as cell phones are now widely available and affordable. As a result, the cost of acquiring data from these devices is lower, not unusual, and consequently is a more generally studied version of the general activity recognition hassle.

It is the goal of human activity recognition software to analyse video or still images to determine what people were doing at a given time. Human activity recognition systems are impacted by this fact and work to accurately categorise input data into the appropriate activity category. To classify human activities, terms like "gestures," "atomic motions," "human-to-object," "group movements," "behaviours," and "events" are used, all of which are based on the degree of complexity they involve. When it comes to gestures, they are considered to be primordial bodily components that may be linked to a certain movement of a person. It is possible to describe an activity in terms of its "atoms," which are the parts that make up the whole. Human-to-human or human-to-item interactions include people interacting with other people or things. Group activities are those that are carried out by a group of people. A person's emotional, personality, and psychological condition may be communicated via their body movements. The ultimate definition of events is that they are excessively staged actions that represent social movements among persons and imply the aim or social standing of those involved.

2 Literature Review

A Double Stacked Autoencoder with implanted grouping (DSAFEC) and a BOW building strategy in view of the DSAFEC to limit computational intricacy and kill determination limitations (B-DSAFEC) has been offered in [1]. The DSAFEC used video feature points to generate BOWs for human activity identification by predicting the probability of feature point cluster assignment. Soft clustering instead of hard clustering gives each feature point to a number of clusters with the greatest probability, rather than assigning each feature point to only one cluster. Trial discoveries on three benchmark human action datasets showed that the B-DSAFEC outflanks five different procedures created utilizing either ordinary bunching techniques or profound grouping strategies.

In [2], PrivHome, a privacy-preserving solution, was proposed. For smart home systems, it offers authentication, safe data storage, and inquiry. For security reasons, PrivHome ensures information classification as well as element and information confirmation for all gatherings associated with the information transmission. Additionally, the gateway offers privacy-preserving queries so that neither the service provider nor the gateway can access the data. The idea of protection safeguarding requests for savvy home frameworks has never been raised, essentially not as far as anyone is concerned. Entity and key sharing protocols as well as a searchable encryption mechanism have been added to the system, were included in the method. Because both protocols rely entirely on symmetric cryptographic techniques, the scheme was practicable. There are comparisons to current smart home security methods based on data collected via experiments and modelling, to demonstrate the efficiency and effectiveness of the method.

Using FMCW radar, a novel Dynamic Range-Doppler Trajectory (DRDT) approach was proposed in [3], for recognising continuous human movements in real time in a variety of situations simulating real-life conditions. Continuous motions can be separated and processed as single events using this method. The backscattered signals were first used to create range-Doppler frames, in which each map is composed of a progression of reach Doppler maps. For constant following of human developments in time and reach and Doppler areas, the DRDT was constructed from these frames. The DRDT map was then used to find and segregate each individual human motion using a peak search approach. Final results were obtained by extracting and integrating range, Doppler, Radar Cross-Section (RCS), dispersion characteristics into a machine learning classifier using a multidomain fusion technique. Even when the distance, angle of view, and direction of the view, as well as the diversity of individuals, were all altered, this method achieved accurate and robust recognition. Extensive testing has been carried out to demonstrate its practicality and superiority. Continuous categorization yielded an average accuracy of 91.9%.

In order to avoid identification of source cameras based on PRNU, the research [4] suggested a number of solutions. Forced seam carving is an example of a new counter forensics approach. By eliminating seams to the point that most uncarved picture blocks are less than 50 × 50 pixels, the method eliminateed the need for PRNU-based identification of the source camera. Even if the image's uncarved blocks are smaller than the 50 × 50 pixel minimum, the results of this research reveal, source attribution can still be done given numerous seams carved photographs from the same camera.

The community broker function was employed in study [5] to integrate community services, providing electronic information services, reducing the burden of community managers, and enhancing connections between the community and its surrounding environment. Integrating various sensors and gadgets in the house to provide energy management, scenario information, and security are all parts of a house intranet that were developed using a decent touch board and a home regulatory framework. Local area PCs, as well as connectivity to other communities and home networks, are part of the end-to-end community solution (e.g., video cameras and building automation equipment). With the usage of standard interface devices, numerous in-home displays may be achieved by separating the logic and user interfaces.

A Light Field Camera as a fresh perspective for detecting face show assaults was described in [6]. A light field camera's interesting ability to make different profundity (or concentration) pictures in a solitary catch is due to its capability to record both the direction and intensity of each incoming ray. A new method for detecting presentation assaults was therefore described, which examined focus fluctuations across multiple LFC-generated depth (or focus) pictures. Using a light field camera, first a fresh face artefact database of 80 individuals was gathered. Facial artefacts were made by mimicking two typical assaults: picture printing and electronic screen attacks. Tests on the light field artefact database have shown that the suggested PAD technique outperforms many well-known state-of-the art strategies.

A recurrent neural network approach is used to analyse the dicriminative training of data analytics was proposed in [7] which performed well in activitity recognition. A most compact and efficient model with feature descriptors and classifiers proposed by Cozar et al. [8] produced a comparatively good results on activity recognition. A deep learning clustering metric learning method was suggested by Law et al. [9] and Dizaji et al. [10]. They obtained the gradient loss function which involves less complexity for training a model. The method learnt future representations and embeddings required for clusters. Detecting human action from video datasets are very complex and the same was addressed in [11] providing spatio-temporal localization. Activities of learners in an online class is monitored by a deep learning method proposed by Vivekanandam [12] in which the emotional behaviors of human is recognised with satisfied computational speed. A Neural Network based suspicious human activity detection system was proposed in [13] where they used temporal and qualitative knowledge in detection of anomaly of a human [14].

To summarize, the activity recognition of human has been discussed and analysed for various reasons like security and detection of any anomaly etc. in many literature using Neural Network and Machine Learning algorithm. Many of the works are lagging in identifying the recognition of human activity interms of accuracy. Also many works used dual stacked auto encoders which for video dataset is insufficient to extract the data.

3 Proposed Model

In this paper, the human activity recognition is determined by which the user's activity in an ATM can be effectively identified by which an abnormal action can be determined. So, the first stage is to gather data from different sources, and then divide this data into preparing and testing datasets, where the preparation dataset is kept discrete and the testing dataset is used to prepare the model. After then, the dataset is supplemented with new data to make it larger. When these datasets have been analysed, they are aligned using various procedures. In order to create the system, the KTH video dataset is employed. After preprocessing the dataset, architectural training can be performed. Feature extraction methods like optical flow and spatiotemporal techniques are being utilized to extract the features. Triple layered auto encoders are used to minimise the size of the data. Using efficient BoW vector feature extraction, data for training the model are gathered from text data. An algorithm like VGG19 is then used to identify and categorise a person's activities. As a camera module is permanently installed in the room, this effective model can be used in ATM monitoring. In the event that the ATM camera detects an abnormality in the human's activities, an alarm is raised to the police. When the ATM camera detects anomalous behaviours of a person attempting to break the ATM's security, an alarm is delivered and live streaming is enabled via a mobile application made using react native. Proposed System Architecture is shown in Fig. 1.

In this research work, VGG19 is used for classifying the activities from obtaining the features as it shows accuracy close to 92%. VGG19 model has faster training speed, fewer training samples per time, and higher accuracy. VGG is a deep Convolutional Neural Network used to classify the images. VGG19 is a variation of the VGG model that has 19 layers in total (16 convolution layers, 3 Fully connected layer, 5 MaxPool layers and 1 SoftMax layer). VGG11, VGG16, and other variations are all forms of VGG. VGG19 has a total of 19.6 billion FLOPs. To train this network, an RGB picture with a fixed size of (224 * 224) is fed, and therefore the matrix is shaped as (224, 224, 3). For preprocessing, simply the mean RGB worth of every

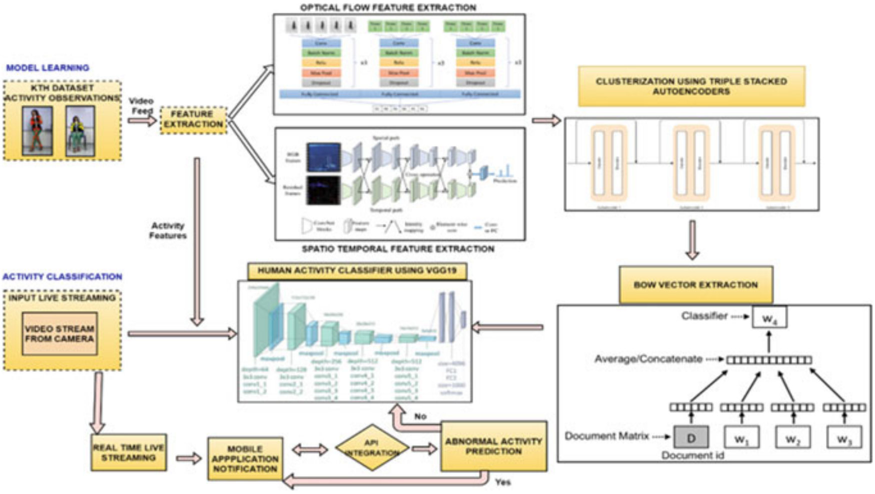


Fig. 1 Proposed system architecture

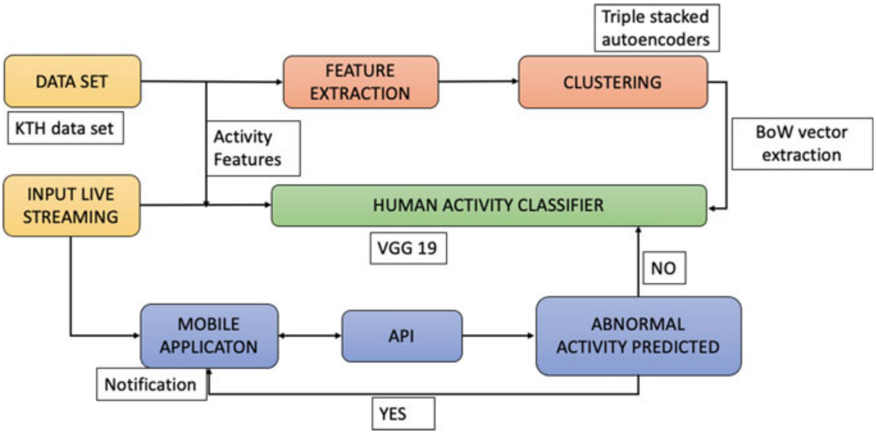


Fig. 2 Algorithm to classify the activity of a human in an ATM

pixel, determined across the entire preparation set, is disposed off. (3 × 3) measured bits with a step size of 1 pixel is utilized, permitting them to traverse the entire picture idea. The picture's spatial goal is safeguarded by the utilization of spatial cushioning. Sride 2 is utilized to create greatest pooling on a 2 by 2 pixel window. In order to increase classification accuracy and speed up computations, the model is further refined by using the Rectified Linear Unit (ReLu), a non-linear function that introduced non-linearity into the model. To classify 1000-way ILSVRCs using 1000 channels, a softmax function is used in the last layer of three fully linked layers, the first two of which are 4096 bytes long. The algorithm for identifying the human activity in this work is given in Fig. 2.

React native is being utilised to create a mobile app for this work. React Native is a framework for creating JavaScript code from a hierarchy of UI components. Using this framework, a native-looking mobile app can be created for both iOS and Android. The mobile industry has grown at a rate that has never been seen before. By the year 2020, mobile apps are expected to produce income of over 188 billion dollars in the United States via app stores, advertising, and in-app purchases. A high-quality app is needed for both personal and corporate usage, with various displays, simple navigation, and a decent design. On the other hand, developing high-quality, high-performing native applications takes a long time compared to developing cross-platform apps, which is quicker but sacrifices quality and support. If you're looking to develop a high-quality app in a short period of time, React Native seems to be a suitable alternative.

The development environment for mobile apps in this paper is Android Studio. It depends on JetBrain's IntelliJ IDEA programming and is custom-made solely for Android improvement. Android Studio is the authority IDE for Google's Android working framework. In 2020, it will be a free download or a membership based assistance for clients of Windows, macOS, and Linux working frameworks. As the major IDE for native Android application development, it replaces Eclipse Android Development Tools (E-ADT).

4 Results and Dıscussıon

The dataset collection involves the process of collecting normal human activity that include walking, posing etc. which a human performs inside an ATM room as shown in Fig. 3i, ii. Apart from the regular activity listed above, other activities like running, jogging and hand waving are also collected for training dataset and performed several times by 25 subjects in four different scenarios: outdoors, outdoors with scale variation, outdoors with different clothes and indoors. The database contains 2391 sequences. All sequences were taken over homogeneous backgrounds with a static camera with 25fps frame rate. The sequences are downsampled to the spatial resolution of 160 × 120 pixels and have a length of four seconds in average. The validation matrics the human activity is shown in Table. 1.

Then these datasets are pre-processed by converting the images into required size format so that it can be made ready for training with the model. Training is performed using CNN algorithm. After training the stacked autoencoder, an accuracy of approximately 99% is achieved. This means that the stacked autoencoders can recreate the original input signal with about 99% accuracy. The decrease in loss during the training and the increse in accuracy during training are depicted in Figs. 4 and 5 respectively.

From the above results, it can be seen that the triple stack autoencoder algorithm has been successfully implemented to automatically recognize the human activity. In the existing system, the system uses dual stacked auto encoders which for video dataset is insufficient to extract the data and the system does not provide an effective technological solution for real time deployment. In order to overcome this problem, a camera module that is permanently installed in the ATM room provides a cost-effective solution for ATM security. It's able to identify a normal human activity as shown in Fig. 6 inside an ATM room, and in the event that the ATM video detects odd behaviour from a person attempting to get into the machine as shown in Fig. 7, an alarm is forwarded to authorities. Hence, the purpose of this project is to develop and deploy an action recognition and gesture recognition system that can detect human activities automatically. Figure 6 explains the application of such human normal activity in an ATM room (drawing money by standing in front of the machine) to identify what is happening inside and it helps to improve the security.

In Fig. 7, when a person tries to perform an unwanted activity like break open the ATM machine etc., it is identified by the algorithm, the video is captured and security alert is sent to nearby Police station to safeguard the ATM. Through live streming, a Police staff can view in the ATM app, which is shown in Fig. 8.

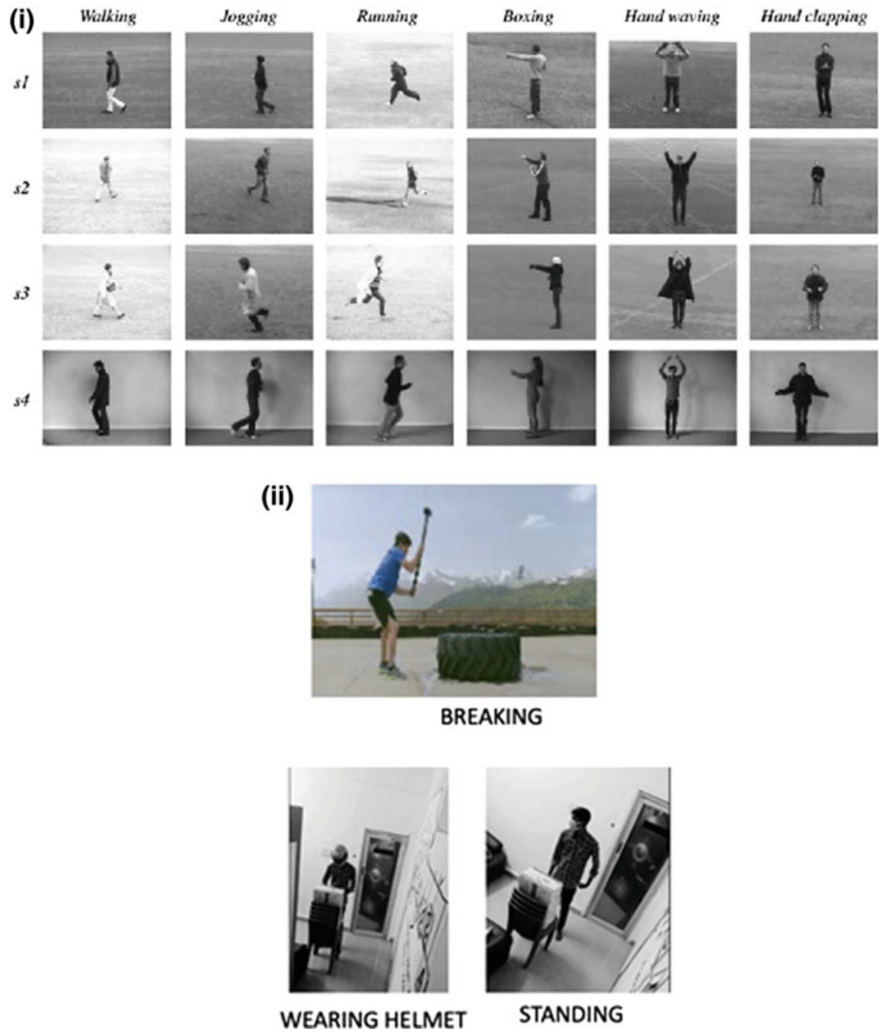


Fig. 3 **i** Dataset collected (1), **ii** dataset collected (2)

5 Conclusion

The Human Activity Recognition model has been created and successfully implemented. It recognizes the human activity in the ATM room and sends the notification to the police if any abnormal activity is detected. The police are provided with an application called ATM app, in which the police is able to watch the live streaming of what is happening inside the ATM. Spatio temporal and optical flow features are

Table 1 Validation matrics

Activity	Precision	Recall	F1-score	Support
Walking	1.00	1.00	1.00	70
Boxing	1.00	1.00	1.00	37
Hand clapping	1.00	1.00	1.00	41
Hand waving	1.00	1.00	1.00	55
Standing	1.00	0.99	0.99	207
Breaking	0.93	0.96	0.95	27
Accuracy			0.99	437
Macro avg	0.99	0.99	0.99	437
Weighted avg	0.99	0.99	0.99	437

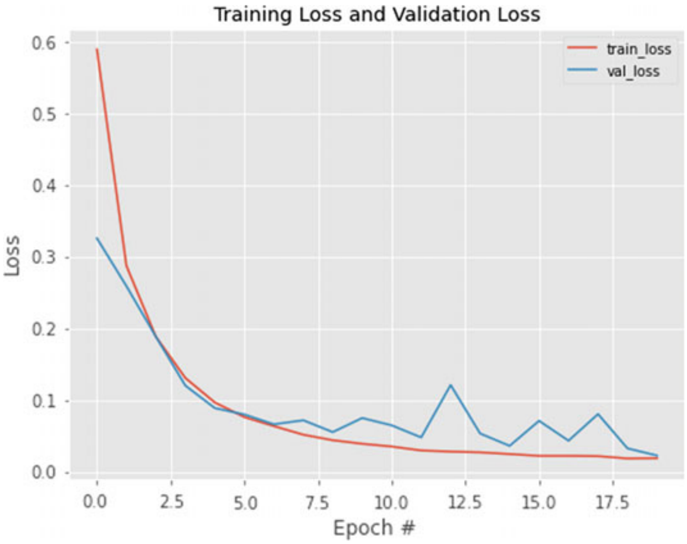


Fig. 4 Decrease in loss during training

extracted. Triple stacked auto encoders are used for clusterization. Deep Learning VGG19 algorithm is used for training the model. Thus the recognition is done. Results are examined and classified as normal and abnormal activity, which may be used in an ATM for real time surveillance.

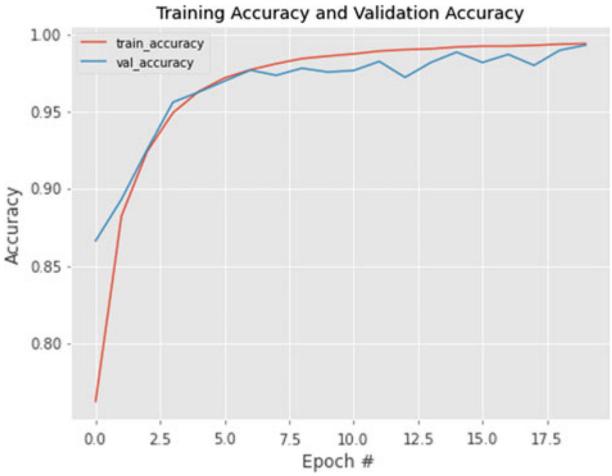
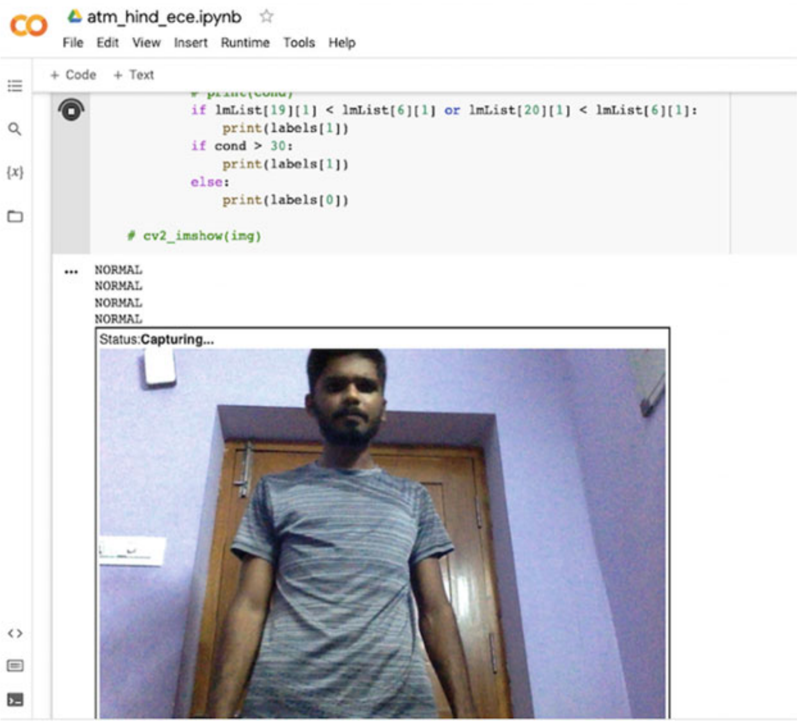


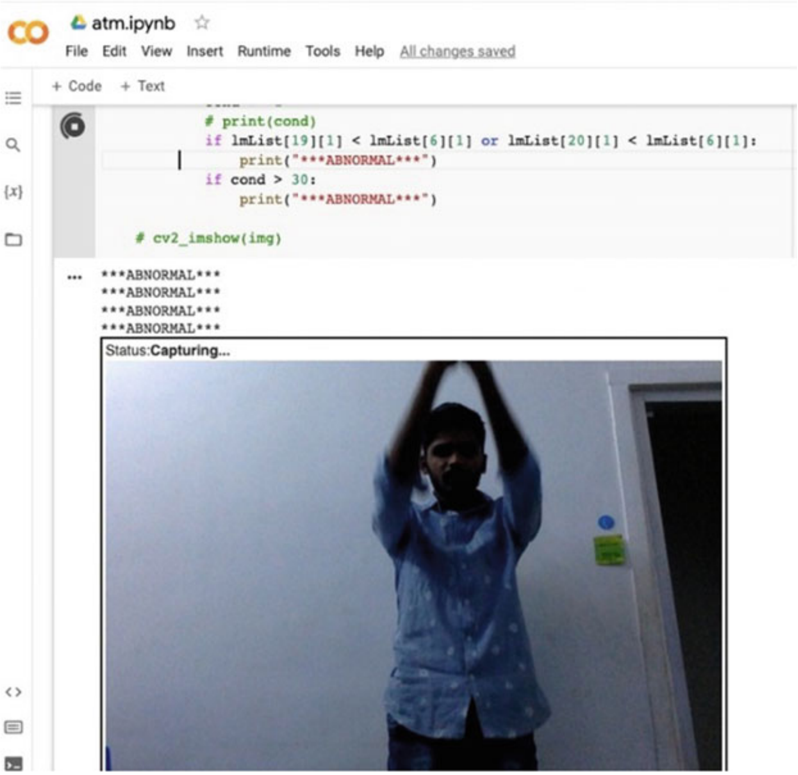
Fig. 5 Increase in accuracy during training



```
        # print(cond)
        if lmList[19][1] < lmList[6][1] or lmList[20][1] < lmList[6][1]:
            print(labels[1])
        if cond > 30:
            print(labels[1])
        else:
            print(labels[0])

  # cv2_imshow(img)
```

```
NORMAL
NORMAL
NORMAL
NORMAL
```

Fig. 6 Normal activity



```
    # print(cond)
    if lmList[19][1] < lmList[6][1] or lmList[20][1] < lmList[6][1]:
      print("***ABNORMAL***")
    if cond > 30:
      print("***ABNORMAL***")

# cv2_imshow(img)
```

```
***ABNORMAL***
***ABNORMAL***
***ABNORMAL***
***ABNORMAL***
```

Fig. 7 Abnormal activity

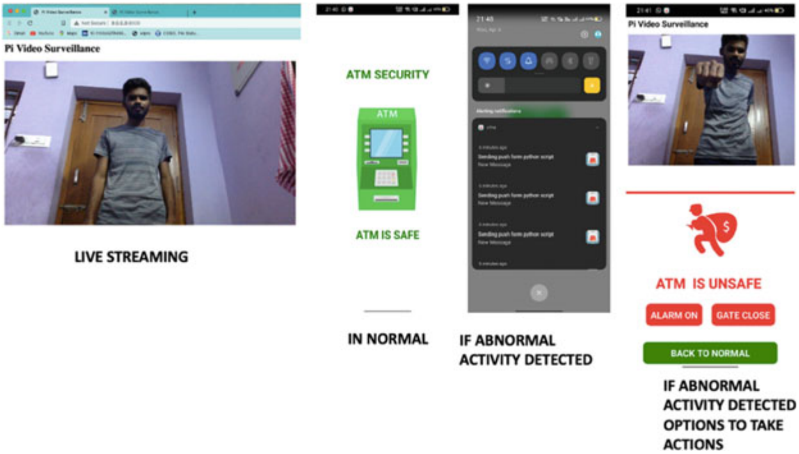


Fig. 8 Live streaming of ATM app

References

1. Wang T, Ng WWY, Li J, Wu Q, Zhang S, Nugent C, Shewell C (2021) A deep clustering via automatic feature embedded learning for human activity recognition. J Latex Class Files. IEEE
2. Poh GS, Gope P, Ning J (2019) PrivHome: privacy-preserving authenticated communication in smart home environment. IEEE Trans Dependable Secure Comput 18(3):1095–1107
3. Ding C, Hong H, Zou Y, Chu H, Zhu X, Fioranelli F, Le Kernec J, Li C (2019) Continuous human motion recognition with a dynamic range-doppler trajectory method based on FMCW radar. IEEE Trans Geosci Remote Sens 57(9):6821–6831
4. Taspinar S, Mohanty M, Memon N (2017) PRNU-based camera attribution from multiple seam-carved images. IEEE Trans Inf Forensics and Secur 12(12):3065–3080
5. Lee YT, Hsiao WH, Huang CM, Seng-Cho TC (2016) An integrated cloud-based smart home management system with community hierarchy. IEEE Trans Consum Electron 62
6. Raghavendra R, Raja KB, Busch C (2015) Presentation attack detection for face recognition using light field camera. IEEE Trans Image Process. Norwegian Biometric Laboratory, Gjøvik University College, Norway
7. Richard A, Gall J (2017) A bag-of-words equivalent recurrent neural network for action recognition. Comput Vis Image Underst 156:79–91
8. Cozar JR, González-Linares JM, Guil N, Hernández R, Heredia Y (2012) Visual words selection for human action classification. In: 2012 International conference on high performance computing simulation (HPCS), pp 188–194
9. Law MT, Urtasun R, Zemel RS (2017) Deep spectral clustering learning. In: Proceedings of the 34th international conference on machine learning, vol 70, pp 1985–1994
10. Dizaji KG, Herandi A, Deng C, Cai W, Huang H (2017) Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization. In: IEEE international conference on computer vision, pp 5747–5756
11. Shao L, Jones S, Li X (2014) Efficient search and localization of human actions in video databases. IEEE Trans Circuits Syst Video Technol 24(3):504–512
12. Vivekanandam B (2020) Evaluation of activity monitoring algorithm based on smart approaches. J Electron 2(03):175–181
13. Bhambri P, Bagga S, Priya D, Singh H, Singh H, Dhiman HK (2020) Suspicious human activity detection system. J IoT Soc Mob Anal Cloud 2(4)
14. Herath S, Harandi M, Porikli F (2016) Going deeper into action recognition: survey. Image Vis Comput 60:4–21

Terror Attack Classification with the Application of Orange Data Mining Tool and Neo4j Sandbox

Ankit Raj, Suchitra A. Khoje, and Sagar Bhilaji Shinde

Abstract There is no universally accepted definition of terrorism. Terrorism and its ramifications have every once in a while caused massive death and destruction around the world. Current cutting-edge technologies, such as machine learning and deep learning, can predict and classify such attacks efficiently. The major difficulties observed in implementing these strategies are a lack of consistent and clean data, as well as programming knowledge in Python and R. Inconsistent data can be resolved by incorporating graph database features into the dataset, and Python programming can be replaced with the orange data mining tool. As a part of data processing and manipulation software, orange data mining tool employs a machine learning model in a non-coding context. This research study has attempted to replicate the results by using the orange tool and Neo4j Sandbox. In this study, a non-coding approach was used to classify terror attacks by using the orange data mining tool, and the use of graph embeddings as dataset features have assisted in eliminating the problems associated with inconsistent data. The dataset was then subjected to machine learning techniques such as Random Forest, Decision Tree, Support Vector Machine, Naive Bayes, Gradient Boosting, KNN, and Adaboost to classify the terror attacks. Random Forest and Gradient Boosting are the models that can achieve an accuracy score, recall, precision, and F1 score greater than 90%.

Keywords Neo4j Sandbox · Orange data mining tool · GTD · Machine learning · Graphs

A. Raj (✉) · S. A. Khoje
MIT–World Peace University, Pune, Maharashtra, India
e-mail: reachankitat@gmail.com

S. A. Khoje
e-mail: suchitra.khoje@mitwpu.edu.in

S. B. Shinde
NMVPM's Nutan College of Engineering and Research, Pune, Maharashtra, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_6

1 Introduction

Relational and non-relational databases are the two main types of databases. A non-relational form of database is known as graph database. A graph database or graph is a higher dimensional data representation where nodes and relationships are used instead of rows and columns [1]. While the characteristics of those nodes are the rows of the relational database, which indicate the number of entries in a dataset and the nodes in a graph are the entities that represent a column or attribute of the relational database [2]. The global terrorism database managed by the University of Maryland is the dataset used in the proposed research study. Working in the field of machine learning requires a strong working knowledge of the python programming language. Without any python programming experience, Orange Tool provides the flexibility to work in the domain of machine learning [3]. Here, GTD is used to develop a graph database for the proposed project. This graph database has millions of relationships between its thousands of nodes. The graph data science library found in the Neo4j Sandbox plugin was used to calculate certain properties of the graph database, including degree, centrality, and node embedding. Seven machine learning models, including decision tree, random forest, gradient boosting, KNN, SVM, Naive Bayes, and Adaboost were applied to the dataset. According to different performance metrics like AUC Score, accuracy, recall, and F1 score, the best model will be selected. The prediction results were displayed by using the confusion matrix in the orange tool [3].

1.1 Dataset

The University of Maryland-owned global terrorism database remain as the source for research data. The dataset is a compilation of every act of terrorism that has taken place around the globe between the year 1970 and 2019 in a relational dataset format. The attributes in the dataset include the timing and location of the assault, the type of weapon used, target type, causation, and more. There are 136 attributes in the dataset with two lakh entries of the terror incidents [4]. We sorted and filtered the data set due to computing resource constraints and selected 10 instances per year from 1970 to 2020. Consequently, there were 500 records in the sample dataset. Figure 1 shows the geographical spots on the world map, where terror attacks have occurred in the past. It can be seen that the South Asia region is the most terror attack-prone region on the globe [5]. The event İD, event location, event time, event date, event day, event month, event year, longitude, latitude, specificity, proximity, attack type, target type, gun type, weapon type, and others are the attributes of the GTD dataset [5].

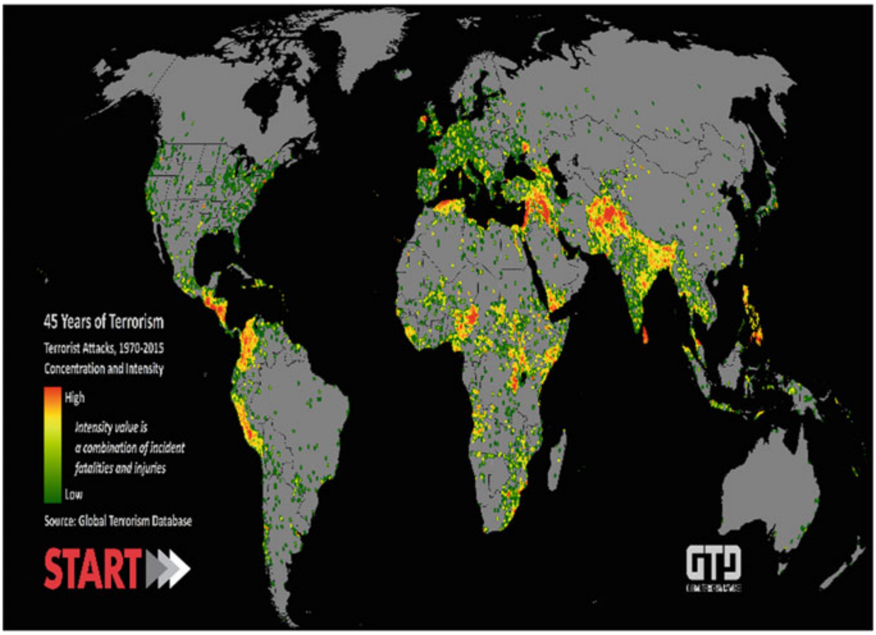

45 Years of Terrorism
Terrorist Attacks, 1970-2015
Concentration and Intensity
High
Intensity value is a combination of incident fatalities and injuries
Low
Source: Global Terrorism Database
START
GTD

Fig. 1 GTD map for showing terror attacks worldwide. *Source* www.start.umd.edu

1.2 Graph and Neo4j

Neo4j is a JavaScript-based tool for creating and manipulating graphs. CQL or cipher query language is used for its operation [6]. The graph data science library in Neo4j may be used to apply various algorithms to the graph. Neo4j's computed embedding may be exported as a CSV file. The machine learning model will use the estimated embedding, degree, and centrality as significant features [6].

2 Literature Survey

Neo4j is a javascript-based tool used for creating and manipulating graph databases that use the Cypher query language, or CQL. The manipulation and mathematical operations on the graph database are made simpler by the preloaded plugins, such as the graph data science library (GDS) and awesome procedures on Cipher (APC). According to Felix Melchor Santos Lopez and Eulogio Guillermo Santos De La Cruz Neo4j gives the database atomicity, consistency, isolation, and durability (ACID), hence it is an excellent substitute for traditional SQL (Relational Database) [1]. For machine learning applications, Orange is considered as a data mining tool that serves as a substitute for the Python and R programming languages. With many machine learning algorithms, including supervised learning techniques and unsupervised learning approaches, Orange tool was most recently released in 2016 by including a huge library for data preprocessing along with the utilization of data imputation block for removing null values from the dataset, PCA (Principle component analysis) block is commonly used for performing dimension reduction, wherein the data preprocessor block is used for data scaling, and support all the preprocessing algorithms. To show the machine learning use of Orange data mining tool, Musa Peker, Osman Ozkaraka, and Ali Sasar implemented five machine learning models on a diabetic dataset obtained from Dalaman State Hospital of Turkey [2]. Today's market offers a variety of data mining technologies, including R programming language, Rapid Miner, WEKA, Orange, and Kinme. Rapid Miner is language agnostic, whereas orange was created using C, C++ , Cython, and Python. Orange offers more freedom to the developers by offering them a load model block so that they may create their own models and send them to the orange tool [3]. The benefits and downsides of various data mining technologies were thoroughly compared by Ranjan et al. [3]. The end-user can implement a variety of machine learning models provided by WEKA using the java programming language. With a data training percentage of 66% and test data percentage of 34%, Ghada M. Tolan attempted to use machine learning models by including Naive Bayes, K-nearest neighbour, C4.5, ID3, and support vector machine. The dataset majorly used for terror attack classification is the global terror attack database, which is a copyright of the University of Maryland. It consists of 136 attributes and two lakh entries of incidents from the year 1970 to 2015. WEKA has both machine learning and deep learning support and is an open-source platform issued under GNU general public license [7]. Two of the most popular algorithms for classifying terror attacks are the decision tree and random forest algorithms. While the decision tree has never demonstrated accuracy above 75%, random forest algorithm with modified hyper-parameters has consistently demonstrated results above 90% [8]. Although GDBMS are now more widely accepted by data analysts, they nevertheless have their own drawbacks, such as high computer power requirements, longer calculation times, and more complicated algorithms when dataset sizes grow. Due to the large number of libraries that are filled with graph data science and their connection with the python programming language, Neo4j, Orient DB, and Titan are considered as the most promising graph database management technologies [9]. Neo4j operates twenty times more efficiently than conventional RDBMS, such as Postgre, when compared to the two types of RDBMS. Both the relational database and the graph database have their own advantages and thus it is impossible to say which is quicker because it relies entirely on the application for which it is being used [6]. More than twenty graph database solutions are now available on the market, including Orient DB, Arango DB White DB, Graph DB, Azure Cosmos DB, Fauna DB, Tiger Graph, Neo4j, Velocity DB, Memgraph, Titan, and many others. Of these, Neo4j and Tiger Graph are the two that perform the best. These graph databases are frequently used in the field of biomedical engineering to record patient names, identification numbers, diagnosis, and treatment information.

The advantages of graph databases over relational databases have also been demonstrated with a thorough comparison of the graph database frameworks by Timon-Reina et al. [10]. The graph database has a variety of uses, including network administration, social connectivity, biology, and the identification of fraudulent conduct. Compared to relational databases, it offers the developer more performance, flexibility, and agility [11]. Hybrid models and ensemble machine learning techniques also produce promising outcomes, with these techniques achieving results ranging from 87 to 97%. ROC, AUC, precision, recall, and F1 score are the performance measures used in the result analysis. This is mostly based on ROC curve analysis for each model [4]. Neural Network is another machine learning algorithm that when trained and tested for twenty epochs was able to give a mean squared error (MSE) of 0.180 by Ghada and Abou-El-Enien. Metaheuristic Optimization algorithms, which help in increasing the prediction accuracy of machine learning algorithms [5]. The GTD codebook gives an overview of the data collection methodology for the global terrorism database. The code is maintained by the University of Maryland as well as it is copyright of the same. The database can be used on an individual basis for study purposes and is provided by the admin on a request basis [12].

3 Methodology

3.1 Graph Creation

An application called Neo4j Sandbox was used to create the graphs. The global terrorism database was utilized as a source for building the graph and was imported into Neo4j by using the LOAD CSV cipher command. Using the CREATE command, several nodes were added to the graphs. Event Timing includes the date, time, and year of the occurrence; Event location includes the neighborhood, location, longitude, latitude, and Specificity, and attack types attacktype1, attacktype2, attack subtype 1, and attack subtype 2 were provided in the event info. Target type was composed of target types 1, 2, and target subtypes 1, and 2. Weapon type 1, weapon subtype 1, weapon type 2, weapon subtype 2, gun name, and gun type were all contained in the weapon type. Property damage, causality, and Ransome type were all the factors in causation. These nodes are all linked together through relationships. A cipher query language was used to generate and modify the graph. A subgraph of two nodes was constructed from the generated graph in order to compute graph embedding [7] (Fig. 2).

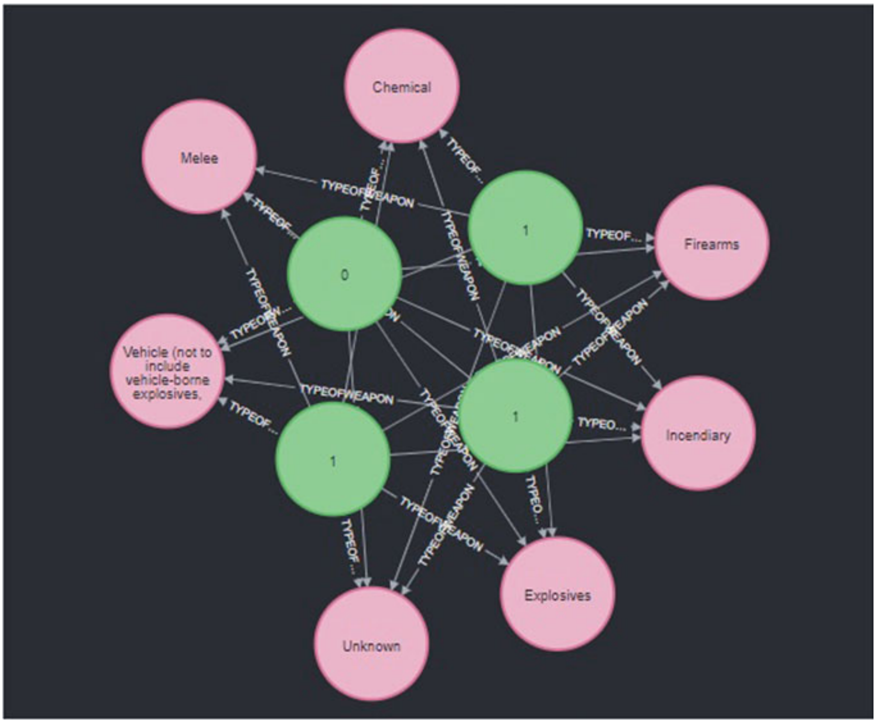


Fig. 2 Sample graph created using Neo4j Sandbox

3.2 Graph Embedding Calculation

Graph a higher dimension data is essentially represented in a lower dimension via embedding. They take the shape of a vector. In our example, embedding was calculated by using the node2vec technique. Node2vec algorithms operate based on random walks in the network. The graph is effectively represented in a lower dimension with a graph embedding by assuming a vector form [10]. The node2vec method was used to calculate embedding in our case. Node2vec methods use random network walks and are largely based on word2vec techniques. Fast random projection, node2vec, and Graphsage are the three techniques offered by Neo4j to compute node embeddings, node2vec is the approach used here. With the use of second-order random walks, the node2vec method creates a list of node identities that, when put together, constitute a sentence. This corpus of sentences is then used to calculate embedding vectors, also known as node embeddings or graph embeddings. Based on random walks, the node2vec method alternates between depth-first search and breadth-first search [11]. Although up to 10 embedding dimensions have been generated in the research, the embeddings between two nodes in a network will be calculated based on the Neo4j platform up to n dimensions.

4 Model Building

The orange data mining tool was used to create many machine learning models. The global terrorism dataset and computed embedding are the first two input datasets that are entered into the orange tool by utilizing the CSV file import block. A Data table block may be used to visualize the CSV's contents [2]. A Data table block may be used to visualize the CSV's contents. The data table block's output was provided as an input to the merge data block, which integrated the separate datasets into one dataset [3]. There are many null values in the merged dataset that can't be directly given as input into the models and thus an imputation block was utilized to eliminate those null values. The select column block was incorporated into the models once null values were eliminated. The selected column block is used to choose and remove characteristics from the dataset as well as to choose and configure the model's target variable. The purge domain block in the orange tool was used to delete and eliminate the redundant characteristics from the dataset since the data frame contains certain redundant attributes that were making the model's prediction accuracy redundant. Figure 3 shows the employed model in the orange tool where X resembles the name of the ML models.

Before feeding the dataset to the model, the dataset was scaled by using a data processor block. Overfitting is a severe problem that affects machine learning models most of the time. Here, principal component analysis (PCA) is used as a dimension reduction approach to solve this problem. The input data was then divided into

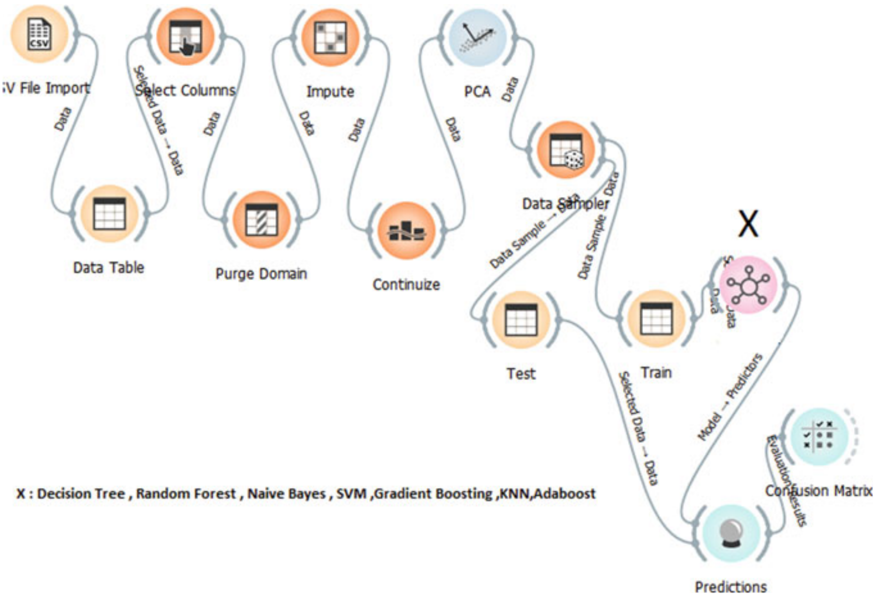
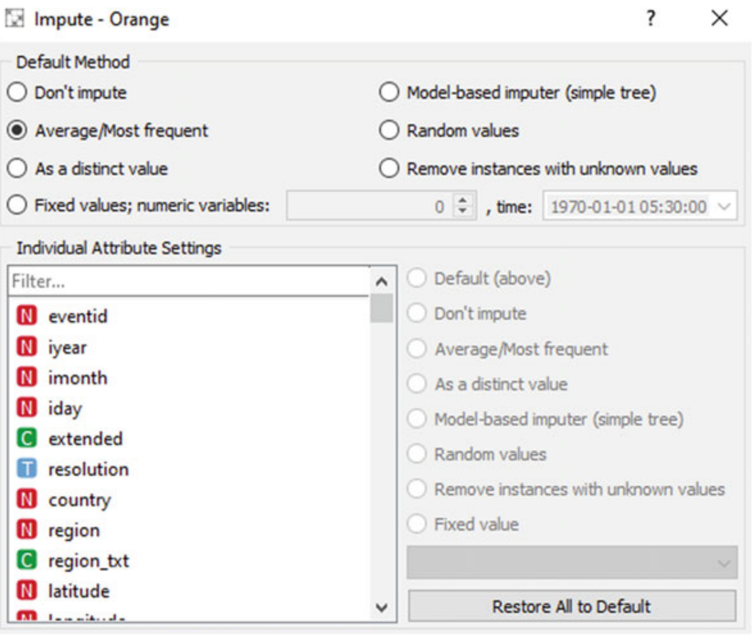


Fig. 3 Applied model in orange tool

training and testing data by using an 8:2 sampling ratio. Eighty percent of the data were used to train the model, and twenty percent were used to test it, according to the sampling ratio of 8:2. The model block was the next to be added, and it received its input from the training database and its output from that block was provided as input to the test and score block. The test data table serves as the second input for the test and score block, which also assess how well the machine learning model performed. Multiple building elements, such as a bar plot, line plot, heat map, and others, can be used to visualize the results. In our instance, a confusion matrix was employed to visualize the outcome.

4.1 Counting Null Values

The orange data mining tool's impute block was used to count and eliminate null values from the dataset. There is a choice to use the average and most frequent imputation algorithm, random value imputation algorithm, model-based imputer method, or fixed value or numeric value imputation algorithm. In this study, the most frequent and average imputation procedure was used to remove null values from the dataset. The data imputation block for removing null values from the dataset is shown in Fig. 4.



Impute - Orange

Default Method

○ Don't impute
● Average/Most frequent
○ As a distinct value
○ Fixed values; numeric variables: 0 , time: 1970-01-01 05:30:00
○ Model-based imputer (simple tree)
○ Random values
○ Remove instances with unknown values

Individual Attribute Settings

Filter...

N eventid
N iyear
N imonth
N iday
C extended
T resolution
N country
N region
C region_txt
N latitude

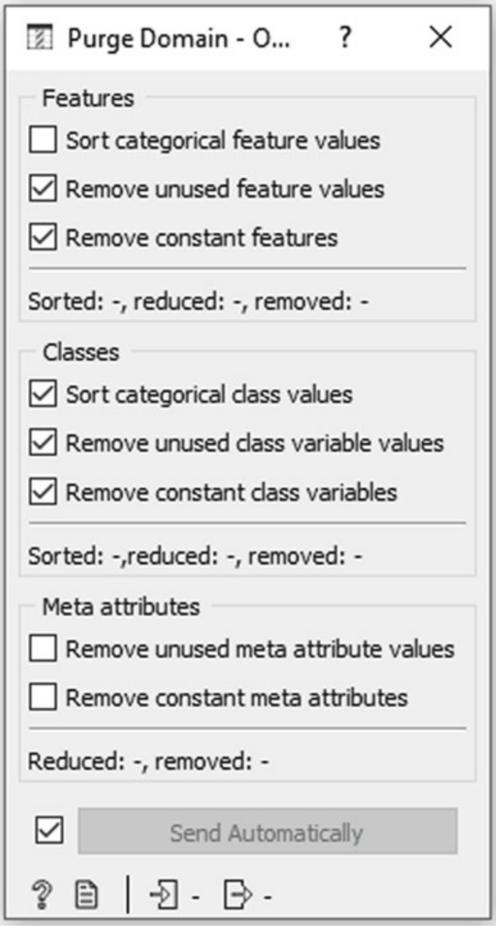
○ Default (above)
○ Don't impute
○ Average/Most frequent
○ As a distinct value
○ Model-based imputer (simple tree)
○ Random values
○ Remove instances with unknown values
○ Fixed value

Restore All to Default

Fig. 4 Data imputation block in orange

4.2 Removal of Redundant Data

The purge domain block in the orange data mining tool is used to eliminate or discard redundant characteristics from the dataset. Three alternatives are provided by the purge domain block to eliminate redundant characteristics from the dataset: features, classes, and meta attributes. The three major functions performed by purge domain block are sorting, reducing, and removing features. The purge domain block for removing redundant attributes from the dataset is shown in Fig. 5.



Features

☐ Sort categorical feature values

☑ Remove unused feature values

☑ Remove constant features

Sorted: -, reduced: -, removed: -

Classes

☑ Sort categorical class values

☑ Remove unused class variable values

☑ Remove constant class variables

Sorted: -,reduced: -, removed: -

Meta attributes

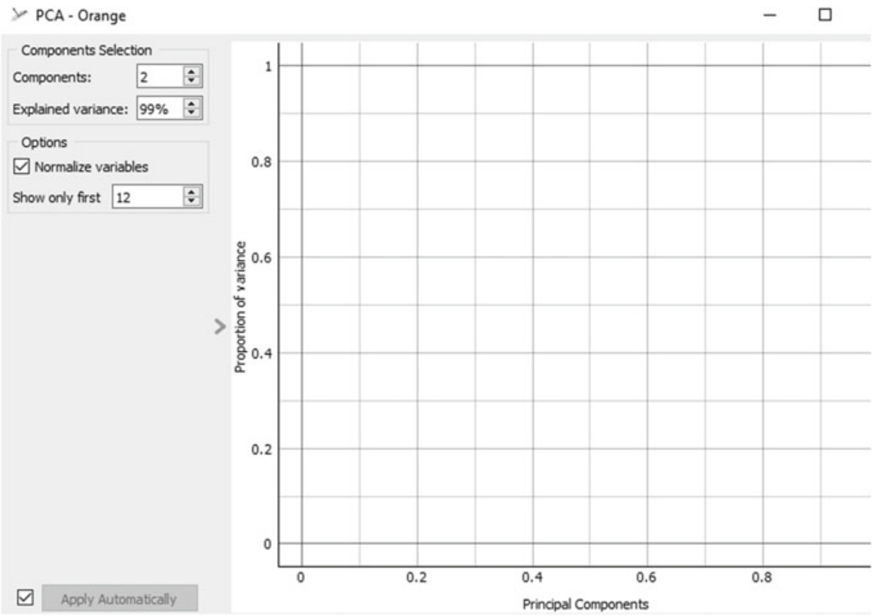
☐ Remove unused meta attribute values

☐ Remove constant meta attributes

Reduced: -, removed: -

☑ Send Automatically

Fig. 5 Purge domain block in orange tool



PCA - Orange

Components Selection

Components: 2

Explained variance: 99%

Options

☑ Normalize variables

Show only first 12

☑ Apply Automatically

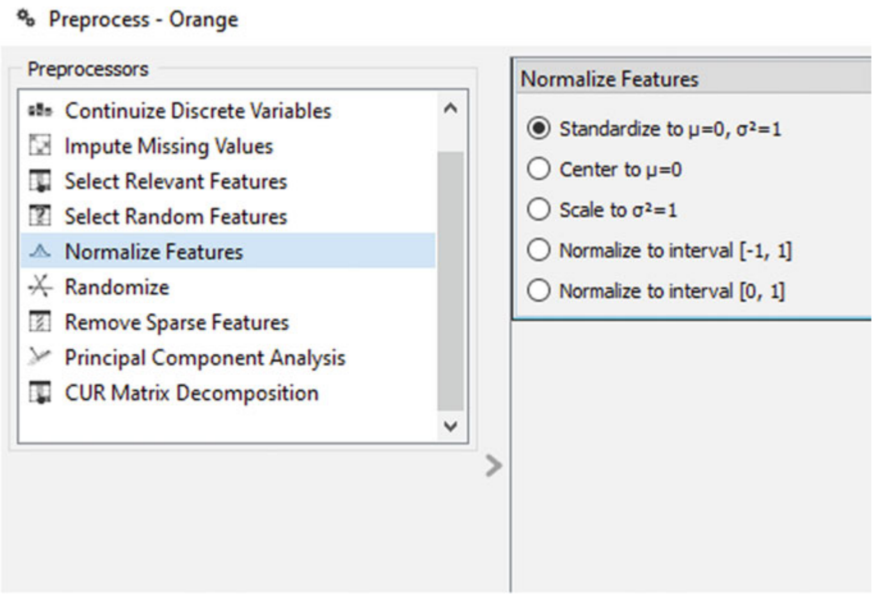
Fig. 6 PCA in orange tool

4.3 Dimension Reduction

Overfitting, which happens as a result of the dataset's many characteristics, is one of the main issues that machine learning models encounter. There are several methods for reducing the number of dimensions, including PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis). PCA (Principal Component Analysis) is used in this study to reduce the dimensions to two components. With a variance of 99%, the PCA block in the orange data mining tool is utilized for performing dimension reduction. The PCA block for dimension reduction from the dataset is shown in Fig. 6.

4.4 Data Scaling

The process of bringing the data into a certain range so the model can quickly learn and categorize is known as data scaling. The orange tool has a number of scaling techniques, including conventional scaling and center scaling. For data scaling, utilize the orange data processor block. Data preparation options available in the data preprocessor block include discretization, continuization, imputation, normalizing, randomization, and principal component analysis. The data processor block for data scaling from the dataset is shown in Fig. 7.



Normalize Features

● Standardize to μ=0, σ²=1
○ Center to μ=0
○ Scale to σ²=1
○ Normalize to interval [-1, 1]
○ Normalize to interval [0, 1]

Fig. 7 Preprocess block in orange tool

5 Results and Analysis

The obtained research findings indicate that Random Forest (RF) with accuracy scores of 0.938, F1 scores of 0.920, precision scores of 0.936, and recall scores of 0.932 is the model that performs the best. SVM is the model that performs the poorest, with accuracy scores of 0.554, F1 scores of 0.830, precision scores of 0.850, and recall scores of 0.880. AUC, precision, recall, and F1 Score are the study's performance indicators. Table 1 shows the results obtained from the proposed research.

Table 1 Results obtained from the proposed research

Model	AUC	F1 score	Precision	Recall
RF	0.938	0.920	0.936	0.932
GB	0.931	0.955	0.954	0.955
KNN	0.743	0.850	0.842	0.870
Tree	0.572	0.903	0.902	0.911
SVM	0.554	0.830	0.850	0.880
NB	0.833	0.710	0.869	0.651
Adaboost	0.862	0.927	0.931	0.925

To determine accuracy, apply the formula below:

$$Accuracy = (TP + TN)\big/(TP + TN + FP + FN)$$

To determine the precision, apply the formula below:

$$Precision = TP\big/(TP + FP)$$

To determine the recall, apply the formula below:

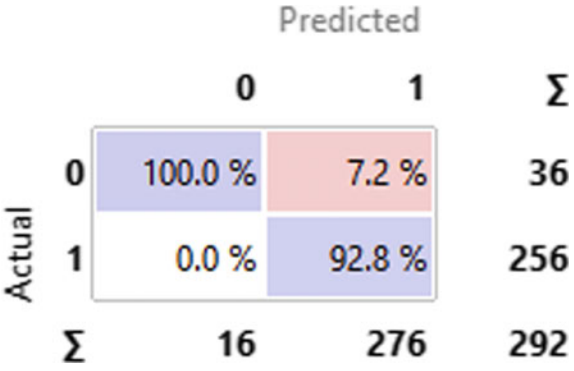
$$Recall = TP\big/(TP + FN)$$

To determine the F1 Score, apply the formula below:

$$F1\,Score = 2 * (Precision * Recall)\big/Precision + Recall$$

where TP is the true positive classified sample by the model, TN is the true negative classified sample by the model, and FN is the false positive classified sample by the model, FN is the false negative classified sample by the model. Figure 8 shows the confusion matrix for the random forest model.

Figure 8 shows the confusion matrix for the random forest model.
Figure 9 shows the confusion matrix for the gradient boosting model.
Figure 10 shows the confusion matrix for the KNN model.
Figure 11 shows the confusion matrix for the decision tree model.
Figure 12 shows the tree diagram for the classification model.
Figure 13 shows the confusion matrix for the SVM model.
Figure 14 shows the confusion matrix for the naive bayes model.
Figure 15 shows the confusion matrix for the Adaboost model.

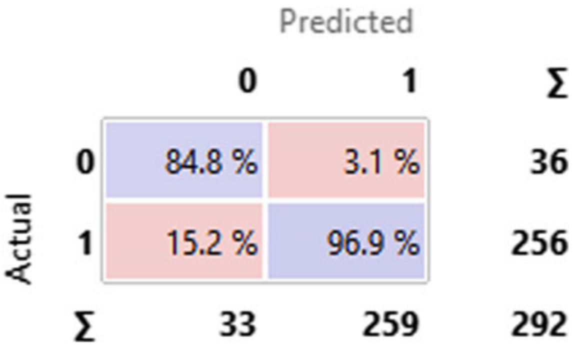


Predicted

		0	1	Σ
Actual	0	100.0 %	7.2 %	36
	1	0.0 %	92.8 %	256
	Σ	16	276	292

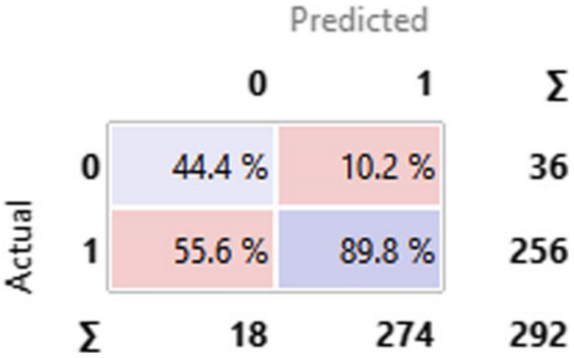
Fig. 8 Confusion matrix for random forest

Fig. 9 Confusion matrix for gradient boosting



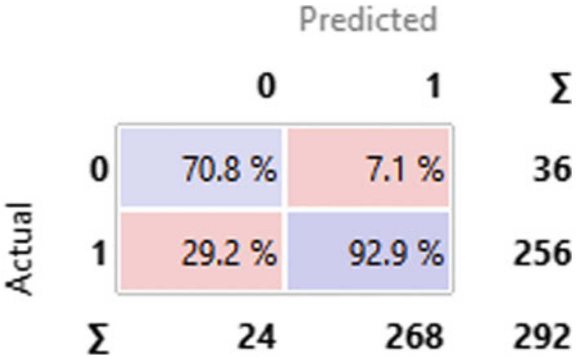
Actual \ Predicted	0	1	Σ
0	84.8 %	3.1 %	36
1	15.2 %	96.9 %	256
Σ	33	259	292

Fig. 10 Confusion matrix for KNN



Actual \ Predicted	0	1	Σ
0	44.4 %	10.2 %	36
1	55.6 %	89.8 %	256
Σ	18	274	292

Fig. 11 Confusion matrix for decision tree



Actual \ Predicted	0	1	Σ
0	70.8 %	7.1 %	36
1	29.2 %	92.9 %	256
Σ	24	268	292

6 Conclusion

Tools like Orange and WEKA can be very useful in the absence of knowledge on python programming language. Graph features such as graph embedding can act as useful features in the classification process. Random Forest (RF) and Gradient

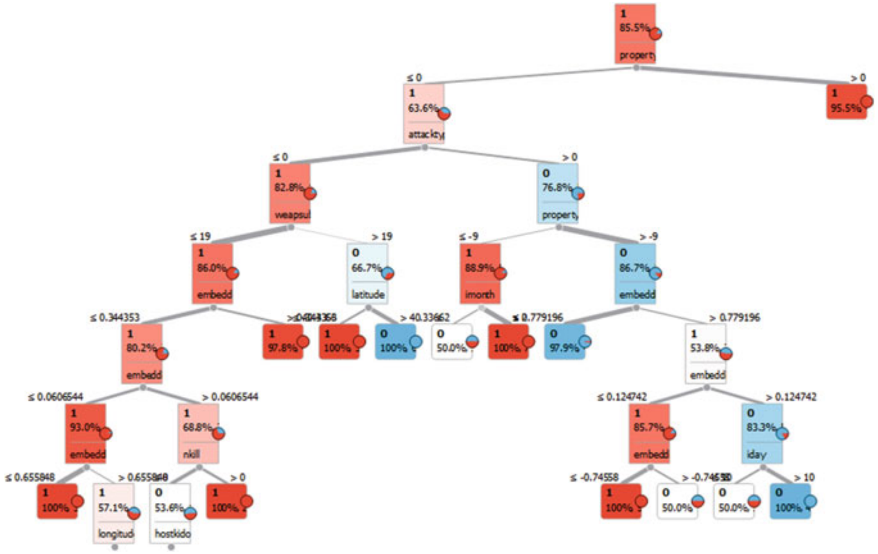
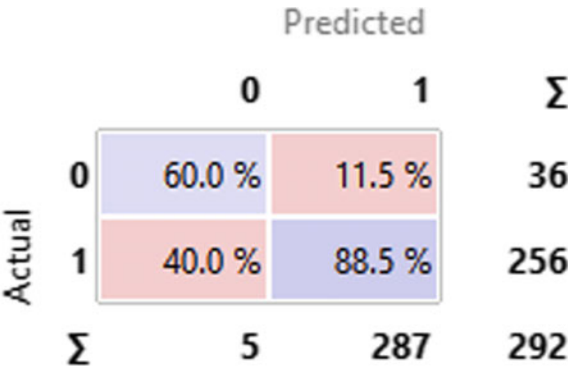


Fig. 12 Tree diagram for decision tree

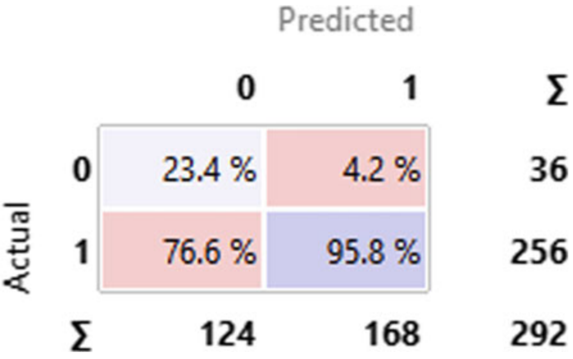
Fig. 13 Confusion matrix for SVM



Predicted

Actual	0	1	Σ
0	60.0 %	11.5 %	36
1	40.0 %	88.5 %	256
Σ	5	287	292

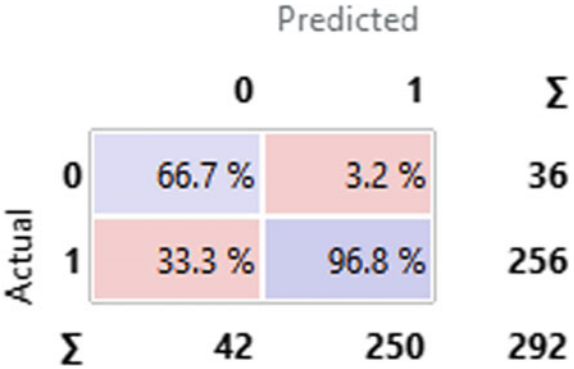
Fig. 14 Confusion matrix for naive bayes



Predicted

Actual	0	1	Σ
0	23.4 %	4.2 %	36
1	76.6 %	95.8 %	256
Σ	124	168	292

Fig. 15 Confusion matrix for Adaboost



		Predicted		
		0	1	Σ
Actual	0	66.7 %	3.2 %	36
	1	33.3 %	96.8 %	256
	Σ	42	250	292

Boosting (GB) techniques are the most promising techniques used for the classification of terror attacks by using the orange data mining tool. According to our findings, Random Forest (RF) can provide an accuracy score of 0.938, F1 score of 0.920, precision of 0.936, and recall of 0.932 on a training and testing ratio of 8:2. The worst performing model was SVM, which gave an accuracy score of 0.554, F1 score of 0.830, precision of 0.850, and recall of 0.880. Although orange is a very diverse tool with its own limitation including less flexibility, a predefined and limited number of algorithms, and less customization of blocks. Graph embedding has compensated the inconsistency in the dataset and improves the prediction accuracy of the model.

References

1. Lopez FMS, De La Cruz EGS (2015) Literature review about Neo4j graph database as a feasible alternative for replacing RDBMS. Revista de la Facultad de Ingenieria Industrial 1560–9146
2. Peker M, Özkaraca O, Şaşar A (2018) Use of orange data mining toolbox for data analysis in clinical decision making: the diagnosis of diabetes disease. In: Expert system techniques in biomedical science practice
3. Ranjan R, Agarwal S, Venkatesan S (2017) Detailed analysis of data mining tools. Int J Eng Res Technol (IJERT) 2278–0181
4. Python A, Bender A, Nandi AK, Hancock PA, Arambepola R, Brandsch J, Lucas TCD (2021) Predicting non-state terrorism worldwide. Sci Adv
5. Soliman GMA, Abou-El-Enien THM (2019) Terrorism prediction using artificial neural network. Revue d'Intelligence Artificielle
6. Macák M, Stovcik M, Buhnova B (2020) The suitability of graph databases for big data analysis: a benchmark. IoTBDS
7. Tolan GM, Soliman OS (2015) An experimental study of classification algorithms for terrorism prediction. Int J Knowl Eng
8. Huamaní EL, Alicia AM, Roman-Gonzalez A (2020) Machine learning techniques to visualize and predict terrorist attacks worldwide using the global terrorism database. Int J Adv Comput Sci Appl (IJACSA)

9. Pokorný J (2015) Graph databases: their power and limitations. In: Computer ınformation systems and ındustrial management
10. Timon-Reina´ S, Rincon M, Martínez-Tomas R (2021) An overview of graph databases and their applications in the biomedical domain. Database J Biol Database Curation
11. ShefaliPatil G, Bhatia A (2014) Graph databases—an overview. Int J Comput Sci Inf Technol
12. Global terroism database. https://www.start.umd.edu/gtd/downloads/Codebook.pdf

Multipurpose IoT Based Camera Using Deep Learning

Urvashi Dube, Sudhish Subramaniam, and G. Sumathi

Abstract The COVID 19 pandemic has given rise to a new normal. This includes wearing masks and maintaining social distance. Nowadays sudents don't focus in offline classes. Also, students with masks in offline proctored exams find ways to roll their eyes at others' work for malpractice. The systems designed to date are not accurate to detect facial features with mask. These problems have motivated us to develop a reliable, robust model to detect mask, eye location, eyeball location, eye status, and head pose of people wearing and not wearing a mask, all at once. We have used 3800 masked, unmasked images to train our model using MobileNetV2, a convolutional neural network, with 99% accuracy. The output of this model is processed using image processing, facial landmark analysis, EAR, and deep learning to detect the facial landmarks accurately. Ultimately, a unique method is used to detect head pose of person.

Keywords Image processing · Deep learning · Face mask · Eye location · Eyeball location · Head pose estimation

1 Introduction

Covid 19 pandemic has changed the style of living in the world. Although the virus is exiting, it has an impact on people. Schools and colleges have opened but people still wear masks to all places. In this masked environment, face image processing has faced a lot of challenges. To overcome all these problems, we have developed a model to detect the face mask accurately and to detect the eye, eye status, eye location, eyeball location and the head pose of the person in front of the camera. Our model is also designed to detect the attentiveness of the person in a classroom. This is one of the major problems faced in the school environment, students are not paying attention in the classroom under the masks, and this decreases the productivity of education in the schools, colleges and classes. Our model can also be used in the

U. Dube · S. Subramaniam · G. Sumathi (✉)
School of Electronics, VIT, Vellore, Tamil Nadu, India
e-mail: sumathi.g@vit.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_7

offline proctored exams where students are not allowed to roll their eyes elsewhere during the duration of the exam. To avoid malpractice, our model can be used in runtime to capture images and process the eye landmarks and head pose of the candidate. The model can be further progressed to end the test of the user if the person has attempted malpractice. We have used mobilenet_v2 to classify the 3800 images of masked and unmasked people. On getting a very high accuracy of 99%, we proceeded to filter out the eyes and other facial features. We ultimately determine mask, eye status, eye location, eyeball location, and head pose irrespective of the person wearing or not wearing a mask.

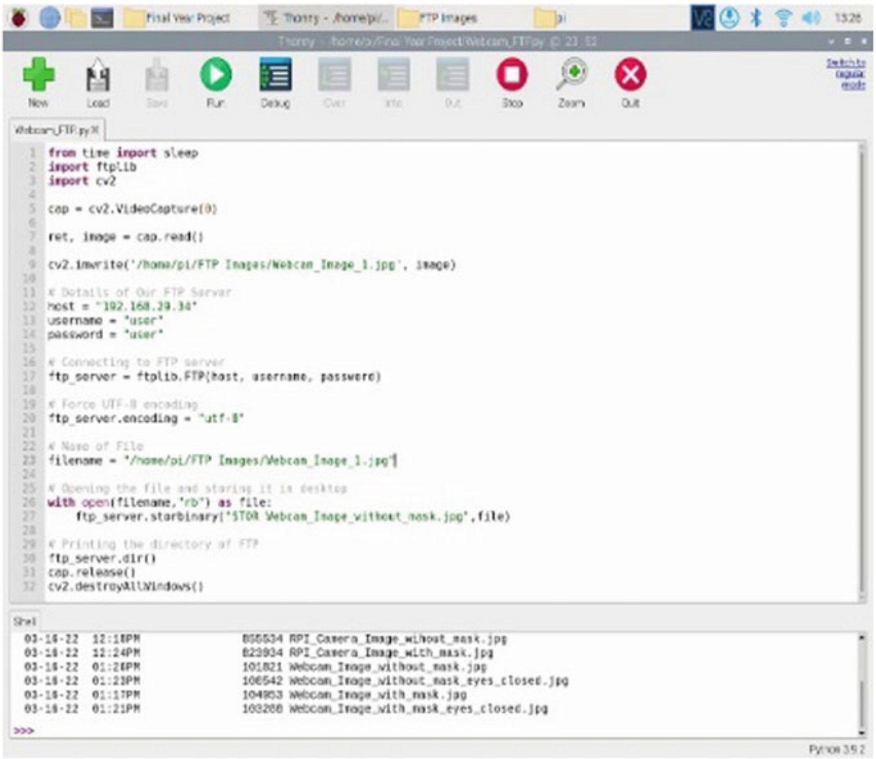
2 Related Work

The masking of a face can be identified by an edge computing-based deep learning algorithm [1], this method is specifically implemented in busses to identify people wearing masks or not. A dataset of masked faces (MAFA) and two CNN algorithms (LLE-CNN algorithm) are applied in [2]. A model comprising of PIR Sensor, microcontroller and smartphone system is proposed in [3] to detect the motion and store the corresponding output video in the cloud. In [4], a generative adversarial network for masked object detection and image completion of the removed masked region is proposed and implemented. Detection of wearing state of face mask by training a custom dataset: a face without a mask, face with the wrong mask, face with the correct mask is done using Context-Attention R-CNN method in [5]. Head Pose estimation can also be done using the HGL method which is a combination of the H-channel of the HSV colour space with the face portrait and grayscale image, this method is proposed in [6] and achieves an accuracy of 87.17%. The paper [7] includes two novel ideas, a residual context attention module for crucial face mask related regions and an auxiliary task using a synthesized gaussian heat map regression method to discriminate features of the face. The authors of [8] have proposed a model to detect candidates wearing mask regions using the transfer model of Faster_RCNN and InceptionV2 structure, in the second stage real facial masks are verified using a broad learning system. A simple and effective facial landmark detection method comprising of a lightweight U-Net model and a dynamic optical flow is proposed in [9] which exhibits better performance than others without requiring heavy computational loadings. The authors of [10] have developed a model using driving environment datasets and eye aspect ratio (EAR), to detect facial landmarks, eye location and state evaluation, they achieved an average accuracy of 93.9%. The authors of [11] used the method of segmentation of pupil and iris images by pixel to determine the eye status of the driver and his fatigue, they achieved an accuracy of 96.72%. In [12] fatigue detection convolutional network (FDCN) based CNN network was built which has a 1.0% accuracy improvement on the ZJU database on fatigue detection. DriCare, a new face-tracking algorithm to improve the tracking accuracy is developed in [13], and it achieved 92% accuracy. Eye status, PERCLOS of both coloured and infrared, fatigue is detected around the clock in [14]. Using logistic regression, EAR and analyzing facial landmarks percent eye-closure over a fixed time window (PERCLOS), blink rate, statistics of blink duration, closing speed, reopening speed and number of yawns are extracted in [15]. In [16], eyes for a frontal face are extracted precisely. Histograms of Oriented Gradient (HOG) descriptors are proved to be better than existing models in the case of human detection in [17]. You Only Look Once (YOLO) method for object detection uses regression models instead of repurposing classifiers is developed in [18], this method outperforms DPM, R-CNN methods. İn [19] overall face detection, facial feature localization, and face comparison is carried out all at once. The authors of [20] have built a model to gather environmental parameters to build a smart campus environment. The parameters include air temperature, light intensity, and humidity. The paper also carries out an in-depth study on how to store real-time data in a standard and organized manner. The work in [21] is about principal component analysis (PCA) used for feature extraction that helps in achieving superior performance. The authors of the paper have worked to achieve a high recognition rate for IoT based image recognition. On analyzing all the pre-existing researches, it is evident that a robust model to detect all the facial features and head pose of a person with a mask all at once, does not exist. This has motivated us to develop a plentiful model which extracts all the features mentioned, from an image.

3 Proposed Work

3.1 Hardware Integration

To test the model's performance with different cameras and different lighting, we have used a raspberry pi camera, webcam, CCTV security camera and laptop webcam. The raspberry pi board can be integrated easily with Rpi camera and can also be used to transfer images from one computer to the other, in the same network using file transfer protocol (FTP). The raspberry pi has a very fast processing speed and capability to run long codes as compared to other boards. Using Rpi OS, the Raspberry pi provides an interface to work on and can be programmed using python, C, etc. The raspberry pi camera or a portable webcam, is interfaced with the raspberry pi. Using a python code, the images are sent periodically through file transfer protocol to the Jupyter notebook for image processing. In the case of the raspberry pi camera, picamera module is used to interface the camera with the raspberry pi. Imwrite function of OpenCV module is used to store the images on the RPI OS. File Transfer Protocol is a set of guidelines that controls how computers transfer data across the internet from one system to another. An FTP server is first set up, by connecting the raspberry pi and laptop to the same Wi-Fi network as shown in Fig. 1. ftplib library of python is used to send images through FTP. In the case of CCTV cameras, the TAPO camera serves Real-Time Streaming Protocol (RTSP). The protocol combines intricate programming, transcoding and client server method to send video through a



```python
from time import sleep
import ftplib
import cv2

cap = cv2.VideoCapture(0)

ret, image = cap.read()

cv2.imwrite('/home/pi/FTP Images/Webcam_Image_1.jpg', image)

# Details of Our FTP Server
host = "192.168.29.34"
username = "user"
password = "user"

# Connecting to FTP server
ftp_server = ftplib.FTP(host, username, password)

# Force UTF-8 encoding
ftp_server.encoding = "utf-8"

# Name of File
filename = "/home/pi/FTP Images/Webcam_Image_1.jpg"

# Opening the file and storing it in desktop
with open(filename,"rb") as file:
    ftp_server.storbinary("STOR Webcam_Image_without_mask.jpg",file)

# Printing the directory of FTP
ftp_server.dir()
cap.release()
cv2.destroyAllWindows()
```

```
03-16-22  12:18PM                 855534 RPI_Camera_Image_wihout_mask.jpg
03-16-22  12:24PM                 823934 RPI_Camera_Image_with_mask.jpg
03-16-22  01:26PM                 101821 Webcam_Image_without_mask.jpg
03-16-22  01:23PM                 106542 Webcam_Image_without_mask_eyes_closed.jpg
03-16-22  01:17PM                 104853 Webcam_Image_with_mask.jpg
03-16-22  01:21PM                 103288 Webcam_Image_with_mask_eyes_closed.jpg
>>>
```

Fig. 1 Using FTP folder with webcam and RPI

network or to the internet using a link. Using this protocol, we have written a python code to display the stream on the Jupyter notebook. By analyzing the frames per second (FPS) we extract the frames periodically and save them in the OS using the OpenCV library. These images are then used for image processing and the results are obtained. In the case of a laptop webcam, images are captured and directly sent to the respective folder where images are extracted one at a time for image processing. The flowchart of working of hardware components of the model is shown in Fig. 2a–c.

3.2 Training the Model

Initially we import ImageDataGenerator, MobileNeyV2, AveragePooling2D, Dropout, Flatten, Dense, Input, Model, Adam, preprocess_input, img_to_array, load_img, to_categorical libraries from their respective Tensorflow.keras libraries. For preprocessing the dataset, we imported libraries from Sklearn. Other imported libraries included utils, matplotlib, NumPy, argparse and OS. Then we initialize

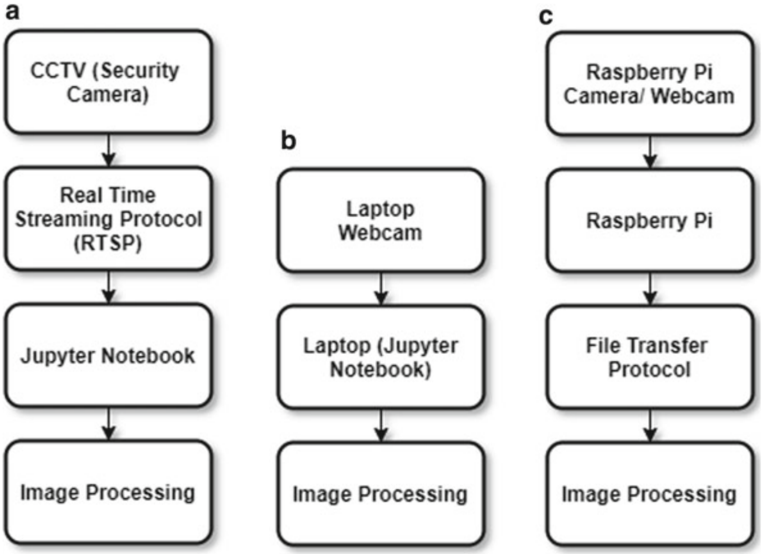


Fig. 2 **a** Flowchart of working of model with CCTV camera and Jupyter notebook, **b** flowchart of working of model with laptop webcam and Jupyter notebook, **c** flowchart of working of model with Rpi and Rpi camera

the initial learning rate, number of epochs to train for, and batch size which are determined through the hit and trial method. Once the dataset (with 3800 images) is imported, the model loops over all the images in the folder by simultaneously labelling them. Data and labels are converted into NumPy arrays and passed to a LabelBinarizer to fit the model. The dataset is split into train and test with 80% for training and 20% for testing. Data augmentation is also applied to increase the accuracy. After creating the base model with the help of MobileNetV2, head Model is made using the layers AveragePooling2D, Flatten, Dense, Dropout and again Dense layer. The head model is then placed on top of the base model. Finally, the model is fit with 20 epochs. The model is tested with the testing dataset and we get the classification report. Using this method, we have achieved an accuracy of 99% for our model.

3.3 Detect and Predict Mask

The extracted frames/images are sent as arguments to this function. FaceNet and MaskNet have defined weights for the convolutional neural network applied. Once the frame/image is obtained, first the height and weight are extracted and saved in shape. A blob image is made using the weights from faceNet. FaceNet contains a module, forward, that detects frontal faces, this is applied to the captured image. If

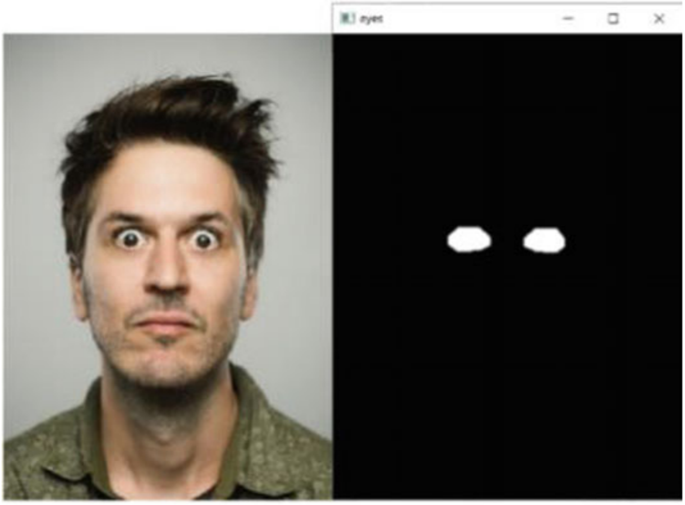


Fig. 3 Input image (left), output image (right), mask applied on the input image to detect eyes (white portion detecting eyes). *Source* Adapted from [21]

the confidence of the detected face is greater than 50%, the coordinates of the face are extracted and saved in locs. If a face has been detected, maskNet is used to detect the face with a mask on it. Once this is predicted, the coordinates are saved in preds. Ultimately a tuple of locs and preds is returned.

3.4 Eye on Mask

To detect the eyes on a face, we have applied a mask on the whole face except on the eyes. This helps to detect the eye location accurately. For this function, the predefined mask and the side of the eyes are sent as arguments. The location of the predefined eyes is saved in points, it is then converted into a NumPy array for processing. Once all the preprocessing is done, the fillConvexPoly () function is used to fill the face with a mask except for the eyes as shown in Fig. 3.

3.5 Eye Open or Closed

We have used the Eye Aspect Ratio (EAR) method to determine the status of the eye. In this method, once the eye location is obtained, the distance between the upper and lower eyelid is calculated. The points used to calculate EAR are shown in Fig. 4 (red points). This is compared with the standard value of eyes open and closed. It helps

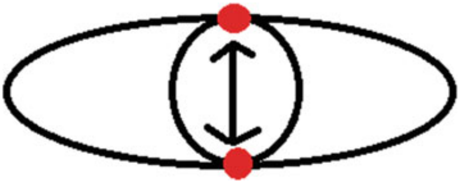


Fig. 4 Pictorial image of an eye marked red points used to calculate EAR

in determining if the eye is closed or opened. The standard values for a masked and unmasked face are different, the functions are defined respectively.

3.6 Eye and Eye Ball Contours

In the case of eye contours, the shape_68.dat is used to extract the coordinates of the eye. Once the coordinates are obtained it is sent as arguments to the eye contour function. The circle() function of OpenCV module takes five arguments, including the source picture, the (x, y) coordinates, the radius, the colour, and the thickness. The circle function draws a circle on the coordinates of the image with the mentioned radius and thickness. Obtained coordinates, radius = 2, respective color and thickness = 2 are passed to the function to draw the contours.

In the case of the eyeball contours, a threshold value is set (different for masked and unmasked faces), the midpoint of the eyeball is calculated and the threshold, midpoint and image are sent with right = False. Now, each eye is taken at a time and a mask is applied to them. Once the mask is applied, contours are detected. These contours are found by adjusting the threshold values sent. The contours detected are eyeballs. Obtained coordinates are marked using the circle function of OpenCV.

3.7 Head Pose of Person

The image obtained from hardware is sent as an argument to the function. The image is first resized to 1000 × 600 pixels and then flipped for processing. Then the image is converted to grayscale, a face mesh is created on the detected image. Using the defined weights, the landmarks of the face are detected. Once detected a loop is iterated on the coordinates. Using the nose coordinates and the other coordinates of the face, rotational and translational vectors are created using the function solvePnP of OpenCV. The rotational vector is then sent to rodrigues function of OpenCV to calculate rmat and jac. Rmat is then sent to RQDecomp 3 × 3 to get the angles (x, y, z) that are roll, pitch and yaw of the face. While testing the algorithm for different images, we found that the images form a pattern, when (x, y, z) is summed and compared. For different images, the summation of roll, pitch and yaw values had different outputs. Hence the method is generalized and it became a distinguishing factor in the head pose estimation. Finally, the summation of the angles is returned.

3.8 Function Calling and Display of Output

On reading the image, it is converted into grayscale and locs and preds are found. We predict if the person has worn a mask or not. On detecting a mask, we find the locs and preds to create a rectangle in which we detect the eye contours and eyeball contours, eye status is found using Eye Aspect Ratio. Ultimately the head pose of the person is found. The output is displayed on the image. Then we display the final output image with all the labels. If the face is unmasked, we use the detector to find the rectangle surrounding the face. Then we find the eye status, eye contours, eyeball contours and head pose. The final output image is displayed with all the labels. The flowchart in Fig. 5 illustrates the process.

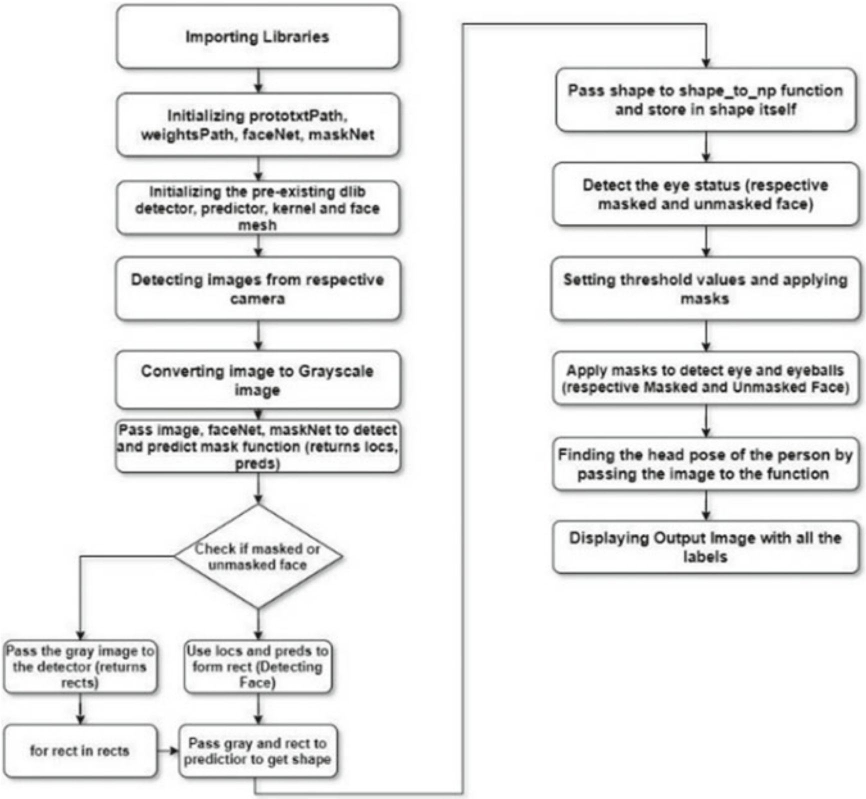

Importing Libraries
Initializing prototxtPath, weightsPath, faceNet, maskNet
Initializing the pre-existing dlib detector, predictor, kernel and face mesh
Detecting images from respective camera
Converting image to Grayscale image
Pass image, faceNet, maskNet to detect and predict mask function (returns locs, preds)
Check if masked or unmasked face
Pass the gray image to the detector (returns rects)
Use locs and preds to form rect (Detecting Face)
for rect in rects
Pass gray and rect to predictior to get shape
Pass shape to shape_to_np function and store in shape itself
Detect the eye status (respective masked and unmasked face)
Setting threshold values and applying masks
Apply masks to detect eye and eyeballs (respective Masked and Unmasked Face)
Finding the head pose of the person by passing the image to the function
Displaying Output Image with all the labels

Fig. 5 Flowchart of application of the developed algorithm

4 Result Analysis

We used a laptop webcam for the input image. The input and output images are shown in Fig. 6.

Part 2: Interfacing raspberry pi camera and laptop using File Transfer Protocol (FTP). The results of image processing are shown in Fig. 7.

Part 3: Interfacing webcam, raspberry pi and laptop for the input image we received accurate results. The processed images are shown in Fig. 8.

Part 4: Interfacing Tapo camera directly with Jupyter notebook by extracting the stream using Real-Time Streaming Protocol (RTSP). We saved the frames of the stream at regular intervals. The images and their respective processed output images are shown in Fig. 9.

Part 5: Combining all parts, the head pose of the person is also displayed on the image as shown in Fig. 10.

The training loss and accuracy versus epoch plot are shown in Fig. 13. Loss and accuracy functions from the deep Learning model, training history, in keras is used to find the training loss and accuracy of the model. It is clearly evident that the accuracy of the model is 99% and the training loss of the model is 1%, when the number of epochs reaches 20. The epoch results with 20 epochs are shown in Fig. 14. The classification report with precision, recall, f1-score and support is shown in Fig. 15. The accuracy of the model increases and validation loss of the model decreases with

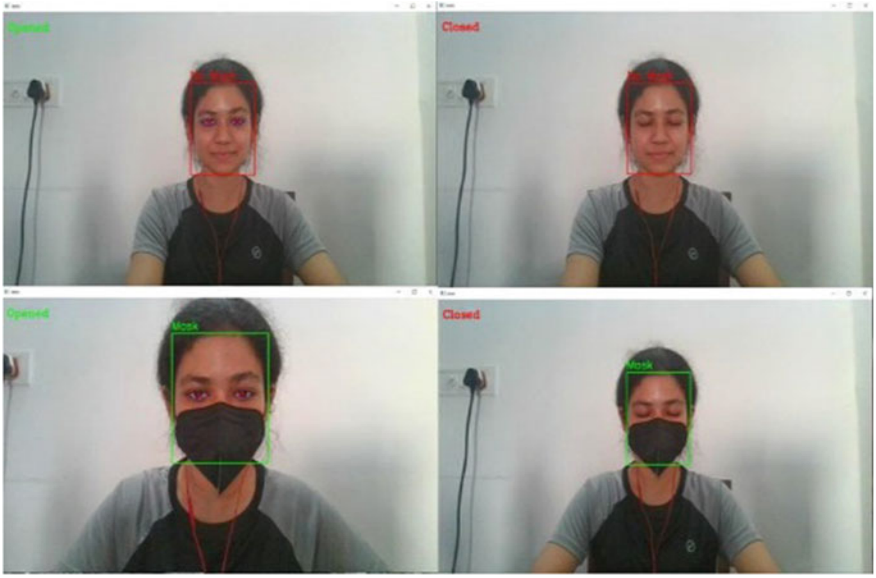


Fig. 6 Input image from a laptop webcam

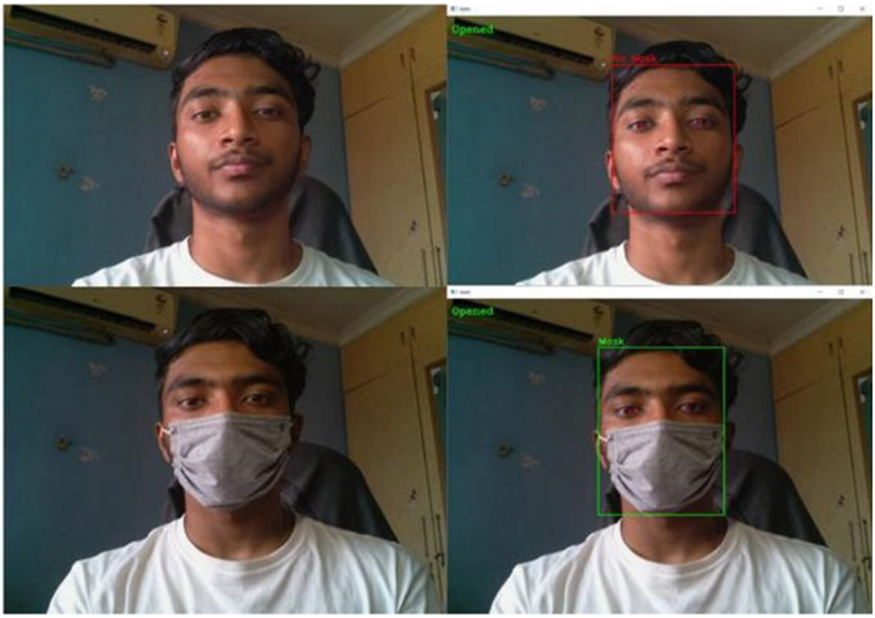


Fig. 7 Using a raspberry pi camera to get input image for masked and unmasked image

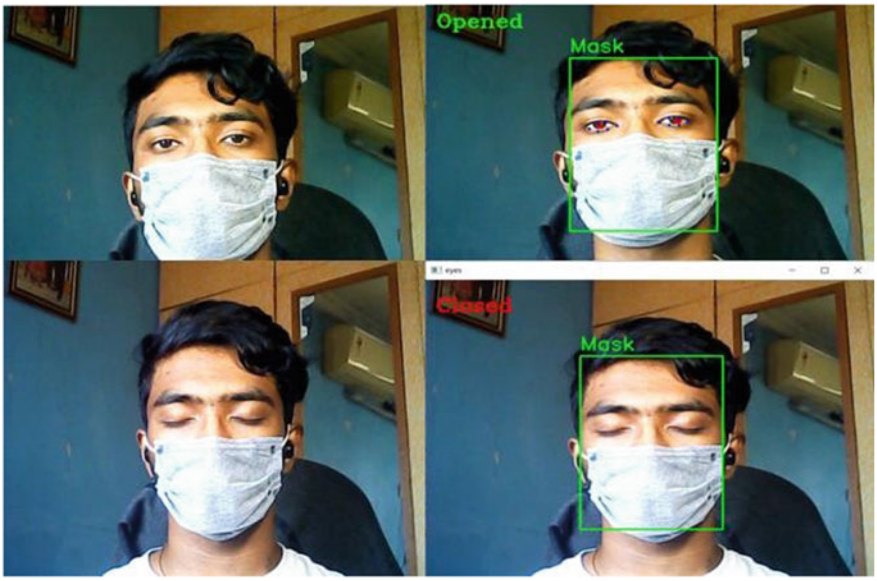


Fig. 8 Using webcam and Rpi to get input image for masked image

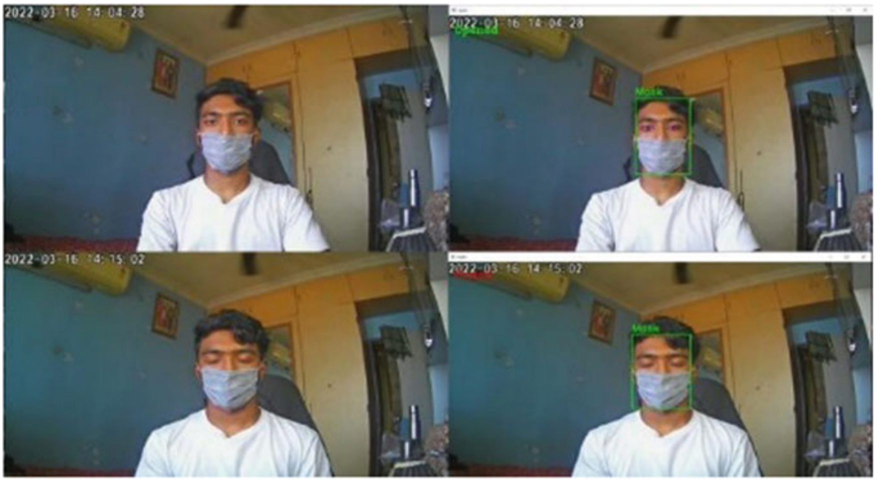


Fig. 9 Using Tapo camera to get input image

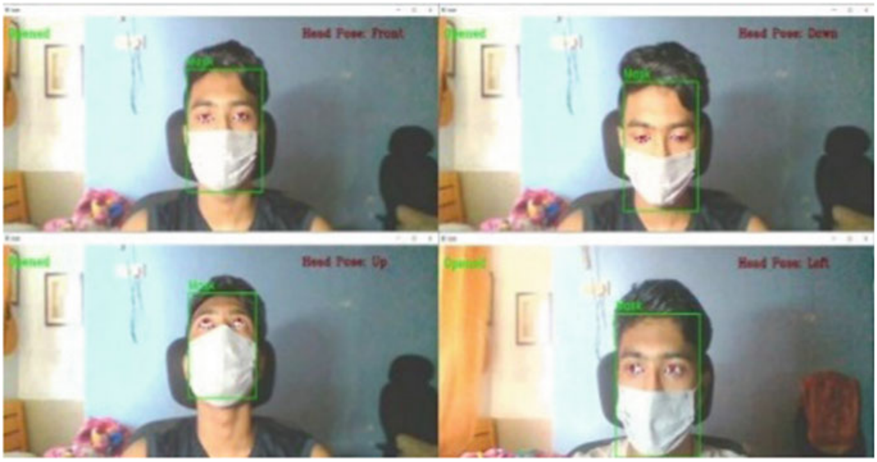


Fig. 10 Final output images with all the features and head pose

increasing epochs. Once the model is trained, it is fit into the input images to get the output. The model achieves 99% accuracy with both mask images and unmask images. The obtained accuracy is better than other existing models.

The accuracy is calculated using the below formula:

$$\text{Accuracy} = \frac{True\ category1 + True\ category2}{True\ category1 + True\ category2 + False\ category1 + False\ category2}$$

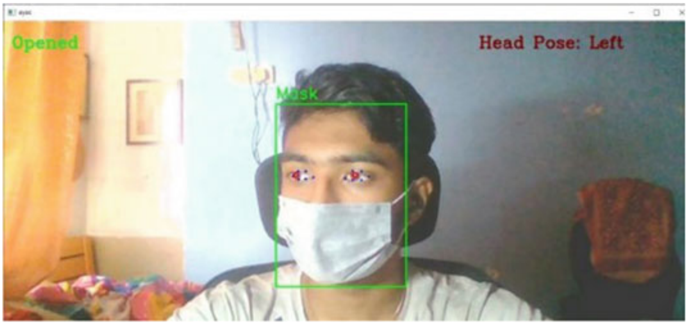


Fig. 11 Final output images with all the features and head pose (left)

Computationally, the training loss is calculated by taking the sum of errors for each example in the training set. The training loss is measured after each batch.

The results obtained are accurate and the model is working perfectly. It can be described as a robust model. The method of detecting each landmark one after the other has increased the performance of the model. The results from the laptop webcam are shown in Figs. 11 and 12. Here the image processing is done with a grayscale image, this increases the performance of the model too. The rotational vector calculated during the head pose estimation differs from one camera to another as the focal length of each camera is different. The model works best when the image processing is done at regular intervals, this reduces the processor requirements. Images are sent at regular intervals to the Jupyter notebook for image processing. Once the image processing is completed, the output image with labels is displayed. The model can be used to alert the alarm system if a person, wearing or not wearing a mask, is not attentive in a class or in a driving system. In a classroom, once the direction of the teacher or the blackboard is set, the model can be modified to detect if the pupil is attentive in the class or not. The major limitation of the model is that it cannot be used in very dim light conditions or at night, where there is minimum or no light. To overcome this problem, night vision cameras can be incorporated, to get the input stream, which can be processed by the model to get the desired results. Another limitation is the hardware interconnections between the camera and the raspberry are prone to wear and tear, and have to be handled with care for accurate results and smooth processing.

5 Conclusion

We have achieved a maximum accuracy of 99% and our model is working better than the pre-existing models. We have also worked on head pose estimation and we achieved better accuracy and precision than the existing solutions. Our model is best

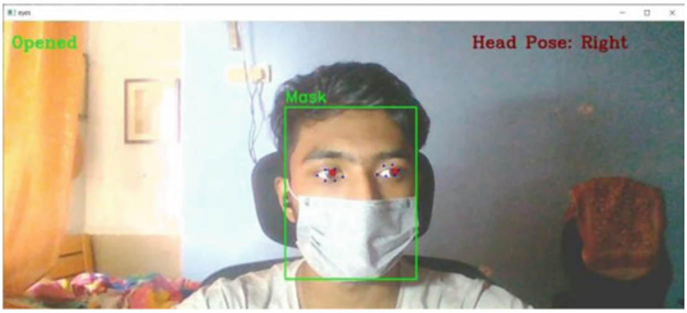


Fig. 12 Final output images with all the features and head pose (right)

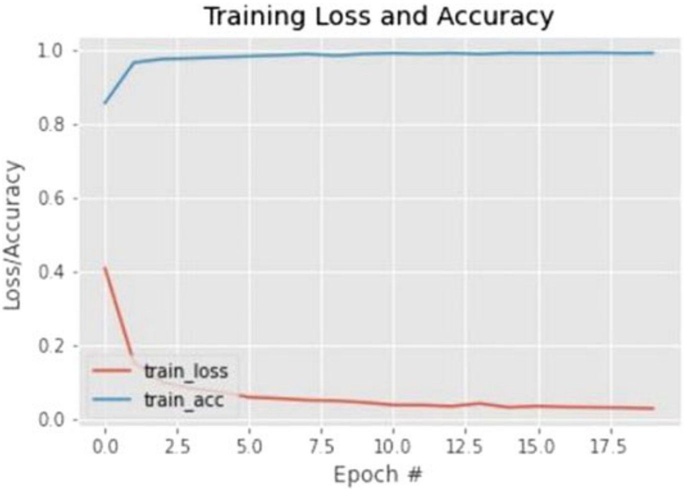
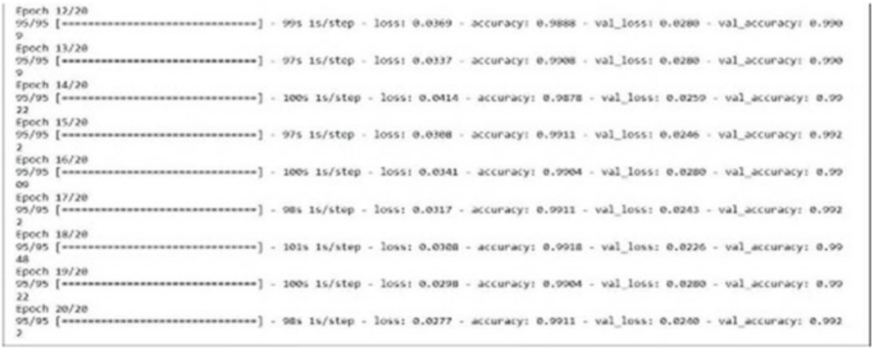
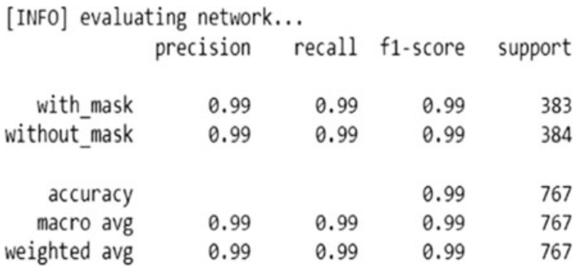


Fig. 13 Training loss and accuracy versus epoch plot



```
Epoch 12/20
95/95 [==============================] - 99s 1s/step - loss: 0.0369 - accuracy: 0.9888 - val_loss: 0.0280 - val_accuracy: 0.990
9
Epoch 13/20
95/95 [==============================] - 97s 1s/step - loss: 0.0337 - accuracy: 0.9908 - val_loss: 0.0280 - val_accuracy: 0.990
9
Epoch 14/20
95/95 [==============================] - 100s 1s/step - loss: 0.0414 - accuracy: 0.9878 - val_loss: 0.0259 - val_accuracy: 0.99
22
Epoch 15/20
95/95 [==============================] - 97s 1s/step - loss: 0.0308 - accuracy: 0.9911 - val_loss: 0.0246 - val_accuracy: 0.992
2
Epoch 16/20
95/95 [==============================] - 100s 1s/step - loss: 0.0341 - accuracy: 0.9904 - val_loss: 0.0280 - val_accuracy: 0.99
09
Epoch 17/20
95/95 [==============================] - 98s 1s/step - loss: 0.0317 - accuracy: 0.9911 - val_loss: 0.0243 - val_accuracy: 0.992
2
Epoch 18/20
95/95 [==============================] - 101s 1s/step - loss: 0.0308 - accuracy: 0.9918 - val_loss: 0.0226 - val_accuracy: 0.99
48
Epoch 19/20
95/95 [==============================] - 100s 1s/step - loss: 0.0298 - accuracy: 0.9904 - val_loss: 0.0280 - val_accuracy: 0.99
22
Epoch 20/20
95/95 [==============================] - 98s 1s/step - loss: 0.0277 - accuracy: 0.9911 - val_loss: 0.0240 - val_accuracy: 0.992
2
```

Fig. 14 Epoch results of the trained model



```
[INFO] evaluating network...
              precision    recall  f1-score   support

   with_mask       0.99      0.99      0.99       383
without_mask       0.99      0.99      0.99       384

    accuracy                           0.99       767
   macro avg       0.99      0.99      0.99       767
weighted avg       0.99      0.99      0.99       767
```

Fig. 15 Classification report of fit model

fit in real-time applications and can be used in different scenarios in driving environments, schools, colleges, and offline proctored exams to examine the attentiveness of a person accurately. The model can be trained and modified to get appropriate results in the night too. Night vision cameras can be incorporated in the model to get images in the night. The night vision cameras could be fit inside a car, where the lighting conditions are very less or negligible, to check the status, eye status and head pose, of a masked driver. In case of any abnormality in the driver status or negligence in driving an alarm could be triggered to aware about a possible accident. Our developed model outperforms the pre-existing models, faster_RCNN, inceptionV2 structure, fatigue detection convolutional network (FDCN), Histograms of Oriented Gradient (HOG), of image processing to detect if a person has worn a mask or not or to detect facial features without a mask. The achieved accuracy of the model makes it robust and fit for all lighting conditions and angles.

References

1. Kong X et al (2021) Real-time mask identification for COVID-19: an edge-computing-based deep learning framework. IEEE Internet of Things J 8(21):15929–15938. https://doi.org/10.1109/JIOT.2021.3051844
2. Ge S, Li J, Ye Q, Luo Z (2017) Detecting masked faces in the wild with LLE-CNNs. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp 426–434. https://doi.org/10.1109/CVPR.2017.53
3. Sanoob AH, Roselin J, Latha P (2016) Smartphone enabled intelligent surveillance system. IEEE Sens J 16(5):1361–1367. https://doi.org/10.1109/JSEN.2015.2501407
4. Din NU, Javed K, Bae S, Yi J (2020) A novel GAN-based network for unmasking of masked face. IEEE Access 8:44276–44287. https://doi.org/10.1109/ACCESS.2020.2977386
5. Zhang J, Han F, Chun Y, Chen W (2021) A novel detection framework about conditions of wearing face mask for helping control the spread of COVID-19. IEEE Access 9:42975–42984. https://doi.org/10.1109/ACCESS.2021.3066538
6. Li S et al. (2020) Multi-angle head pose classification when wearing the mask for face recognition under the COVID-19 coronavirus epidemic. In: 2020 International conference on high performance big data and intelligent systems (HPBD&IS), pp 1–5. https://doi.org/10.1109/HPBDIS49115.2020.9130585
7. Fan X, Jiang M, Yan H (2021) A deep learning based light-weight face mask detector with residual context attention and Gaussian heatmap to fight against COVID-19. IEEE Access 9:96964–96974. https://doi.org/10.1109/ACCESS.2021.3095191

8. Wang B, Zhao Y, Chen CLP (2021) Hybrid transfer learning and broad learning system for wearing mask detection in the COVID-19 Era. IEEE Trans Instrum Meas 70:1–12, Art no. 5009612. https://doi.org/10.1109/TIM.2021.3069844
9. Wu B-F, Chen B-R, Hsu C-F (2021) Design of a facial landmark detection system using a dynamic optical flow approach. IEEE Access 9:68737–68745. https://doi.org/10.1109/ACCESS.2021.3077479
10. Ling Y, Luo R, Dong X, Weng X (2021) Driver Eye location and state estimation based on a robust model and data augmentation. IEEE Access 9:67219–67231. https://doi.org/10.1109/ACCESS.2021.3076365
11. Zhang F, Su J, Geng L, Xiao Z (2017) Driver fatigue detection based on eye state recognition. In: Proceedings of the international conference on machine vision and information technology (CMVIT), pp 105–110
12. Huang R, Wang Y, Guo L (2018) P-FDCN based eye state analysis for fatigue detection. In: Proceedings of the IEEE 18th international conference on communication technology (ICCT), pp 1174–1178
13. Deng W, Wu R (2019) Real-time driver-drowsiness detection system using facial features. IEEE Access 7:118727–118738
14. Sun Y, Yan P, Li Z, Zou J, Hong D (2020) Driver fatigue detection system based on colored and infrared eye features fusion. Comput Mater Continua 63(3):1563–1574
15. Cheng Q, Wang W, Jiang X, Hou S, Qin Y (2019) Assessment of driver mental fatigue using facial landmarks. IEEE Access 7:150423–150434
16. Colak ME, Varol A (2020) Easymatch-an eye localization method for frontal face images using facial landmarks. Tehnički Vjesnik 27(1):205–212
17. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition (CVPR), pp 886–893
18. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition (CVPR), pp 779–788
19. Colmenarez AJ, Huang TS (1998) Face detection and recognition. In: Face recognition. Springer, Berlin, Heidelberg, pp 174–185
20. Sungheetha A (2021) Assimilation of IoT sensors for data visualization in a smart campus environment. J Ubiquit Comput Commun Technol 3(4):241
21. Jeena JI, Darney PE (2021) Design of deep learning algorithm for IoT application by image based recognition. J ISMAC 3(03):276–290

Dr. Watson AI Based Healthcare Technology Project

N. Suresh Kumar, S. Ganesh Karthick, K. P. Aswin Kumar, S. Balaji, and T. Nandha Sastha

Abstract The aim of the design is to implement Artificial Intelligence in the Healthcare domain and find suitable results. This work is an AI based web application which enables four features 1. Medicine Prescriber 2. Diabetic Analyzer 3. Covid-19 Predictor and 4. AI based chatbot. Provide online solution for the patients like prescribing medicine based on the symptoms, Analyzing the blood sugar (Mg/Dl) and suggesting food diet based on the age and sugar level, Predicting Covid-19 Positive or negative according to the X-ray of the chest and AI based Chatbot which acts as a support agent guides the user in using this application and tells interesting facts on Covid-19. The UI design of the web application is crafted using Adobe XD. Machine learning and Deep learning techniques are used to predict results for these features, these Machine learning and Deep Learning models are deployed as Web application using a framework called Flask. IBM Watson Assistant which is used to create the chatbot, allows you to integrate conversational interfaces into any app, device, or medium, as well as add a natural language interface to the app to automating conversations with your customers.

Keywords Dr. Watson AI based healthcare technology · Medicine prescriber · Covid predictor · Diabetic analyser · Chatbot

N. Suresh Kumar
Department of IT, Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India

N. Suresh Kumar · S. Ganesh Karthick (✉) · K. P. Aswin Kumar · S. Balaji · T. Nandha Sastha
Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India
e-mail: ganeshkarthick.1905024@srec.ac.in

K. P. Aswin Kumar
e-mail: aswinkumar.1905010@srec.ac.in

S. Balaji
e-mail: Balaji.1905011@srec.ac.in

T. Nandha Sastha
e-mail: nandha.1905053@srec.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_8

1 Introduction

Patients, doctors, and hospital managers' life are made easier by artificial intelligence, which performs activities normally performed by people in a relatively short time and at a fraction of the investment [1]. The AI sector, which was estimated at around $600 million in 2014 and is expected to hit $150 billion by 2026, is one of the world's fastest-growing sectors [2]. In the simplest sense, AI is something that whenever computers as well as any other machines can study, analyze, and decide things or take actions in the same way that humans can [3]. The use of machines to evaluate and act on medical information, usually with the purpose of predicting a specific outcome, is referred to as AI in healthcare [4]. AI in healthcare can improve patient outcomes through improving preventative care and livability, as well as producing more accurate diagnosis and treatment regimens. By data analysis from the government, hospitals, as well as other resources, AI can also anticipate and track the spread of contagious diseases [5]. As a result, AI has the potential to be a critical instrument in the fight against diseases and pandemics in global public health. Artificial intelligence in medicine uses Machine learning algorithms to search medical data for ideas that can assist increase health and treatment outcomes [6]. Clinical decision assistance and image analysis are now the most prominent uses of AI in medical contexts [7]. Clinical decision support tools assist practitioners in making therapy, medicine, mental health, and other patient-related decisions by providing quick access to relevant information or research. AI, unlike humans, does not require sleep [8]. Machine learning algorithms could be used to monitor the vital signs of critically ill patients and notify clinicians if specific risk indicators rise.

Artificial Intelligence is used to spot patterns in behaviour that lead to either high or low blood sugar levels in diabetes patients [9]. Continuous glucose monitors used by those with diabetes collect a huge amount of data that has previously not been used efficiently. Type 2 diabetes is observed to be more frequent than type 1. Type 1 diabetes affects more than one out of every ten persons, whereas type 2 affects the rest. With age, the percentage of diabetics rises. Diabetes affects approximately 10.5% of the general population. In the 65 and older age categories, the rate is as high as 27%. Here the AI analyse the blood sugar value and personalized healthcare recommendations like food diet. This project also focusses on predicting Covid-19 positive or negative according to the X-ray soft copy of the chest, this prediction is made based on the Deep learning technique where large amount of chest X-ray images dataset of soft copy of covid positive and negative images are trained [10]. An AI based virtual chatbot is developed using IBM Watson Assistant is a white label cloud service that enables business software developers to integrate an artificial intelligence virtual assistant into their product and brand it as their own. The IBM Cloud is used to deliver the service, which allows customers access to Watson AI. Here this chatbot is developed to tell facts on Covid-19 and guides the user in using this Web application.

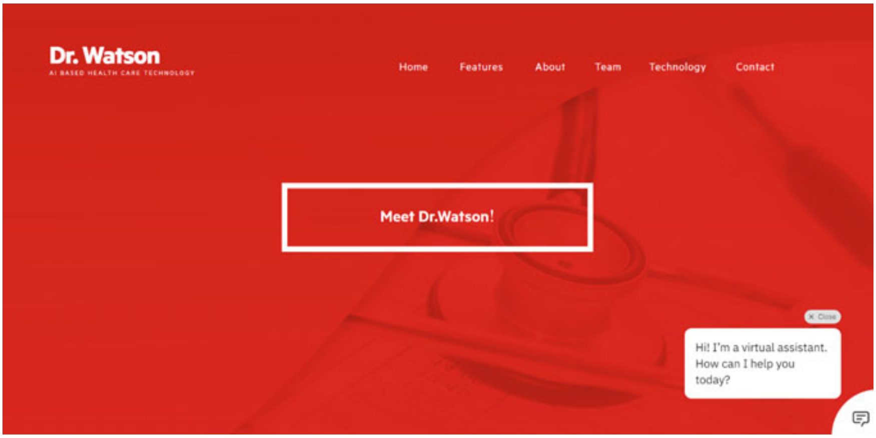
2 Methodology

The entire Web application runs on a Framework called Flask, where this application contains four features. 1. Medicine Prescriber 2. Diabetic Analyzer They have a unique dataset for each feature and Machine Learning Models are created for each feature with their respective dataset. Here the libraries used to train the machine learning model are Pandas, NumPy, SciKit, Pickle. This Machine Learning model is created using Decision Tree classifier which by constructing a decision tree, the categorization model is created. Every node in the tree represents a test on a variable, and each branch falling from that node represents one of the property's possible values. Each branch represents one of the instance's class labels. The training set's instances are identified by routing them from the base of the tree to a leaf, based on the results of the tests all along way. Each node in the tree divides the instance space into consists of two subs based on an attribute test condition, starting with the root node. After that, a new node is produced by moving down the tree branch matching to the attribute's value. This cycle continues for the subtree rooted just at new node, until the training set's records have all been classified. The decision tree is normally built from the top down, with each step selecting the optimal attribute test condition for splitting the data. There are a variety of methods for determining the optimal reason to divide the records.

3. Covid Predictor, a large amount of chest X ray is used as dataset and trained using Convolutional Neural Networks (CNN) are plug neural networks that are used to assess visual images by data processing in a grid-like layout. CNN is selected to train a deep learning model since it trains the input images into many hidden layers and brings out the trained classified output. A ConvNet is another name for it. To discover and classify items in an image, a convolutional neural network is employed. An accuracy of 88.7% is attained in this deep learning model. This Deep Learning model is created using TensorFlow, Keras and OpenCV libraries modules. Pickle is a library which is used to save the machine learning and deep learning model and deploy those saved model as a web application using the framework called Flask. Flask is a python scripted language microframework for creating small websites, and creating Restful APIs in Python is quite simple. The UI of the web application is crafted using Adobe XD and built as a web application using front end frameworks like HTML, CSS, JavaScript, Bootstrap.

3 Implementation

The home page which contains the Logo and header menus, and a chatbot below. To view the features user, need to click Meet Dr. Watson! Button. The chatbot which helps the user to guide how to use these features and get interesting facts on covid-19. Dashboard screen is a graphic representation of all of the characteristics. Although it can be utilized in a variety of ways, its primary goal is to enable quick access



Hi! I'm a virtual assistant. How can I help you today?

Fig. 1 Home page

to information. This enables user to view three features 1. Medicine prescriber 2. Diabetic Analyzer 3. Covid Predictor. Medicine prescriber feature allow user to enter their symptom and get medicine prescription. Diabetic analyser page which helps user in analysing their diabetic and get proper food diet based on their age and sugar level (mg/dl). The user can also download the result for future purpose.

Covid predictor feature predicts covid positive or negative based on the X-ray soft copy of chest. User is required to upload the soft copy of the chest X-ray and click submit to view the result. Team is a page where the team members names are mentioned in. to navigate to team simply click team button in header. Technology is a page where the technology used in this project are mentioned in a descriptive manner. To navigate to this page user simply need to click the Technology button in the header.

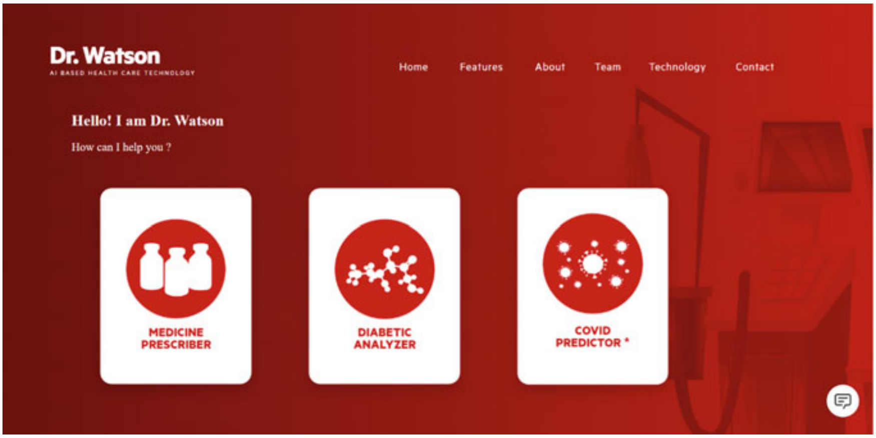
4 Results

4.1 Home Page

Select “Meet Dr. Watson” button and enter the Dashboard page (Fig. 1).

4.2 Dashboard Page

Dashboard screen will help the user to get directed to the features (Fig. 2).



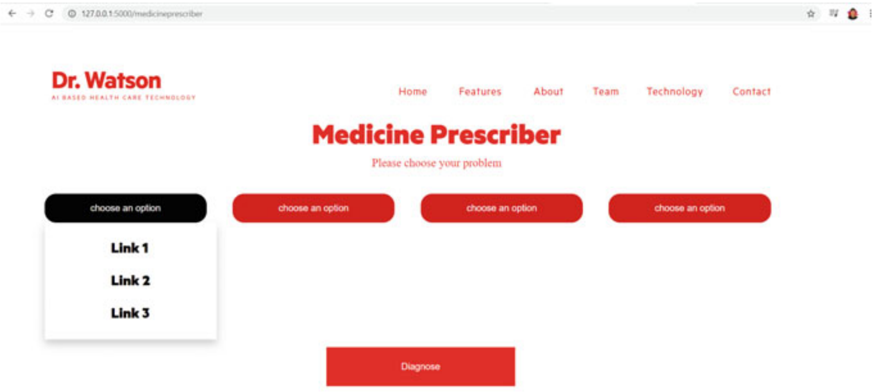
Hello! I am Dr. Watson

How can I help you ?

Fig. 2 Dashboard page

4.3 Medicine Prescriber

This feature prescribes medicine based on the user symptoms. The decision tree classifier algorithm predicts the best suitable medicine for the symptoms and allow the user to download and print the medicine as a medical prescription (Fig. 3).



Please choose your problem

Fig. 3 Medicine prescriber

4.4 Diabetic Analyzer

This feature analyses the user input which is their age and blood sugar level (mg/dl) of before food and after food and make predictions using the machine learning algorithm and provide the diabetic condition status and suggest proper food diet based on the user input. The user can download and print that food diet to kept a note of it (Fig. 4).

4.5 Covid Predictor

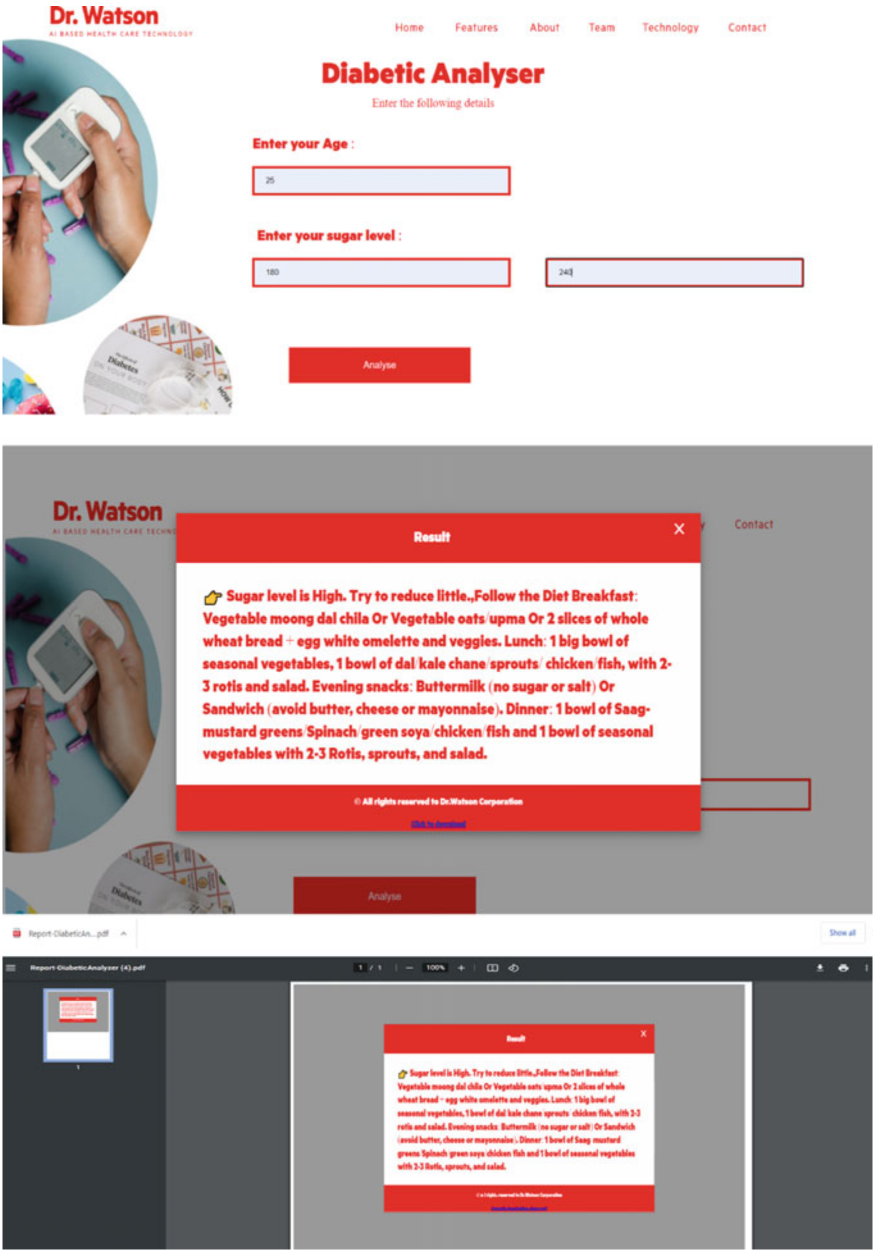
This feature allows user to upload the chest X-ray softcopy and check whether Covid positive or negative, when the softcopy of the chest X-ray scan which is provided as a user input the Convolutional Neural Network model predicts best suitable results. An accuracy of 88.7% is attained in this work, enables user to experience a fast response (Fig. 5).

4.6 Watson Assistant Chatbot

This feature guides the user to use this web application and acts as a support agent for this application and also capable of answering questions related to Covid-19. These Covid facts are trained based on the World Health Organization (WHO) data using Natural Language Processing (NLP) (Fig. 6).

5 Conclusion

Health care is progressively expanding into the house, involving a mix of individuals, a variety of jobs, and a wide range of instruments and technologies; it also occurs in a variety of residential settings. Rising health-care costs, rising numbers of seniors, rising prevalence of chronic disease, improved sustenance rates of various illnesses, injuries, as well as other conditions, massive groups of returning veterans from war with serious injuries, and a broad range of new technologies are all driving this migration. The quality and cost of the health care provided as a result varies greatly in terms of safety, efficiency, and efficiency as well as performance and price.



Diabetic Analyser

Enter the following details

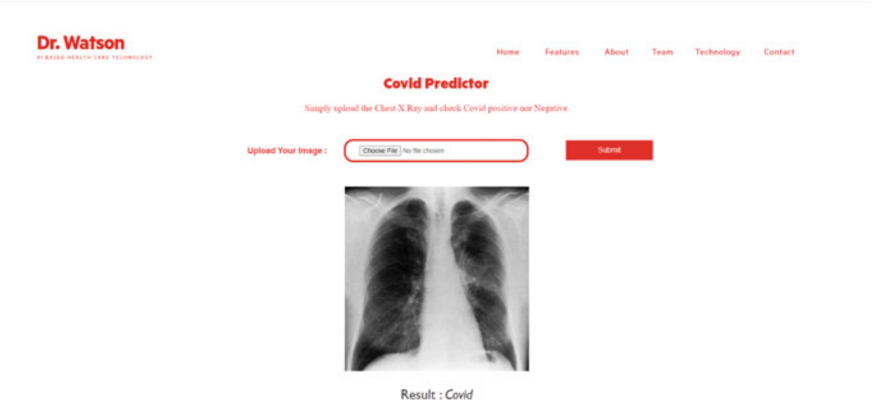
Enter your Age : 25

Enter your sugar level : 180 240

Sugar level is High. Try to reduce little.,Follow the Diet Breakfast: Vegetable moong dal chila Or Vegetable oats/upma Or 2 slices of whole wheat bread + egg white omelette and veggies. Lunch: 1 big bowl of seasonal vegetables, 1 bowl of dal/kale chane/sprouts/ chicken/fish, with 2-3 rotis and salad. Evening snacks: Buttermilk (no sugar or salt) Or Sandwich (avoid butter, cheese or mayonnaise). Dinner: 1 bowl of Saag-mustard greens/Spinach/green soya/chicken/fish and 1 bowl of seasonal vegetables with 2-3 Rotis, sprouts, and salad.

Sugar level is High. Try to reduce little.,Follow the Diet Breakfast: Vegetable moong dal chila Or Vegetable oats/upma Or 2 slices of whole wheat bread + egg white omelette and veggies. Lunch: 1 big bowl of seasonal vegetables, 1 bowl of dal/kale chane/sprouts/ chicken/fish, with 2-3 rotis and salad. Evening snacks: Buttermilk (no sugar or salt) Or Sandwich (avoid butter, cheese or mayonnaise). Dinner: 1 bowl of Saag-mustard greens/Spinach/green soya/chicken/fish and 1 bowl of seasonal vegetables with 2-3 Rotis, sprouts, and salad.

Fig. 4 Diabetic analyzer

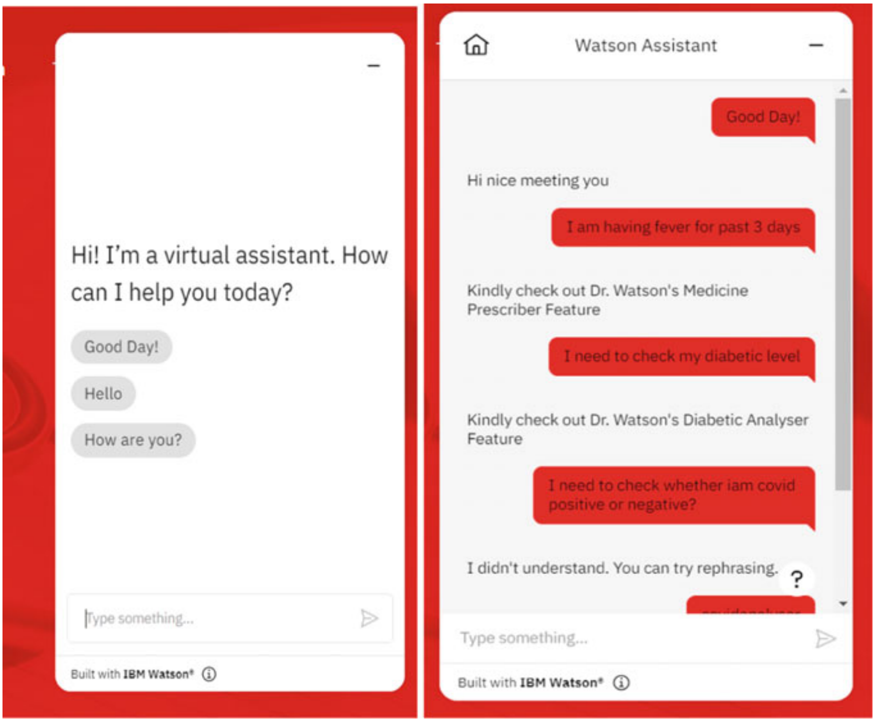


Covid Predictor

Simply upload the Chest X Ray and check Covid positive nor Negative

Result : *Covid*

Fig. 5 Covid predictor



Hi! I’m a virtual assistant. How can I help you today?

Good Day!

Hello

How are you?

Good Day!

Hi nice meeting you

I am having fever for past 3 days

Kindly check out Dr. Watson's Medicine Prescriber Feature

I need to check my diabetic level

Kindly check out Dr. Watson's Diabetic Analyser Feature

I need to check whether iam covid positive or negative?

I didn't understand. You can try rephrasing.

Fig. 6 Watson assistant chatbot

References

1. Chen X, Xie H, Zou D, Hwang GJ (2020) Application and theory gaps during the rise of artificial intelligence in education. Comput Educ Artif Intell 1:100002
2. Kandlhofer M, Steinbauer G, Hirschmugl-Gaisch S, Huber P (2016) Artificial intelligence and computer science in education: from kindergarten to university
3. Tobore I, Li J, Yuhang L et al. (2019) Deep learning intervention for health care challenges: some biomedical domain considerations
4. Rathore PS, Sharma BK (2022) Improving healthcare delivery system using business intelligence. J IoT Soc Mob Anal Cloud 4(1):11–23
5. Mason A, Morrison A, Visintini S (2018) An overview of clinical applications of artificial intelligence
6. Rajkomar A, Dean J, Kohane I (2019) Machine learning in medicine. N Engl J Med
7. Gajane P, Pechenizkiy M (2017) On formalizing fairness in prediction with machine learning
8. Adamson AS, Smith A (2018) Machine learning and health care disparities in dermatology
9. Collins GS, Moons KGM (2019) Reporting of artificial intelligence prediction models
10. Smys S (2019) Survey on accuracy of predictive big data analytics in healthcare. J Inf Technol 1(02):77–86

Empirical and Statistical Comparison of RSA and El-Gamal in Terms of Time Complexity

Ankita Kumari, Prashant Pranav, Sandip Dutta, and Soubhik Chakraborty

Abstract In this paper, two algorithms are compared based on their time complexity. The time complexity is defined by encryption and decryption of different message lengths. Time varies for different lengths of messages. We statistically analyzed the time complexity of the algorithm and compared their results.

Keywords RSA · MANET · El-Gamal · Cryptography · Empirical

1 Introduction

An algorithm is a set of instructions that are executed sequentially. The instructions are followed by a set of rules. Algorithms are used for solving complex problems step by step. An algorithm helps to find out the time complexity, and the space complexity of a program. In cryptographic algorithms, many computational operations are used In MANET cryptographic algorithms are used to find to secure the communication between two nodes. One need to find how much time is taken for encryption and decryption of message while moving in the network so that the one with less complexity can be used.

A. Kumari (✉) · P. Pranav · S. Dutta
Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi, India
e-mail: ankitakmr33@gmail.com

P. Pranav
e-mail: prashantpranav19@gmail.com

S. Dutta
e-mail: sandipdutta@bitmesraac.in

S. Chakraborty
Department of Mathematics, Birla Institute of Technology, Mesra, Ranchi, India
e-mail: soubhikc@yahoo.co.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_9

It is extremely hard to develop a new design using composite cryptographic techniques without sound security analysis behind it, usually based on cryptographic reasoning. One way to achieve this is by employing cyber confidentiality.

Mathematics and number theory are inseparably linked to cryptography. As a result, it's impossible to come up with a new design that uses composite cryptographic approaches without first doing a thorough security study, which is usually based on cryptographic reasoning. Learning from others is one technique to achieve this goal. Analyzing current MANET/WSN security measures, as well as gaining a better understanding of the network to improve it. Learn how cryptographic approaches work in conjunction with MANETs to create a secure service. Network performance, scalability, storage, and synchronization should all be acceptable [1].

Cryptography is classified into two classes one is symmetric cryptography, and the other is asymmetric cryptography. Symmetric cryptography encompasses one key which is used for encryption as well as decryption. Asymmetric cryptography has two keys one is a public key another is private; one is used for encryption, and the other is used for decryption.

Algorithm evaluation is an important part of the computational complexity idea, which offers theoretical estimation for the specified sources of an algorithm to clear up particular computational trouble. Analysis of algorithms is the evaluation of the amount of time and space required to execute it.

An entire analysis of algorithms for running time of a set of rules includes the subsequent steps:

- Implement the set of regulations.
- Determine the time required for each primary operation.
- Identify unknown quantities that may be used to explain the frequency of execution of the number one operation.
- Analyzed the unknown portions, assuming the modeled input.
- Calculate the complete strolling time by multiplying the time utilizing the frequency for every operation, then adding all the products.

2 Literature Review

In [2] Author shows hybrid cryptographic algorithms for records protection in MANETs. They worked on two algorithms RSA and DSA algorithms. The generated trace documents and scripts the everyday overall performance of the proposed routing approach is evaluated and in comparison, with the traditional comfortable routing approach. The comparative typical overall performance of the proposed and traditional model is summarized. In their proposed technique the Energy is Low, Packet Delivery Ratio and Throughput are immoderate as have a study to Traditional technique.

The awareness [3] at the displacement between nodes. The set of rules now not most effective requests the blanketed nodes as regularly as possible, but the connections with dominator nodes are also stronger and greater solid. Thus, in dynamic

MANET surroundings, the algorithm can better find the appropriate dominator node to assemble the CDS. However, the algorithm does not offer a development to the electric power levels offering increases in radio range and the quantity of the included nodes. It remains viable to boom the power consumption quickly, leading to extra radio noise, and the battery lifetime for cellular gadgets stay a vital yet unaccounted for factor.

In [4] authors have presented a couple of constraint algorithms for multicast visitors engineering in MANET. The proposed algorithm is a new edition of more than one constraint QoS multicast routing optimization set of rules in MANET primarily based on GA (MQMGA). The proposed MQMGA optimizes the maximum hyperlink usage, the value of the multicast tree, the choice of the lengthy-life route, and the average delay and the most end-to-end-to-quiet put-off. The performance evaluation of our proposed techniques is completed thru modeling and simulation. The simulation effects exhibit that the proposed approach is an accurate and green technique for estimating and comparing the direction stability in dynamic mobile networks.

In [5] author proposed work to decrease the chance of network degradation provoked by the strength exhaust of node. The use of the system learning algorithm allows us to determine the ultimate route in phrases of header range; specifically, they showed some properties for acquiring the optical course in phrases of the header quantity which refers to the mobility of the node. An approximate expression for the greatest sum of header variety is realized. Finally, based totally on our findings, they proposed a method to discover and select the course that accounts for the anticipated statistics transfer time over the direction permits enhancing the throughput, packet delivery ratio, and overhead of reactive routing protocols.

In [6] delay optimization method in MANETs for multimedia transmission. The method maintains the sequence of in-order packets transmitted closer to the vacation spot which is received in a haphazard way to the buffer. Decrease the delay of packets with the aid of maximizing the in-order packets and minimizing the disorganized packets inside the buffer. A dynamic Knapsack set of rules is used to decrease packet loss due to the random motion of packets inside the buffer. Moreover, the method will increase the available space within the buffer by keeping a greater wide variety of in-order packets. Additionally, they proposed the mathematical relation of buffer size, packet size, and delay this is a critical metric for real-time multimedia applications.

In [7] proposed a delay-sensitive segment scheduling algorithm (DSSSA) for presenting well-timed P2P streaming services in MANETs. To use the wireless useful resource more correctly, DSSSA adopts the approach of scheduling the segments flippantly transmitted into a MANET based totally on the playback price of the well-timed P2P streaming services. The method can use the limited bandwidth of MANET greater correctly. On the alternative hand, considering that DSSSA is performed irregularly and the facts the use of for the section schedule is maintained by using the history data received from the previous real section transmission, DSSSA is adaptive to host mobility and peer interaction.

Statistical similarity among compounding procedures for RSA, a well-known cryptographic set of principles, is discussed in [8]. The two procedures used in encryption and decryption, namely electricity and modular mathematics, have been analyzed, and the results show that they are statistically distinct. However, whereas the modular arithmetic operation is predicted to be an even more dominating of the two, this is not the case in practice. In addition, the fundamental theorem of the finite distinction method was applied to empirically assess the running time of both processes separately. The empirical run time for power operation is O (n), but the empirical run time for the mod operation is O (n) (1).

In [9] comparing the overall performance of the LUC set of rules, Elgamal and RSA in the method of encrypting secret messages in textual content form have reached pretty a few conclusions. In phrases of the calculation manner, the LUC algorithm, Elgamal, and RSA have something in common, i.e. have a crucial generator within the structure of pinnacle numbers and the results of randomizing public and personal, and systematic keys. Based on the take a look, the quickest encryption system time is the RSA algorithm with 0.19 ms and 0.17 ms. In addition, the fastest decryption way is the LUC algorithm with 0.31 ms and 0.17 ms.

In [10] RSA, single encryption algorithms are used. ECC may be carried out in different ways. ECC uses mathematics algorithms as the primary goal operations for excessive-degree safety functions consisting of encryption for gaining confidentiality and a virtual signature for authentication. ECC can be carried out in software and hardware. ECC follows a usual process where parties agree on publicly-recognized information items, and each consumer generates their public and private keys.

In [11] paper offers relative analyses of AES, Twofish, RSA, and ElGamal cryptosystems. Predicated on the algorithms, new hybrid cryptosystems are Twofish RSA, AES RSA, and AES ElGamal. Memory consumption, translated report length, safety position, and encryption speed; those standards have been used to estimate the above-proposed algorithms and hybrid fashions. After exploration, among the surpassed new hybrid models, AES RSA takes all blessings from the use of symmetric and asymmetric systems, so it is considered; however, Twofish RSA hybrid cryptosystem is quick. For destiny work, proposed hybrid fashions can be analyzed through the entropy index. With entropy, studies will be feasible to estimate the resistance of each set of rules against one-of-a-kind sorts of assaults, ordinarily in opposition to cipher text frequency evaluation.

RSA [12] is a well-known public-key algorithm utilized by many applications. Security of RSA lies in the issue of factoring large numbers into prime factors. ElGamal Algorithm is a public key set of rules used for virtual signatures. Security of the ElGamal Algorithm lies in the difficulty of calculating the discrete logarithm. The proposed set of rules, a mixture of RSA and ElGamal Algorithm, has double protection; that's the problem of factoring massive numbers into prime factors and calculating the discrete logarithm. Based on the work that has been performed, the aggregate of RSA and ElGamal for the important thing technology set of rules has decreased the computing time required. But, the security component is not pretty confirmed. It may be proved via doing some piercing exams to aspect r and calculate discrete logarithms.

In this paper [13] they have virtualization of space in cloud computing for reduction of cost-efficient and making it more prominent for the research area. For securing the data in the cloud they first give the user authentication, they encrypt the stored data with the help of the Elgamal encryption technique then they added the fuzzy rules-based integer for adding more security to the data. They have implemented their proposed work in python and evaluated it with the help of Matlab. The performance result is good in terms of various aspects like security, execution time, and cost association as compared to the existing methodology.

3 Run Time Comparison of RSA and EL Gamal Protocols Used in MANET

The theoretical time complexity of RSA is based on message length and time taken for encryption and decryption of the message. For the encryption and decryption process, it will take different times so the complexity will also change. The time complexity for encryption will be calculated as $C = M^e$ mod n is O (log (N) ^{2}) this is for fixed message length. Where e is the same order as N then the complexity of encryption will change to O (log (N) ^{3}). For the decryption of the message, the complexity will change. $M = C^d$ mod n this formula is used for decryption of the message. The time complexity is O (log $(N)^3$).

El-Gamal algorithms are based on asymmetric cryptography in which two keys are used for encryption and decryption. The time complexity of El-Gamal is Big O which is written as O(n). The time complexity of algorithms depends on many factors like the length of the message, hardware of the system, and operating system. While algorithms are executing in the system it depends upon the number of operations it is performing. Length of the message increases then the execution time will increase.

4 Fundamental Theorem of Finite Difference

Let **f** be a real-valued function and let **a and b** be integers such that **a ≤ b**.

$$\text{If}\,F = \Delta - \text{if}\,\text{then} \sum \text{bn} = \text{af}(n) = F(b + 1) - F(a).$$

The fundamental theorem says that if the nth variation of an nth degree polynomial is constant and the higher differences are 0 then,

$$\Delta^n P_n(x) = C = \text{Constant}$$
$$\Delta^{n+r} P_n(x) = 0, r = 1, 2, 3 \ldots$$

Here Δ is the forward difference operator, state that Δ f(x) = f(x + h) − f(x) for the values of f(x) calculated as x = a, a + h, a + 2 h … h being the interval of differencing.

We analyzed the time complexity of RSA and El Gamal empirically using the fundamental theorem of finite difference. We run both algorithms 3 times and took the mean of the 3 trails as the final execution time. Then we calculated the first and the second difference in the mean execution time.

5 Result and Analysis

In this encryption and decryption table of RSA as in Tables 1 and 2, we have taken three trails for different message lengths. The time for different lengths of the message is different. After taking the number of trials time varies for encryption and decryption of the message. Then we have taken the mean of three different trails. Then take the first difference of mean was and then take the second difference. The result of the second difference is almost constant which shows Big O which means the worst case of the algorithm. The time complexity of RSA for encryption is $O_{emp}(n^2)$.

Table 1 Encryption time for RSA

Encryption time RSA						
Message bit	Trial 1	Trial 2	Trial 3	Mean	First difference (Δ x)	Second difference (Δ x^2)
8	0.000330	0.00050	0.00050	0.00044	0.00012	0.00032
16	0.000931	0.00040	0.00036	0.00056	0.00044	– 0.00025
32	0.000486	0.00130	0.00123	0.00101	0.00019	0.00036
64	0.000799	0.00095	0.00184	0.00120	0.00055	0.00229
128	0.002654	0.00141	0.00115	0.00174	0.00283	0.00918
256	0.000885	0.01129	0.00155	0.00457	0.01201	– 0.01201
512	0.003270	0.00321	0.04328	0.01659		

Table 2 Decryption time for RSA

Decryption time RSA						
Message bit	Trial 1	Trial 2	Trial 3	Mean	First difference (Δ x)	Second difference (Δ x^2)
8	0.00075	0.00039	0.00042	0.00052	0.00002	0.00074
16	0.00121	0.00022	0.00019	0.00054	0.00076	– 0.00060
32	0.00028	0.00184	0.00178	0.00130	0.00016	0.00007
64	0.00049	0.00039	0.00348	0.00145	0.00023	– 0.00013
128	0.00183	0.00175	0.00145	0.00168	0.00010	– 0.00002
256	0.00140	0.00209	0.00185	0.00178	0.00008	– 0.00008
512	0.00173	0.00143	0.00242	0.00186		

Table 3 Encryption time for EL GAMAI

Encryption time EL GAMAI						
message bit	Trial 1	Trial 2	Trial 3	Mean	First difference (Δ x)	Second difference (Δ x^2)
8	0.009293	0.007745	0.006064	0.0077007	0.00563	– 0.00594633
16	0.014014	0.015139	0.010836	0.0133297	– 0.00032	0.00530733
32	0.014075	0.012214	0.012748	0.0130123	0.00499	– 0.00330100
64	0.017437	0.017086	0.019484	0.0180023	0.00169	0.00151067
128	0.019169	0.019341	0.020564	0.0196913	0.00320	– 0.00024600
256	0.023099	0.023005	0.022569	0.0228910	0.00295	– 0.00295367
512	0.025083	0.026891	0.02556	0.0258447		

Table 4 Decryption time for EL GAMAI

Decryption time EL GAMAI						
Message bit	Trial 1	Trial 2	Trial 3	Mean	First difference (Δ x)	Second difference (Δ × 2)
8	0.004315	0.003949	0.005093	0.004452	0.00294	– 0.00267
16	0.007807	0.008239	0.006132	0.007393	0.00027	0.00288
32	0.00791	0.007559	0.007518	0.007662	0.00315	– 0.00202
64	0.010452	0.010262	0.011724	0.010813	0.00113	0.00095
128	0.01231	0.01174	0.011792	0.011947	0.00208	– 0.00008
256	0.015112	0.013585	0.013387	0.014028	0.00200	– 0.00200
512	0.015088	0.017087	0.015912	0.016029		

In El-Gamal algorithms as in Tables 3 and 4, we have taken three different trails for different message lengths for encryption and decryption. The Time taken for encryption and decryption of messages is different for different message lengths. After taken of three trials then taking the mean of three trials and then calculated the first difference and then taken second difference. The time variation is almost constant which shows that El-Gamal encryption and decryption time complexity is Big O which shows the worst case of the algorithm. The time complexity of the algorithm is $O_{emp}(1)$. The validated results are shown in the Figs. 1, 2 3 and 4.

6 Conclusions and Future Work

This paper analyzed that theoretical time complexity is different from real-time complexity. RSA and El-Gamal both algorithms are asymmetric cryptography.

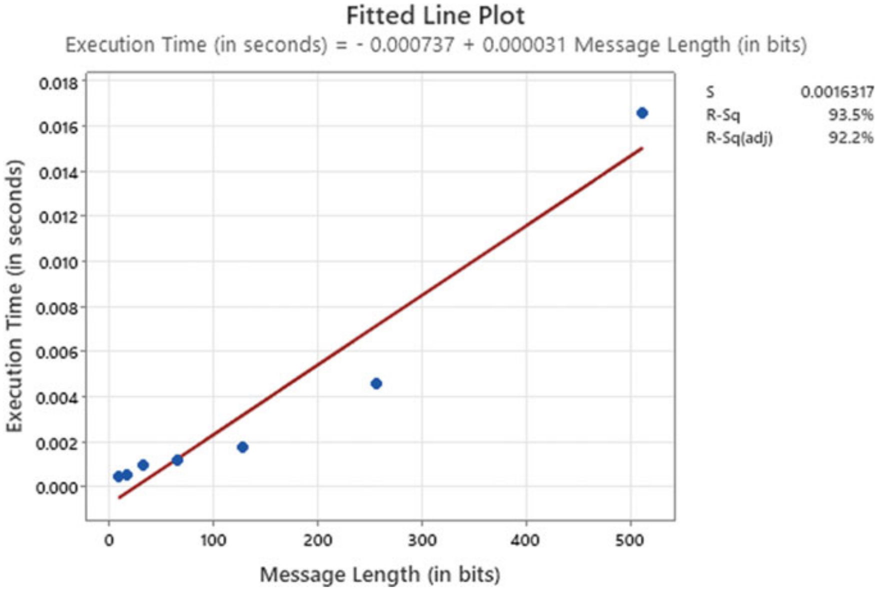


Fig. 1 Fitted line plot of encryption execution time of RSA

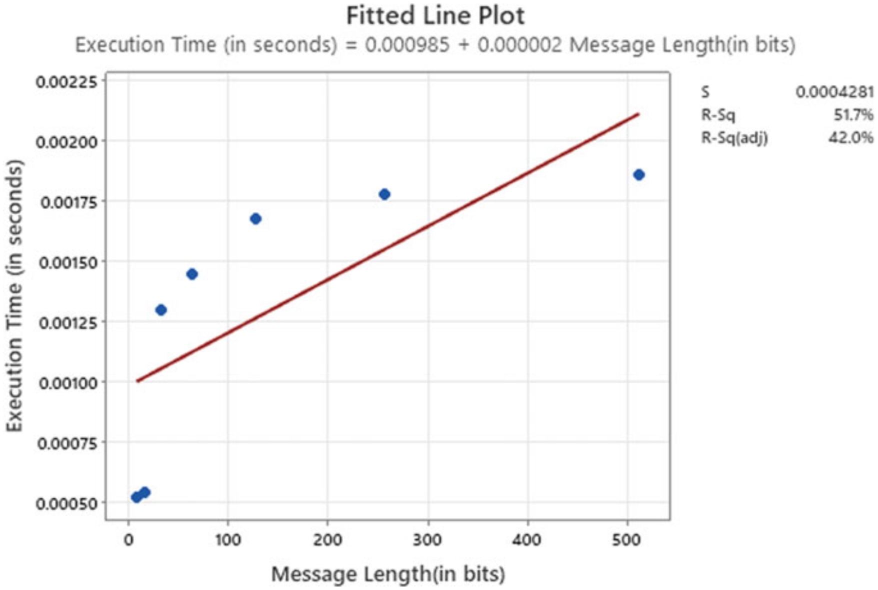


Fig. 2 Fitted line plot of decryption time of RSA

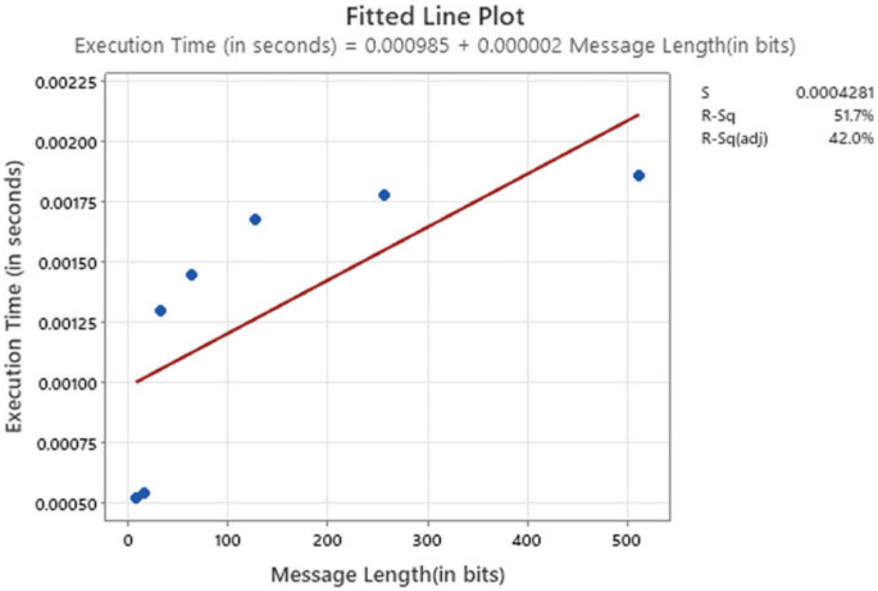


Fig. 3 Fitted line plot of encryption time is non-linear of EL-GAMAL

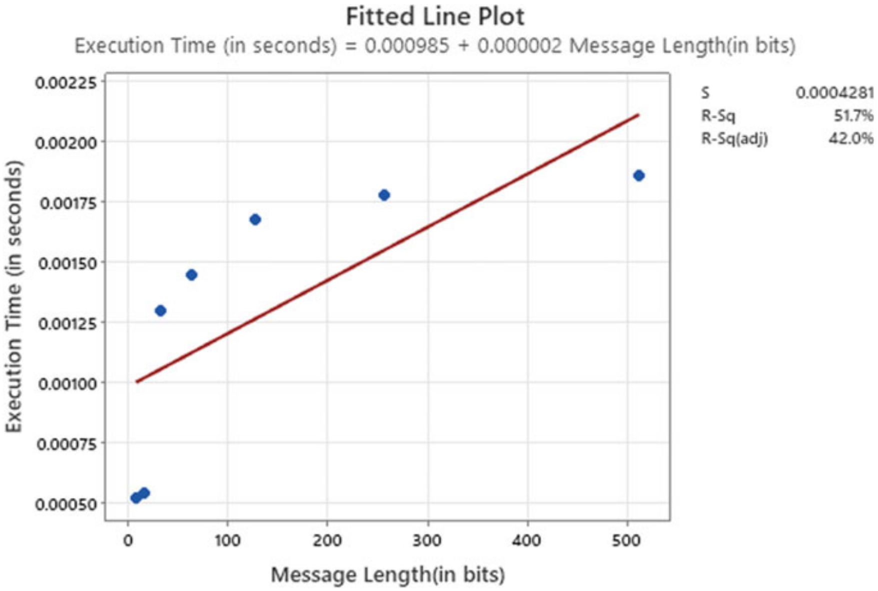


Fig. 4 Fitted line plot of decryption time of EL-GAMAL

Encryption and decryption time is different for different algorithms. The theoretical time complexity of RSA is O(n^3) and for El-Gamal is O(n) where n varies. According to the practical implementation of RSA and El-Gamal, the time complexity is different. By using the fundamental theorem of the finite difference method, we analyzed that time complexity is different. The empirical time complexity of RSA is $O_{emp}(n^2)$ and for El-Gamal is $O_{emp}(n^2)$. According to empirical analysis, the time complexity is different from theoretical.

References

1. Jianmin C, Wu J (2005) A survey on cryptography applied to secure mobile ad-hoc networks and wireless sensor networks. In: Proceedings of CRYPTO, pp 27–56
2. Sharma A, Bhuriya D, Singh U (2015) Secure data transmission on MANET by hybrid cryptography technique. In 2015 International conference on computer, communication, and control (IC4), pp 1–6
3. Leu S, Chang RS (2012) A weight-value algorithm for finding connected dominating sets in a MANET. J Netw Comput Appl 35(5):1615–1619
4. Sun B, Pi S, Gui C, Zeng Y, Yan B, Wang W, Qin Q (2008) Multiple constraints QoS multicast routing optimization algorithm in MANET based on GA. Prog Nat Sci 18(3):331–336
5. Arora SK, Monga H (2016) Performance evaluation of MANET based on knowledge base algorithm. Optik 127(18):7283–7291
6. Ahmad SJ, Reddy VSK, Damodaram A, Krishna PR (2015) Delay optimization using the Knapsack algorithm for multimedia traffic over MANETs. Expert Syst Appl 42(20):6819–6827
7. Hu CC, Lai CF, Hou JG, Huang YM (2017) Timely scheduling algorithm for P2P streaming over MANETs. Comput Netw 127:56–67
8. Pranav P, Dutta S, Chakraborty S (2021) Empirical and statistical comparison of intermediate steps of AES-128 and RSA in terms of time consumption
9. Sari PP, Nababan EB, Zarlis M (2020) Comparative study of Luc, ElGamal, and RSA algorithms in encoding texts. In: 2020 3rd international conference on mechanical, electronics, computer, and industrial technology (MECnIT). IEEE, pp. 148–151
10. Mallouli F, Hellal A, Saeed NS, Alzahrani FA (2019) A survey on cryptography: a comparative study between RSA vs ECC algorithms, and RSA vs El-Gamal algorithms. In: 2019 6th IEEE International conference on cyber security and cloud computing (CSCloud)/2019 5th IEEE international conference on edge computing and scalable cloud (EdgeCom). IEEE, pp 173–176
11. Jintcharadze E, Iavich M (2020) Hybrid implementation of Twofish, AES, ElGamal, and RSA cryptosystems. In: 2020 IEEE East-West design & test symposium (EWDTS). IEEE, pp 1–5
12. Iswari NMS (2016) Key generation algorithm design combination of RSA and ElGamal algorithm. In: 2016 8th International conference on information technology and electrical engineering (ICITEE). IEEE, pp 1–5
13. Pandian AP (2020) Development of secure cloud-based storage using the Elgamal hyper elliptic curve cryptography with fuzzy logic-based integer selection. J Soft Comput Paradigm 2(1):24–35

IoT Communication to Capture and Store Data to Thingspeak Cloud Using NodeMCU and Ultrasonic Sensor

Priya J. Payyappilly and Shweta Dour

Abstract Internet of Things is a domain which has gained quite a lot of momentum in the past decade. The ability of things communicaiting via a network or the internet is both a fascinating and a challenging concept. The authors in this study aim to introduce users to the concept of Internet of Things and demonstrate a use-case wherein data can be uploaded to the cloud; specifically, here the Thingspeak cloud is used as a storage for the data that is uploaded from an ultrasonic sensor via a microcontroller NodeMCU and WiFi standard. This usecase will help new users understand the scope and capability of communication in the Internet of Things. The exeprimental setup shows successful communication between nodeMCU and thingspeak cloud through the internet.

Keywords Internet · Internet of things · Thingspeak · Communication · Cloud · NodeMCU · Sensor · Ultrasonic sensor

1 Introduction

The Internet revolution [1] in the early nineties brought about a new approach for remote sharing and storing of data. A person practically in any part of the world could get access to this data via a worldwide network of interconnected computers forming a kind of sophisticated information superhighway. With the advancement and subsequent amalgamation of different technologies like integrated circuit technology, mobile technology and computing power, the Internet has developed magnificently in a short span of four decades and has become a part of routine life like a utility. The Internet initially was conceptualized to facilitate sharing of information by interaction between machines. This concept was further expanded by Ashton [2],

P. J. Payyappilly (✉) · S. Dour
Navrachana University, Vadodara, Gujarat, India
e-mail: priyaitce@gmail.com

S. Dour
e-mail: shwetad@nuv.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_10

in the late nineties, to include things into the Internet domain. He envisioned a world where physical and virtual 'things' [2] could be connected via Radio Frequency Identification (RFID) technology and make decisions on their own to accomplish certain tasks. He coined the term Internet of Things [2], abbreviated as IoT to distinguish this technology from traditional networking and internet. Since then, this concept has been taken up by various organizations for implementation and standardization. However, a shift from conventional networking and internet to IoT requires lot more efforts involving combining of a wide range of devices and technologies. It can be stated that IoT is still in an evolutionary stage, involving myriad levels of complexities and challenges. The authors in this article discuss the concept of IoT in depth and highlight the main aspects of this study.

On generalizing from the perspective of technology, IoT is a combination of the elements as shown in Fig. 1. It consists of hardware, software, network, cloud and analytics.

- *Hardware*—Hardware devices used for IoT implementation include microcontrollers and microprocessors like Arduino, Beagle Bone and Raspberry Pi. Microcontrollers and microprocessors act as processing elements of the system. These devices perform computation and execute programs to achieve some goal. These devices are interfaced with sensors and actuators. Sensors are instruments that measure certain physical quantities and represent them in digital form. Actuators are devices that manoeuver on getting a signal. The devices mentioned above can

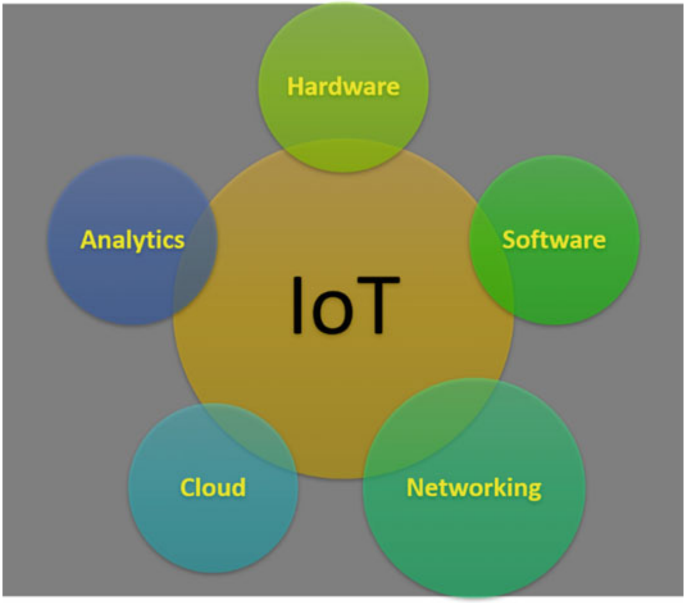


Fig. 1 IoT, a technological overview

be embedded into a larger system or can function individually depending on the application that they are used in. Sensors and actuators work in tandem according to the signal from processing elements.

- *Software*—It's the soul of the system, software encompasses the operating system, programs for performing operations according to input, the commands and also may include APIs for interfacing various technologies.
- *Network*—Network is the collection of technologies and protocols that facilitate connection between different devices. It includes wired and wireless connecting elements, protocols for enabling the connections like MQTT, HTTP, routing protocols, etc. It also includes the server which is a device for housing files.
- *Cloud*—Cloud is a collection of servers and networking resources offered to users on a pay as you use basis to enable ease of storage and access of data.
- *Analytics*—It is the scientific method of analyzing data that is either in transit on the network or stored on the cloud. It uses various algorithms entailing machine learning and deep learning to generate insights from available data. It is important to note here that all of these technologies cannot exist as solitary, there is a great degree of overlap among all of them. Also, there are other protocols and technologies that are used in assistance to the ones mentioned above but here, we are including only the primary technologies. Figure 1 is designed to show its general idea and simplicity from a technological perspective. IoT can also be depicted in the form of a layered approach. Layered approach is a way of representing any technology by dividing it into layers. Layers indicate demarcation points as well as modules of a large technology. Each layer performs a dedicated task and provides a set of services to its adjacent layers or higher layers. Generally, there are three major layers as shown in Fig. 2, the device layer, the network layer and the cloud. Each of these layers have their associated sub layers and perform functions that are explained below:

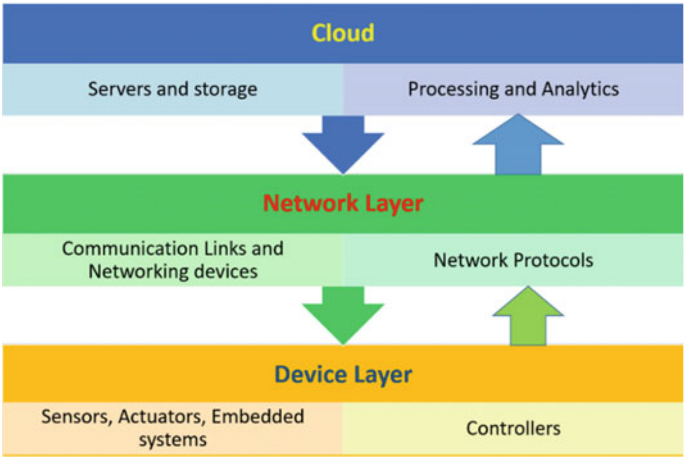


Fig. 2 Layered architecture of IoT

Device Layer: Consists of the devices that make up the implementation of IoT. This layer generally has sensors, actuators or a combination of both in the form of embedded systems and most importantly controllers. The term controllers is a broad classification. Controller can be broadly include microcontrollers and microprocessors.

Network Layer: Consist of the physical network between devices that are set up using various networking standards and protocols. It also consists of networking devices like routers and switches which facilitate the transmission of packets.

Cloud: The cloud is the term used to describe cloud computing. Cloud computing is the on-demand delivery of compute power, database, storage and infrastruce. It is a service that is provisioned by various companies on a pay by use policy. The users only pay for the resources they use. This practice greatly diminishes the requirement of configuring dedicated hardware and resources; saving money in the process. Cloud in IoT is used for one additional function apart from storing the data, it also analyzes the data to get useful insights. These insights play a major role in organizational decision making as the IoT solution will always be implemented by an organization seeking to service customers. As IoT is a confluence of various technologies, the authors through this study aim to: Introduce the readers to the concept of Internet of Things, explain the layered architecture of IoT, present the design and working of NodeMCU microcontroller and an ultrasonic sensor, familiarize the readers to the Thingspeak cloud platform and finally deploy an application to store the data obtained from the ultrasonic sensor to thingspeak via the internet.

2 Related Work

Various attempts at studying as well as implementing the idea of IoT using the concepts stated in the aforementioned sections are in the budding stage [3]. The primary focus of this literature survey by the authors is to explore different projects that are implemented using NodeMCU and Thingspeak. To summarize the contributions of various researchers in this domain, Ramdevi et al. [4] have implemented an indoor air quality and noise monitor using NodeMCU and Blynk application. NodeMCU is interfaced with DHT 11 which is a temperature and humidity sensor [5]. In addition to this, NodeMCU is also interfaced with MQ-3 sensor which is a gas sensor and SEN-12642, which is noise sensor. The cloud platform used to store data and obtain its visualization is the blynk platform [4]. Using the sensors and microcontroller, the system is able to sucessfully produce an outcome in the form of an alert having pollution and noise level values. Additionally, they add a buzzer as an alarm to notify the users apart from e-mail and message notifications. Juan et al. [6] develop a modular and economical energy management system for ameliorating the electricity usage caused by electric water heaters. The authors state that energy consumption due to electric water heaters is around 30c/o of the total energy consumption; therefore, it is imperative that some measures have to be taken to curb the energy usage so that it is lighter on the expenses. The complete system consists of NodeMCU microcontrollers connected to various electronic appliances and communicating with Raspberry Pi to provide status and energy usage of those appliances. The results show promising projections for annual savings using this energy management system. Ramaiah et al. [7] in their implementation have designed a tracking system for patient medication using NodeMCU interfaced with temperature sensor, RTC, LCD and servo motor. It notifies the patient at the time of taking their medicine. The box has a buzzer attached to it that persists ringing until the box is opened and it doesn't allow the box to be opened on a time other than the alloted time for medicine consumption. Jayaysingh et al. work along similar lines in their implementation of patient monitoring system [8] which monitors the vital signs of a patient using NodeMCU, pulse rate sensor, the cloud and smartphone. This is a very useful application of IoT which can redefine the healthcare realm and patient care solutions. Kharade et al. [9] implement a prototype for checking fire safety and air quality. The authors state that the rise of industrialized enviornment and emission of gases from different mechanized devices taint the atmosphere, so they execute a safety mechanism which has nodeMCU interfaced with flame sensor, DHT11 sensor and air quality sensor. These are interconnected to a cloud to complete the prototype of an IoT system. The cloudplatforms used here are blynk and Thingspeak. Blynk is used to alert the user while Thingspeak is used to store the data. Lufyagila et al. [10] use nodeMCU interfaced with various sensors effectively to monitor enviornmental conditions in poultry farms. The case study is specifically related to Tanzania, Africa, in which sample data has been taken from the farmers of Kilimanjaro and Arusha region. Conventional method of maintaining poultry is replaced by cost-effective IoT solution of enviornmental condition monitor. It has a NodeMCU connected with various electrical devices using relay. The real-time statistics are sent to a database server and raspberry Pi via a wireless access point. Raspberry Pi here is the gateway server which communicates to the end devices. The data is periodically synchronized to the end devices. If there is no connection to the internet, the data is stored on the local gateway and synchorized on availability of internet. The available information can then be visualized on the end devices for necessary action on the user's part. The results of this implementation show that the proposed system saves 84% time and saves 66.7% labour costs. Kaushik et al. [11] have implemented an IoT based system to monitor the power consumption in buildings; enabling the concept of smart buildings. The IoT solution detects human presence in a room using thermal sensor and Machine learning technique and SVM. Based on the detection, it switches the light on or off. It uses raspberry Pi and nodeMCU to compute and perform the necessary action with or without human presence. From the logged data, the reason for power consumption in a building can be known. The aforementioned realization of IoT solution helps in mitigating the useless expending of energy. Gupta et al. [12] have developed a testing system for monitoring and acquiring data out of a Photovoltaic cell in harsh enviornmental conditions. They have used Photovoltaic cell that is connected to four NodeMCU boards interfaced with different sensors like DHT22, dust sensor, wind sensor, pyranometer and INA 219 sensor. The software part of these devices is connected to the Thingspeak cloud and blynk application via the internet. There is a WiFi switch for enabling and disabling WiFi. This feature is added so that energy can be saved. The experimental results show that the measurements obtained by the aforementioned system is better than traditional methods and is resilient in challenging enviornmental conditions. A part of home automation system is presented by Garg et al. [13]. The authors propose a system that has a NodeMCU connected to different sensors and continuously monitors surrounding parameters like gas, temperature, humidity and light. Through IFTTT and Blynk application. Through the use of this system, real-time data generated at home can be monitored remotely from anywhere by internet connectivity. Vijaylakshmi et al. [14] in their work have developed an intelligent system to study different parameters that affect ease of learning in a classroom enviornment. Arduino microcontroller is interfaced with different sensors which measure sound levels, temperature, etc., in a room and show visualization on Thingspeak platform about the different parameters. The results are then displayed on LCD monitor. The visualization is helpful in getting to know the enviornment and help in increasing the comfort in the enviornment. Amala et al. [15] have devised a technology for continuously checking soil moisture content using Arduino which is then published to Thingspeak. They have classified the whole system into four layers, namely: sensor layer, Middleware, Communications layer and Cloud and application layer. The comparative analysis of each of the work presented above has been highlighted in Table 1.

As shown in Table 1, it can be inferred that a very few projects have used combination of NodeMCU and Thingspeak. Also the use of ultrasonic sensor has not been seen in amy of the aforementioned articles. In line with the existing literature, the

Table 1 Comparative analysis of communication methodologies in the literature survey

Article	Project	Technology
[4]	Indoor air quality and noise monitor	nodeMCU and Blynk, DHT11, MQ3
[6]	Electric energy tracker	NodeMCU and Raspberry Pi
[7]	Patient's medicine tracking system	NodeMCU, RTC, LCD, servo motor
[8]	Patient's vital signs monitor	NodeMCU, cloud, pulse rate sensor and smartphone
[9]	Air quality monitoring	NodeMCU, blynk and Thingspeak
[10]	Enviornmental conditions monitoring in poultry farms	NodeMCU, Raspberry Pi
[11]	Smart Buildings	NodeMCU, Raspberry Pi, thermal sensor
[12]	Procuring data from a photovoltaic cell	NodeMCU, Thingspeak, Blynk, photovoltaic cell
[13]	Home automation	NodeMCU, Blynk, IFTTT, gas, temperature, humidity and light sensors
[14]	Learning comfort in classroom	Arduino, Thingspeak
[15]	Soil moisture sensing	NodeMCU and Thingspeak, moisture sensor

present article focuses on establishing connection between the cloud and a device. The device here is a combination of microcontroller NodeMCU and ultrasonic sensor.

3 Proposed Work

The setup consists of NodeMCU interfaced with ultrasonic sensor which in turn is connected to the Thingspeak cloud via the internet. If we try to explain it in terms of the layered architecture of IoT as shown in Fig. 6. The mapping of various layers for this experimental setup can be according to the one shown in Fig. 3. The device layer is the lowest layer which consist of the apparatus making up the foundation of IoT technology. It consists primarily of the microcontroller NodeMCU and ultrasonic sensor.

NodeMCU is an open source platform that is executed in conjunction with the ESP8266 WiFi SoC [16]. It can be programmed using the Lua scripting language although, the programming of it using Arduino IDE is widespread due to the developer communities efforts. An ultrasonic sensor which has two pins namely trigger pin and echo pin is also used. Ultrasonic sensor calculates distance of an object based on the reflected signal that is sent. This makes up the hardware part of the system. NodeMCU and ultrasonic sensor are as shown in Figs. 4 and 5. NodeMCU has been chosen as it is compatible, easy to interface and economical for prototyping.

The network layer consists of the communication technologies which broadly speaking is the internet and WiFi. Thingspeak is used as the platform for cloud. It stores the data in a unit called channel. A channel is the basic building block of

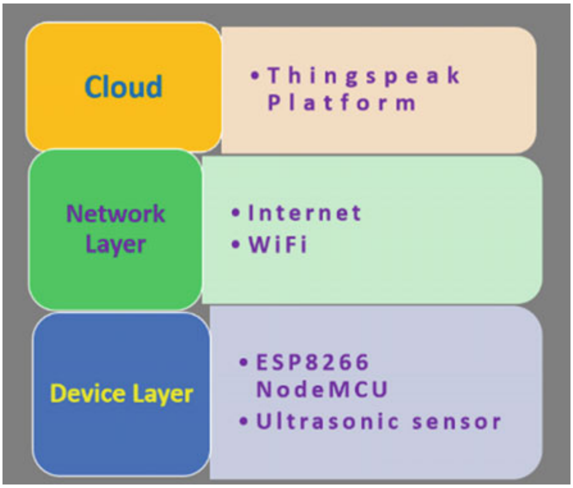


Fig. 3 Layered architecture of IoT mapping with experimental setup

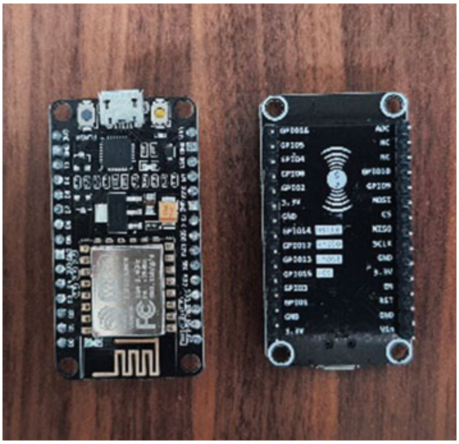


Fig. 4 NodeMCU ESP8266

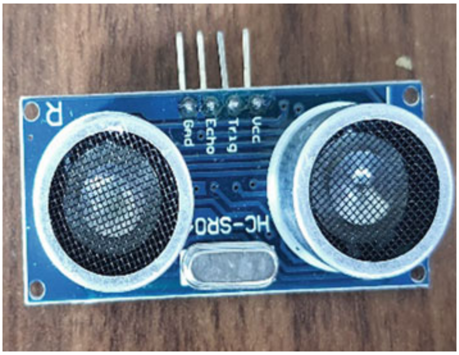


Fig. 5 Ultrasonic sensor

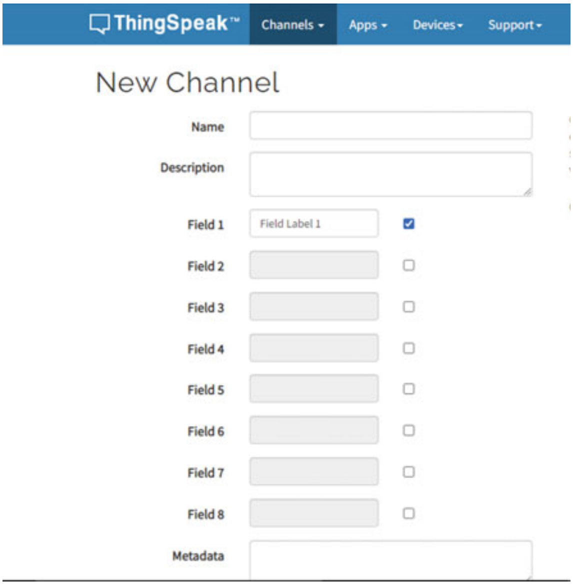
Thingspeak. It consists of various attributes which a user can update to store relevant data.

Initial Configuration of Thingspeak Channel

Example of a Thingspeak channel is shown in Fig. 6 and 7. The attributes that are to be updated need to be activated first. Then they can be named and values for them can be entered.

In the attributes of channel contain field numbers ranging from one to eight. The field is analogous to variable names in a program. As variables are named location of memory, similarly, field name is the named tag given for attribute storage. User can create upto 8 fields. Before naming a field, the user is supposed to activate it by selecting checkbox on the right side of the field. Only the activated field will store the value for any attribute.

Fig. 6 Thingspeak ınitial configuration: attributes for code



New Channel

Name

Description

Field 1 Field Label 1 ☑

Field 2 ☐

Field 3 ☐

Field 4 ☐

Field 5 ☐

Field 6 ☐

Field 7 ☐

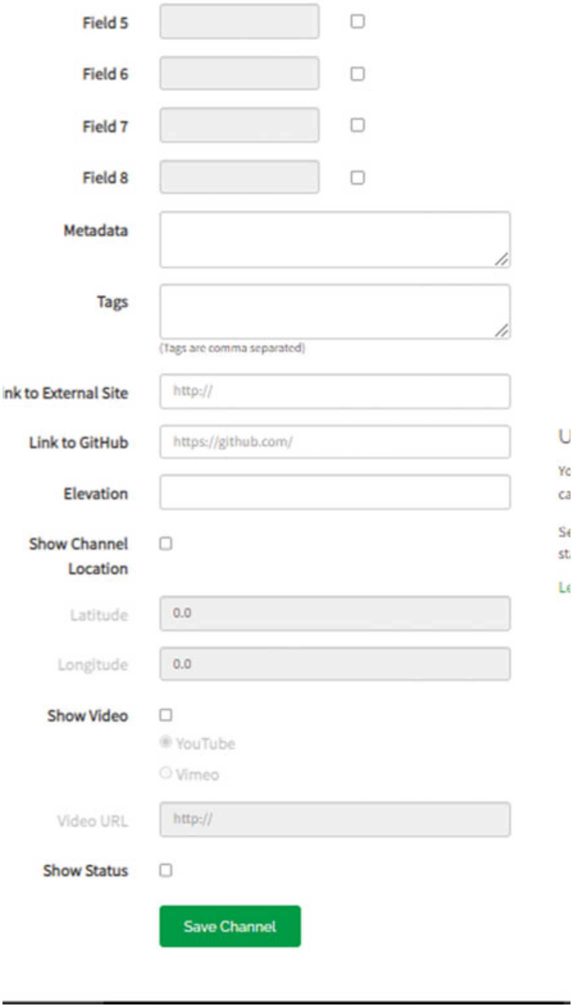
Field 8 ☐

Metadata

Initially, all attributes for storing data are not selected in the thingspeak channel so they can be seen as inactivated. For the proposed method, since only the distance has to be stored, a single field named distance has been created as shown in Fig. 8. The checkbox on the right has to be selected, at that time, the attribute will be activated. On entering the values and saving, the channel statistics is as shown in Fig. 8 is displayed. It contains channel ID which is a unique identifier of the channel. The name field is the identifier of the channel from the user perspective. Description field is analogus to comments in C program. The programmer can add a description to depict information about the channel for understanding what the use of the channel is. After that comes field 1 which is the variable holding value of distance of object from ultrasonic sensor. Since this is a simple experiment, a single field is only used to represent value of distance.

Figure 9 shows the experimental setup where nodeMCU is interfaced with ultrasonic sensor.

The nodeMCU and ultrasonic sensors are in their default configuration and no changes are done to their configuration. On connection with the computer system, the presets are selected automatically when nodeMCU board is selected in the Arduino IDE. The ultrasonic sensor has been selected because of its ease of interfacing with the NodeMCU board. The rationale for choosing NodeMCU for controller is that it is economical to purchase, is open source and can be used in conjunction with Arduino IDE easily.



Field 5 □

Field 6 □

Field 7 □

Field 8 □

Metadata

Tags

(Tags are comma separated)

nk to External Site http://

Link to GitHub https://github.com/

Elevation

Show Channel Location □

Latitude 0.0

Longitude 0.0

Show Video □

● YouTube

○ Vimeo

Video URL http://

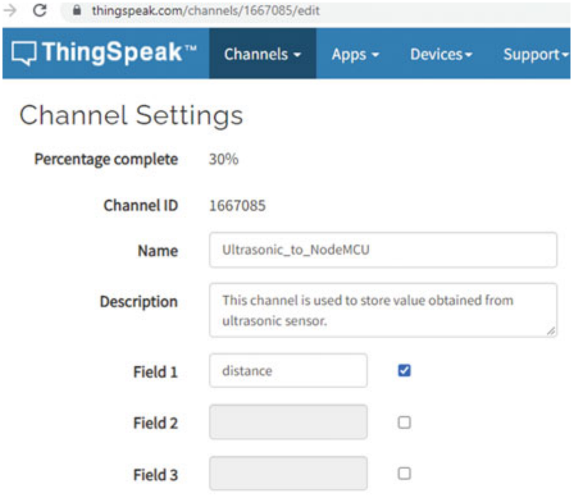
Show Status □

Save Channel

Fig. 7 Thingspeak ınitial configuration: other fields

The code of the proposed setup can be summarized through the flowchart shown in Fig. 10. The variables and hardware are initialized in the begining. Then an attempt to connect to WiFi is done. If it is sucessful, the loop function in Arduino IDE (Integrated Development Enviornment) executes continuously until there is some error or the connection no longer exists. The output is continuously printed on the serial monitor and is uploaded to Thingspeak every 15 s. The system continuously checks for broken communication or un-initialized variables. If any of the conditions are not fulfilled, exit message is printed on the serial monitor and the process is ended there. It is shown by terminator labelled End in the flowchart.

Fig. 8 Thingspeak configuration for IoT communication



Channel Settings

Percentage complete 30%

Channel ID 1667085

Name Ultrasonic_to_NodeMCU

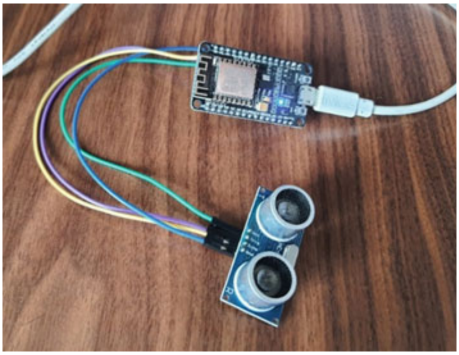
Description This channel is used to store value obtained from ultrasonic sensor.

Field 1 distance ☑

Field 2 □

Field 3 □

Fig. 9 Experimental setup



The implementation in the form of code for this application is shown in Figs. 11 and 12. The system and its components are initialized by declaring variables and pins as shown in Fig. 11. Initially, the header files are included, then the constants are declared in the global declaration section. The calculation parameters and its formulae are declared thereafter.

The rest of the code for communication with Thingspeak as well as displaying the values on serial monitor is given in Fig. 12. There are two main functions in Arduino program viz setup and loop. Setup function consists of all the functions that facilitate setup of a system. Here, as shown in Fig. 12, configuration with serial monitor, Thingspeak cloud, WiFi connection and configuration of pins on nodeMCU is done. Once these are successful, the loop function is used to continuously read values from ultrasonic sensor and upload to cloud. The aforementioned points are demonstrated in Figs. 11 and 12.

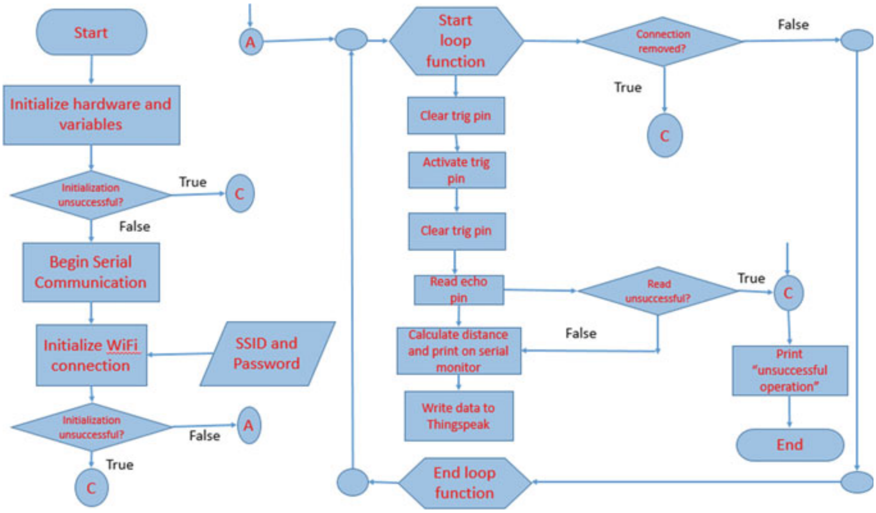
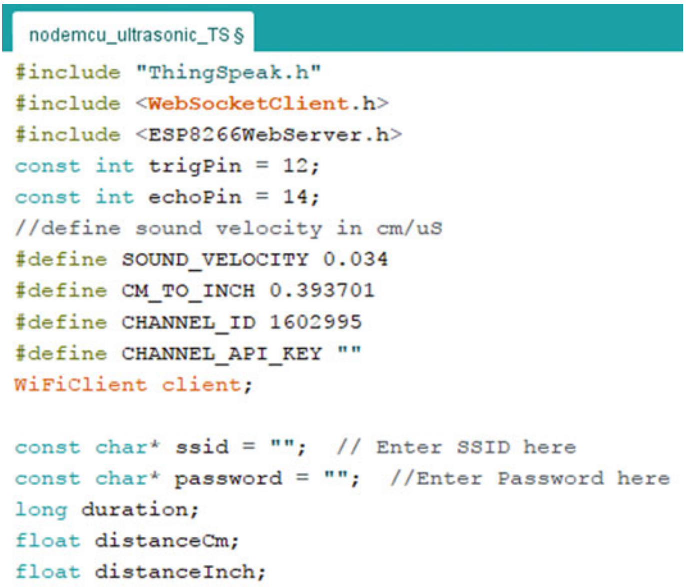
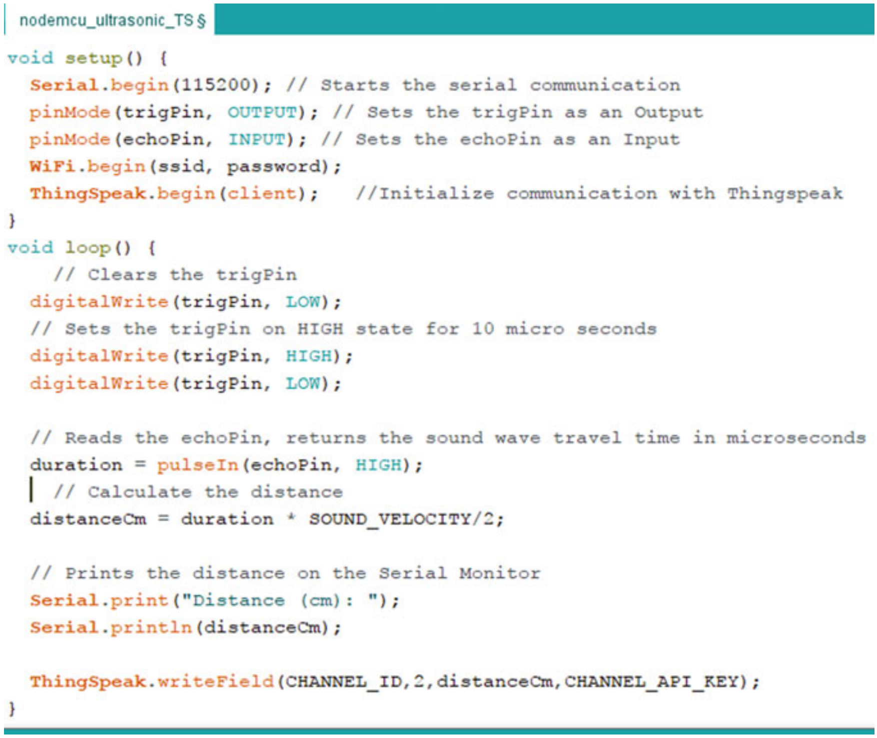


Fig. 10 Flowchart for the proposed method implementation



```
#include "ThingSpeak.h"
#include <WebSocketClient.h>
#include <ESP8266WebServer.h>
const int trigPin = 12;
const int echoPin = 14;
//define sound velocity in cm/uS
#define SOUND_VELOCITY 0.034
#define CM_TO_INCH 0.393701
#define CHANNEL_ID 1602995
#define CHANNEL_API_KEY ""
WiFiClient client;

const char* ssid = "";  // Enter SSID here
const char* password = "";  //Enter Password here
long duration;
float distanceCm;
float distanceInch;
```

Fig. 11 Code for IoT arduino IDE communication in with thingspeak: ınitialization



```
void setup() {
  Serial.begin(115200); // Starts the serial communication
  pinMode(trigPin, OUTPUT); // Sets the trigPin as an Output
  pinMode(echoPin, INPUT); // Sets the echoPin as an Input
  WiFi.begin(ssid, password);
  ThingSpeak.begin(client);   //Initialize communication with Thingspeak
}
void loop() {
    // Clears the trigPin
  digitalWrite(trigPin, LOW);
  // Sets the trigPin on HIGH state for 10 micro seconds
  digitalWrite(trigPin, HIGH);
  digitalWrite(trigPin, LOW);

  // Reads the echoPin, returns the sound wave travel time in microseconds
  duration = pulseIn(echoPin, HIGH);
    // Calculate the distance
  distanceCm = duration * SOUND_VELOCITY/2;

  // Prints the distance on the Serial Monitor
  Serial.print("Distance (cm): ");
  Serial.println(distanceCm);

  ThingSpeak.writeField(CHANNEL_ID,2,distanceCm,CHANNEL_API_KEY);
}
```

Fig. 12 Code for IoT arduino IDE communication in with thingspeak: setup and loop function

4 Results and Discussion

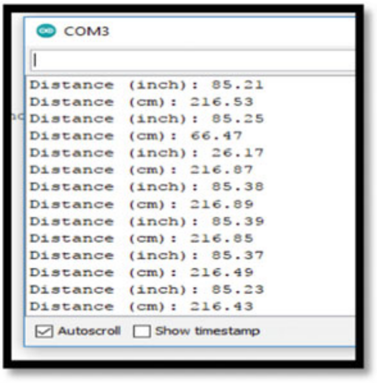
Figures 13 and 14 show the results on execution of the program. Figure 13 shows the serial monitor in Arduino on which the values of distance of an object from centimeter to inches is displayed.

The visualization shown in Fig. 14 is the line graph of the values that are obtained while the sensor is active and communicating to the microcontroller.

Suitability of the proposed model in experiment with other sensors and application wise necessity

The proposed model is suitable to be implemented with most of the other sensors like DHT sensor, gas sensor, touch sensor etc.; since nodeMCU is an open source technology, it can be easily interfaced with other sensors. If the other sensors are from a different vendor, the related header file for that sensor is to be downloaded in the Arduino IDE and it code can be easily uploaded without much hassle. For instance, if nodeMCU is interfaced with DHT sensor, only DHT header library has to be downloaded and included in the current program. Once this process is done, the program executes smoothly.

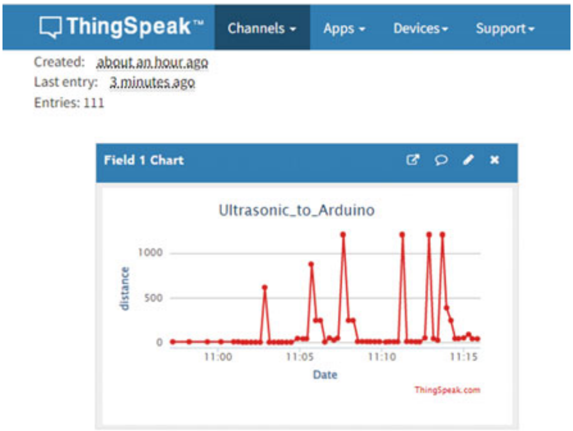
Fig. 13 Output on arduino monitor



```
Distance (inch): 85.21
Distance (cm): 216.53
Distance (inch): 85.25
Distance (cm): 66.47
Distance (inch): 26.17
Distance (cm): 216.87
Distance (inch): 85.38
Distance (cm): 216.89
Distance (inch): 85.39
Distance (cm): 216.85
Distance (inch): 85.37
Distance (cm): 216.49
Distance (inch): 85.23
Distance (cm): 216.43
```

☑ Autoscroll □ Show timestamp

Fig. 14 Visualization on thingspeak



Created: about an hour ago
Last entry: 3 minutes ago
Entries: 111

Ultrasonic sensors can be used in various applications like object detection, distance measurement and liquid level sensing, the application suggested in the present article can be used in conjunction with the aforementioned uses to yield better results and insights for the data that is produced.

5 Conclusion and Future Scope

Internet revolution has unfolded new avenues for device communication on the network. This study attempts to familiarize the reader with the concept of internet of things and elicit an example of the same. The example is shown in terms of communication between an ultrasonic sensor, nodeMCU and Thingspeak cloud. The experimental setup and results show that data is sucessfully uploaded to the cloud using API keys avalable on Thingspeak channel. However, the delay in uploading of data on the cloud platform can be a challenge when real time data has to be given to the cloud and if any decisions have to be made using that real-time data. Also a limitation is security of the devices as well as the communication process. There is ample scope for exploration in the area of security of devices and communication. Futhermore, this study can be extended to include notification through IFTTT if any event occurs and also generating insights from the uploaded data through MATLAB on thingspeak. The study can also incorporate different sensors along with nodeMCU in different applications to upload data to Thingspeak.

References

1. Kleinrock L (2010) An early history of the internet [History of Communications]. IEEE Commun Mag 48(8):26–36
2. Goyal P, Sahoo AK, Sharma TK (2021) Internet of things: architecture and enabling technologies. Mater Today: Proc 34:719–735
3. Smys S, Raj JS (2019) Internet of things and big data analytics for health care with cloud computing. J Inf Technol 1(1):9–18
4. Ramdevi M, Gujjula R, Ranjith M, Sneha S (2021) IoT evaluating ındoor environmental quality check of air and noise. Mater Today: Proc. ISSN 2214-7853
5. Duraipandian M, Vinothkanna R (2019) Cloud based internet of things for smart connected objects. J ISMAC 1(02):111–119
6. Tejero-Gómez JA, Bayod-Rújula AA (2021) Energy management system design oriented for energy cost optimization in electric water heaters. Energ Buildings 243(111012). ISSN 0378-7788
7. Challa R, Yamparala R, Kanumallı SS, Kumar KS (2020) Advanced patient's medication monitoring system with ardunio UNO and NODEMCU. In: 2020 4th International conference on electronics, communication and aerospace technology (ICECA), pp 942–945. https://doi.org/10.1109/ICECA49313.2020.9297420
8. Jayaysingh R, David J, Raaj MJM, Daniel D, BlessyTelagathoti D (2020) IoT based patient monitoring system using nodeMCU. In: 2020 5th International conference on devices, circuits and systems (ICDCS), pp 240–243. https://doi.org/10.1109/ICDCS48716.2020.243588
9. Kharade M, Katangle S, Kale GM, Deosarkar SB, Nalbalwar SL (2020) A nodeMCU based fire safety and air quality monitoring device. In: 2020 International conference for emerging technology (INCET) 2020, pp 1-4. https://doi.org/10.1109/INCET49848.2020.9153983
10. Lufyagila B, Machuve D, Clemen T (2021) IoT-powered system for environmental conditions monitoring in poultry house: a case of Tanzania. Afr J Sci Technol Innov Dev. https://doi.org/10.1080/20421338.2021.1924348
11. Kaushik S, Srinivasan K, Sharmila B, Devasena D, Suresh M, Hitesh Panchal R, Ashokkumar KK, Sadasivuni & Neel Srimali, (2021) Continuous monitoring of power consumption in urban buildings based on internet of things. Int J Ambient Energy. https://doi.org/10.1080/01430750.2021.1931961
12. Gupta V, Sharma M, Pachauri RK, Dinesh Babu KN (2021) A low-cost real-time IoT enabled data acquisition system for monitoring of PV system. Energy Sources Part A: Recovery Utilization Environ Eff 43(20):2529–2543. https://doi.org/10.1080/15567036.2020.1844351
13. Garg S, Yadav A, Jamloki S, Sadana A, Tharani K (2020) IoT based home automation. J Inf Optim Sci 41(1):261–271. https://doi.org/10.1080/02522667.2020.1721581
14. Vijayalakshmi R, Jayasimman L (2021) The design of IoT based smart learning environment for learner's comfort level monitoring. Mater Today: Proc. ISSN 2214-7853

15. Praveen AAA, Tamilnesan P, Muthukumaran M, Udayakumar MD (2021) Experimental analysis of moisture content with involuntary irrigation structure in soil, Mater Today: Proc 45, Part 2, 1893–1897. ISSN 2214-7853
16. Nooruddin S, Islam MM, Sharna FA (2020) An IoT based device-type invariant fall detection system, Internet of Things 9(100130). ISSN 2542-6605

A Comprehensive Study on Cloud Computing: Architecture, Load Balancing, Task Scheduling and Meta-Heuristic Optimization

Shruti Tiwari and Chinmay Bhatt

Abstract Cloud computing (CC) is evolving computing model with a vast array of heterogeneous autonomous systems by modular computational architecture. Load balancing of activities on the cloud environment is an essential part of distributing services from the data center. CC is agonized by overloading demands because of dynamic computing through the internet. Load balancing must be done to ensure maximum use of the resources in all virtual machines (VM). Task scheduling is a crucial step for improving cloud computing's overall efficacy. Task scheduling is therefore significant to minimize energy usage and increase service providers' benefit by reducing the time required. This work provides a detailed study about the cloud computing architecture, load balancing (LB) mechanism, task scheduling (TS) framework in the cloud environment. Various meta-heuristic optimization techniques have been implemented to manage the load over virtual machines using task scheduling and load balancing terminologies. Various research gaps and issues have been identified from the literary work done by various researchers. This comprehensive study has motivated and provided us future direction to do work in this field.

Keywords Cloud computing · Virtual machines · Task scheduling · Load balancing · Meta-heuristic optimization

1 Introduction

Cloud Computing [1, 2] is an evolving software deployment and maintenance trend that is being embraced by industries like Microsoft, IBM, Google, and eBay. IBM-Blue Cloud framework, eBay Cloud, Google App Engine, as well as Distributed Computing Platform are many conceptual systems and frameworks.

S. Tiwari (✉) · C. Bhatt
Department of Computer Science Engineering, RKDF College, Bhopal, (M.P.), India
e-mail: shruti.tiwari08@gmail.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_11

Cloud Computing is seen as the upcoming development which would influence organizational organizations and also how they handle their IT infrastructure and services. A key researching area is infrastructure and technology which cloud services and deployed models have presented.

The Internet has always been a driving factor in the advancement of different technologies. Cloud Computing is arguably the most debated amongst all these. The cloud computing model has seen an immense shift toward this usage over the last several months and has become a standard in the IT space as it offers its users and suppliers substantial cost savings including new business potential [3]. The benefits that use cloud computing are including:

(1) Minimization in the maintenance cost and its hardware cost.
(2) Accessible to all.
(3) Flexible including automatized procedures in which the client does not have to think about complex problems such as software up-gradation [4, 5].

Although the concept of cloud computing has various variations, this new computing model is defined by certain fundamental concepts. Cloud Computing offers technical resources that are provided on-demand as-a-service through the use of the Web, typically managed across premises. Big Evolution in the field of computer science, a service. CC can efficiently and securely deliver various IT services and resources according to the requirement of the user. Cloud computing offers usage-driven applications. It provides services like Infrastructure as a Services (IaaS), Platform as a Services (PaaS), and Software as a Services (SaaS) [6]. Clients of such resources don't own infrastructure in the remote cloud, yet pay for resources on the per usage, provided that a 3rd party operates and maintains the public cloud. The main principle, therefore, is the virtualization of resources. They rent the physical infrastructure, systems, and software inside a shared structure in a realistic situation. From virtual networks, computing systems, integrated data center, end-user web apps, and web services to immense computing-oriented service, security features will differ.

In many fields of ITs, cloud computing can be used to resolve issues such as GIS, Science Research, Decision Making, ERP, e-Governance Systems [7], Web App Creation, Mobile Technologies, etcetera. As the main back-end computing infrastructure, cloud computing depends on data centers [8]. With the need for cloud computing more dramatically in recent years, geographically dispersed data centers are offering more and more cloud resources to improve the reliability and consistency of services. A huge amount of energy is essential to operate these geographically dispersed data centers, which represent around 15% of the overall cost of a data center (DC) [9].

However, considering data centers as supercomputers with shared resources or making the second stage simple, consistent with servers, new approaches to geographical load-balancing (GLB) as in Fig. 1 [10] mostly concentrate on the first phase, i.e. the proportion of work demands that are assigned to each data center, of employment planning. Moreover, the existing GLB approaches often rely on a specific resource dimension (for example, CPU) and consistent resource need,

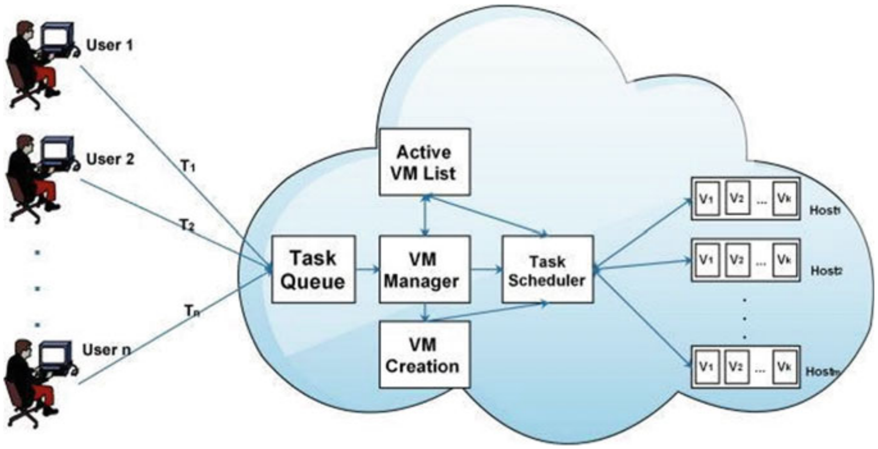


Fig. 1 Cloud load balancing

without taking into account the requirements of many separate jobs and heterogeneous resources. This is significant, in particular, because modern data centers are usually built from a variety of server groups with various processing capacity, size of memory, and storage spaces specifications [11]. The diversity of the market profile for workloads for different occupations exacerbates such heterogeneity further, e.g. CPU intensive numerical computing activities are generally required while high memory support is normally required for database operations [12].

Load must be spread among the cloud-based networks such that there must be no over-utilized or underutilized nodes within the networking process. LB is a common cloud problem that makes it difficult for applications adjacent to QoS (quality of service) measurement to sustain performance according to the SLA (service level agreement) document essential by enterprise cloud providers. Cloud providers have difficulty distributing similar workloads through servers. Effective LB technology can maximize and guarantee high user satisfaction by effectively using VMs resources [13].

In the cloud environment, a load must be balanced through various techniques. Algorithms of LB may be split into Algorithms for dynamic LB and static LB. Load is spread by previous knowledge and information in the Static LB Algorithm. Static algorithms at the same time do not take account of the current workload [14]. These algorithms are for the less work-loaded cloud. Dynamic LB Algorithms take into account current work Cloud load. Centralized and Semi-distributed LB algorithms can be categorized into dynamic LB algorithms [15]. Intelligent strategies like GA (Genetic Algorithms), PSO (Particle Swarm Optimization), ACS (Ant Colony System), and ABC (Artificial Bee Colony) can address problems of load balancing [16]. Various analysis was undertaken in the area of LB and TS in the cloud environments.

The task scheduling [17], known for providing essential cloud service efficiency, currently represents a related hot topic. However, due to inappropriate scheduling, the dilemmas of resources that are underused (unloaded) and overused (overloaded), may emerge, leading respectively to either wastage of cloud resources or a decrease in services. So, the idea to incorporate meta-heuristic algorithms into TS has arisen so that complicated and varied incoming tasks (cloudlets) can be easily distributed within a reasonable time to available resources. The meta-heuristic techniques have shown themselves to be very able to solve scheduling problems, which are met by giving a detailed overview of traditional and heuristic approaches before deeply expanding into common cloud task meta-heuristic methods, then comprehensive systematic review, which includes new taxonomy and advantage of those methods.

More specifically, major contribution keys of this study may be organized as follows:

(1) Comprehensive study about cloud computing evolution and architecture
(2) Studying the mechanisms of existing load balancing.
(3) Providing an overview about task scheduling in cloud computing
(4) Identify important areas where new research can be performed with the optimization principle to better load balancing algorithms.
(5) A systematic analysis is presented on cloud meta-heuristic TS.
(6) Identifying research gaps and issues exist that can be further hurdle to researchers for LB algorithms

The below is paper systematized: Section 2 elaborates the cloud computing concept with their architecture. It also consists of data center networking for VMs. At last, it talks about motivation of study. Section 3 elaborates the review of load balancing techniques, and task scheduling in CC in Sect. 4. Metaheuristic optimization algorithm comparison shows in Sect. 5. Section 6 explains related work also identifies some research gaps and issues and Sect. 7 concludes this work.

2 Cloud Computing: Architecture

Web-based tools and technologies can be found in cloud computing resources. This allows the users to work remotely because the cloud can be used as “Internet”. Therefore, it is not processed as traditional outsourcing. It is also called Massive Computing. In this, the allocation of applications must be dynamic. No hardware or software has to be installed. Cloud computing aims to enable people who have no deep knowledge of information about all technology and applications [18].

2.1 Overview

"Cloud" is a virtualized reusable resource computing pool. It can change or handle a variety of workloads. The "Cloud' term is a tarn of enormous infrastructure possessions, offering various utilities and hardware implementations as well as device software to end-users, alternatively enabling end-users to access their computing needs inconsistent way. The user needs no information about where to discover the device requirement and how the cloud operates. Cloud providers manage all obtainable tools to consumers where 'computing' is defined on an 'SLA'. Cloud computing offers inevitable benefit in sharing over the internet of many configurable system properties and higher-level administrations that may be built with very little board effort. Cloud computing has been developed to promote the use of virtualization technology to allow end-users to use virtual services without infrastructure at a reasonable cost [19]. Cloud Computing Evolution in IT in shown in the Fig. 2.

Now we are describing different modes of cloud computing as follows [21]:

(1) **Public Cloud**: They are managed by CSP which owns facilities and data centers. Infrastructure is located on-site and companies may use pay-as-you-go and on-demand services.
(2) **Private Cloud**: It is only created and operated by certain enterprises however 3rd party companies have control on behalf of the cloud owner to handle it.
(3) **Hybrid Cloud**: It combines an only selection of all kinds of cloud deployments, such as public, private, or community cloud. Core operations are conducted in a private cloud, while a public cloud offers fewer basic resources.
(4) **Community Cloud**: Several organizations or institutions of a shared purpose are sharing the community cloud. Universities that do this for learning and research are typical examples.

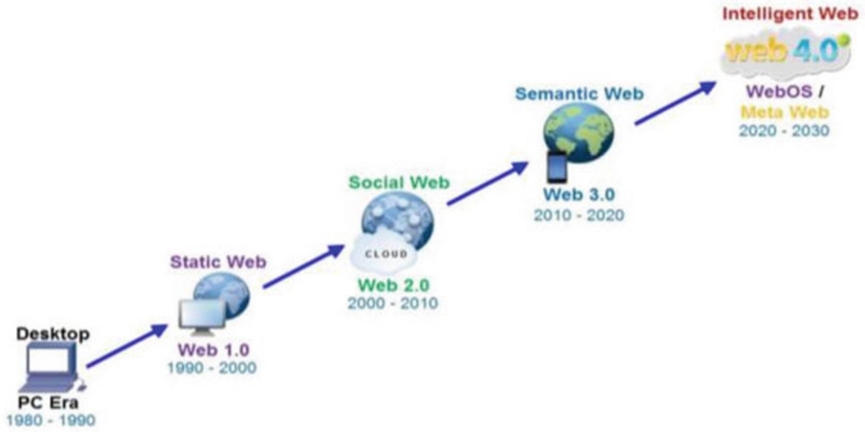


Fig. 2 Cloud computing evolution in IT [20]

2.2 Features of Cloud Computing

Compared with other computing paradigms, CC provides a range of new features [22] and advantages. This section is briefly identified.

- **Scalability and On-Demand Services**: Cloud computing connects customers on request with resources and services. Resources may be scaled across several data centers.
- **QoS (Quality of Service)**: In terms of CPU performance or hardware, memory capacity, and bandwidth, CC may provide QoS for users.
- **User-Centric Interface**: Cloud interfaces are locational isolated and well-defined interfaces including web services and web browsers enable them to be accessed.
- **Autonomous System**: CC applications are user-friendly, transparently controlled decentralized systems. But, software and data inside clouds may be reconfigured and merged automatically into a basic, user-specific platform.
- **Pricing**: Cloud computing requires no investment at the start of the project. There is no need for capital expenses. Users can pay for services and capacities or pay for them, as they require them.

2.3 Architecture

The applications, data, and utilities are all maintained in the cloud over the Internet and run applications, as well as stored data through the provision of software resources on-demand services in CC architecture. The back end and front end of a cloud architecture can be separated. The front end is available to the user by internet connections, enabling user interactions with the system [23]. The back end includes different cloud service models [24], like SaaS, PaaS, and IaaS. Often referred to as 'Layered computing model' is cloud computing architecture [25]. CC architecture can be classified into 4 layers that are hardware layer, infrastructure layer, platform layer, application layer as seen in Fig. 3.

Description of each layer is defined as follows [26]:

- **Hardware Layer**: Cloud is handled with physical resources. Controlling physical servers, switches, routers, the power system is the responsibility of the hardware layer. The implementation of the hardware layer is provided in a data center. There are several servers interconnected by routers and switches in the data center. Hardware layers have several issues including fault tolerance, hardware configuration, traffic management, and management of resources.
- **Infrastructure Layer**: The virtualization layer is named as well. Cloud computing is an important aspect. Infrastructure layers focused on core aspects such as the use of virtualized technologies for the dynamic assignment of resources. The infrastructure layer uses virtualization techniques to collect processing and storage resources and divide physical resources. E.g. Xen, VMware.

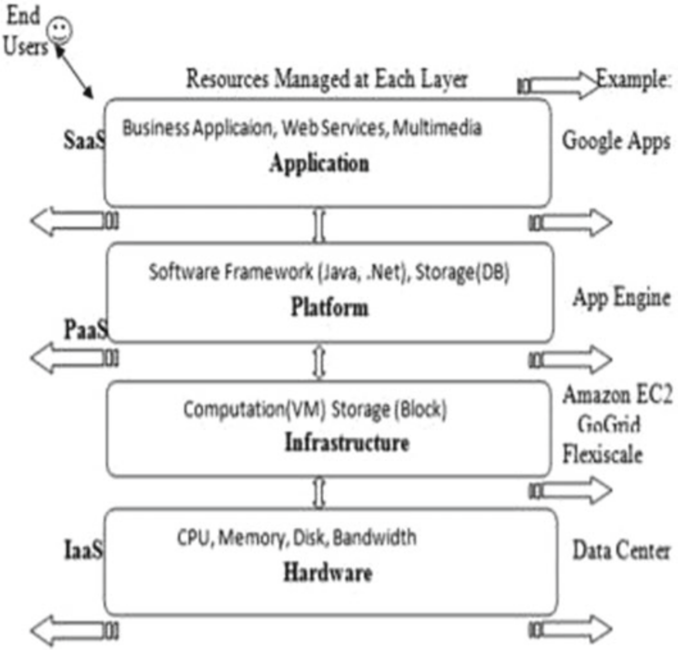

End Users
Resources Managed at Each Layer
Example:
SaaS
Business Applicaion, Web Services, Multimedia
Application
Google Apps
Software Framework (Java, .Net), Storage(DB)
Platform
App Engine
PaaS
Computation(VM) Storage (Block)
Infrastructure
Amazon EC2
GoGrid
Flexiscale
IaaS
CPU, Memory, Disk, Bandwidth
Hardware
Data Center

Fig. 3 Architecture of cloud computing

- **Platform Layer**: This layer is made up of operating system and application framework. It is installed on top of the infrastructure layer. The key principle of the platform layer is CC to reduce overhead for direct deployment in VM containers. E.g., Google App Engine runs on a platform layer to allocate API supports for data storage of various web applications.
- **Application Layer**: It is built on the top level of cloud architecture. It is composed of an actual cloud application. Cloud applications have essential features to achieve better performance, lower operating cost, availability, and scalability.

Thus, this architecture is more modular than other architecture (traditional architecture). Loosely coupled concepts are used in each layer. This architecture allows cloud computing to meet a vast variety of applications and to decrease total costs. No need for the high-power computer to run web-based applications is available in the cloud computing infrastructure.

2.4 Data Center in Cloud Network

Cloud migration platforms are a new and rapid trend. Cloud offers a standardized front-end interface, enabling a large number of applications to be executed on a similar hardware platform. DCs are the backbone of cloud computing development networks. Cloud output depends on the data centers' computing, storage, and network availability [27]. The demands for cloud working load in today's data centers, thus, is reflected in demands for resources.

The data center consists of servers, devices and storage, cooling systems, network devices, power systems, etc. [28]. DCs are for large-scale systems for services like online enterprises, smart grid and mathematical computations. DC is being used to model core services on the cloud infrastructure network. It comprises a group of hosts who handle a group of VM whose roles are to handle "low-level" processing, and a minimum of 1 DC should be set up to commence simulation.

Virtualization technologies [29, 30] are additionally supplied to data centers which allow multiplexed and shared multiple resources between huge no. of users by various time-varying access patterns [31]. DC management is difficult. There are approximately two types of challenge:

(1) *Management of resources,* which focuses on dynamic workloads management provided a resource pool and
(2) *Planning of capacity,* which concentrates on the provision of resources.

DCN (Data Center network) comprises DC also offers data center connections, as defined in its networking topology, the routing and switching equipment as well as the protocols it uses [32]. for the following purposes DCN provides some features to better organize cloud computing [33]:

- DCN allows thousands of data center servers to be efficiently connected such that cloud computing can extend its operation easily by adopting the DCN topology.
- In massive machine-to-machine connectivity, DCN provides traffic reliabilities and efficiencies that generate activities of cloud computing as working loads distributed on servers at data centers.
- DCN embraces different techniques of virtualization that help DCN build virtual machines (VMs), virtual networks, and virtual functions. The DCN should be scalable to isolate and migrate to large numbers of virtual instances.
- Existing DCN research has produced applications in some use cases, including green computing and DC backup, that may also address challenges of cloud infrastructure.

2.5 Virtual Machines in Cloud Computing

Cloud computing includes parallel processing principles and distributed computing to provide shared services via physical server hosting VMs. The service-orientated architecture reduces connectivity costs to collect customer information deliver more flexibility and demand-driven services, etc. The premise behind the generated cloud computing concept is that the processing of information is a public service, which can be achieved more effectively in massive computer farms and storage systems that can be made available worldwide through the Internet. Effective management of

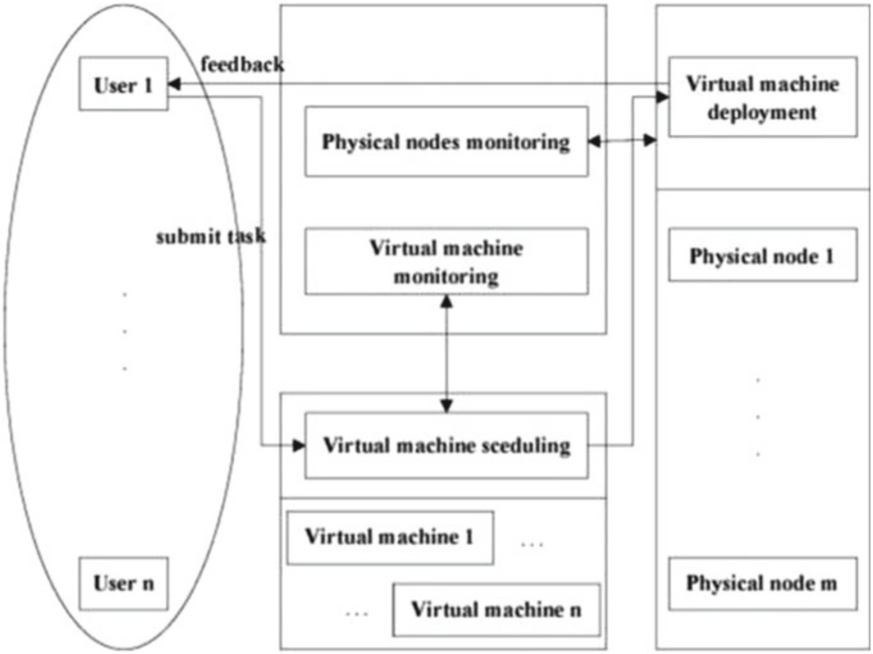

User 1
feedback
submit task
User n
Physical nodes monitoring
Virtual machine monitoring
Virtual machine sceduling
Virtual machine 1
…
…
Virtual machine n
Virtual machine deployment
Physical node 1
Physical node m

Fig. 4 Process of VM scheduling [35]

VMs [34] specifically affects the use of the system's resources and QoS. The amount of VM distributing across physical servers seems to be inconsistent over a certain period, given the dynamic nature of the cloud environment. In this case, VMs have to be moved from an overwhelmed server to a load-based server to balance the load.

Figure 4 presents the general cloud data center VM scheduling process. Scheduling can be widely categorized into user applications, control of resources, and scheduling mode.

VM's component handles the allocation of various hosts to various virtual machines so that the computing cores can be allocated (by the host) to VMs. This configuration is based on method, and the default VM assignment policy is 'first-come, first-serve.' Virtualization allows the live migration of VMs [36] with several VMs loaded on several physical machines (PMs) referred to as VMs. An efficient strategy to reduce energy consumption, operational costs, hardware costs, conformance/violation of SLAs, CO_2 emissions also enhance hardware and service reliability, efficiency and hardware life, LB and use of a CCS may be a VM consolidation algorithm. VM consolidation in the cloud environment can essentially reduce energy use and QoS.

2.6 Motivation of Study

The modern cloud computing model offers sophisticated benefits and benefits compared to previous computing paradigms and is adapted, migrated, and adopted by many organizations. Cloud computing has evolved from a prospective logic over the past few years; business is the idea of virtualization to a rapidly increasing sector in the IT sector. Today, recession-hit firms are becoming increasingly aware that they can easily tap into the cloud and quickly access best-of-breed business applications or drastically enhance their infrastructure services at marginal costs. But, some issues, difficulties and effects are still found which scholars, academics and business intelligence (BI) practitioners are presently addressing.

Clouds still present enterprises with security concerns [37] by using cloud computing. Upon details and IT critics behind the firewall, users are still concerned about the susceptibility to attack. Computing in clouds also provides reliability around the clock. There were some cases of outages for a few hours from cloud computing services. Contrary to the conventional computing model, cloud computing uses virtual computing technologies, as users have access to cloud computing services, they can leak hidden information. Attackers will evaluate crucial jobs based on the user's computing task [38]. The growth of cloud computing needs to open standards. The interpretation of most cloud providers with an [39] API, usually well-documented but often special to its execution and thus not interoperable. Heavy transaction-oriented also other applications of data-intensive in which CC can miss sufficient efficiency can be the key to performance. In addition, users far from cloud-based services suffer from high latency and delay. Software and hardware can be saved for companies offering the CC; however, higher latency costs can be incurred. Bandwidth costs may be small for smaller internet applications and not resource-intensive, nonetheless data-intensive applications can rise significantly. Users can be certain that even certain cloud computing services providers can never become irrelevant, or get bought and swallowed up by some bigger company. "Cloud potential providers can retrieve the data and it is imported into replacement application through either format" -Gartner [40]. Cloud service providers are required to comply with legal issues concerning data responsibility and ownership for loss or misuse of data. Legal problems vary from those caused by traditional hosting or outsourcing [41]. These problems and challenges encourage us to focus more on this subject of cloud computing problems.

3 Load Balancing in Cloud Computing

LB is a key problem and issue in cloud environments [42]. The method of allocating and reassigning the load between usable resources is designed to optimize performance, minimize cost and response times, improve efficiency, resource use and save energy. Excellent load balancing strategies can include SLA and user satisfaction.

Therefore, it is a prerequisite to the performance of CC environments to ensure reliable algorithms and processes for load balancing.

LB model is demonstrated in Fig. 5, LB accepts user requests and executes load balancing algorithms to allocate applications to VMs. Load balancer determines that VM's next request must be allocated. The controller of DC is responsible for task management. LB algorithm provides tasks to delegate tasks to appropriate VM. VM manager is responsible for VMs. Summary of LB policies are shown in Table 1.

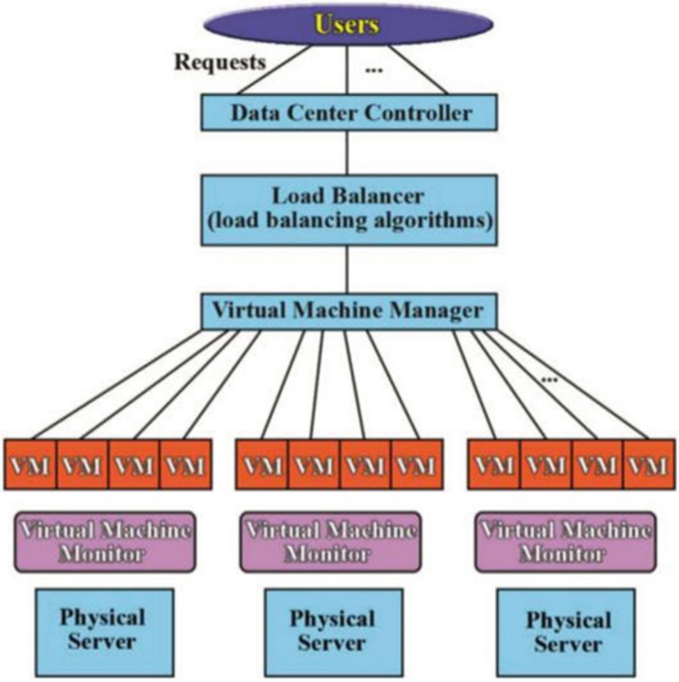


Fig. 5 Load balancing model [43]

Table 1 Summary of LB policies

S. No.	Location policy	Information policy	Selection policy	Transfer policy
1	Find the right partner for transfer	Determine how long information about nodes must be collected	Selecting factors for transferring a task: migration overhead	Includes: task re-scheduling task migration
2	Task monitor availability of resources needed for partner migration	There are three kinds: demand-driven, periodic, and state policies information policy	A no. of remote system calls time of task execution	Depends upon thresholds in terms of load units

Virtualization is leading CC technology. Virtualization aims mostly to share costly hardware between VMs. VM is computer software execution on which operating systems and applications will work. The user requests are processed by VMs. Users are found around the globe and send their requests arbitrarily. For processing, requests must be delegated to VMs. The distribution of tasks is also an important problem in CC. If several VMs are overloaded when others are idle or have some work to do, QoS is reduced. By reducing QoS, consumers are not happy with the system and will never return. The VM is created and managed by a hypervisor or VMM (Virtual Machine Monitor). Multiplexing, suspension (storage), provision (resume) and life migration [44] are four operations of VMM. For load balancing, certain operations are required. In [45] it was specified that load balance would take account of two tasks: allocation of resources and scheduling of tasks. The effect of these 2 tasks is high infrastructure availability, increased use of power, energy savings, reduced cost of resource use, maintaining cloud storage elasticity, and reducing carbon emissions.

3.1 Classification of LB Techniques

This section describes basic testing activities using various approaches to load balancing. The study of the various methods of load balancing for edge computing is explained as follows [46].

Different forms of strategies of LB are seen in Fig. 6. Different types are introduced to provide an improved balance of loads in edge computing, like security, traffic-load based, heterogeneous, optimization-based, joint-load based, heuristic and multi-access based, dynamic load-based, allocation-based, and allocation-based approaches.

3.2 Load Balancing Metrics

The metrics for LB in CC are reviewed in this section and summarized as follows [47]:

- **Response time:** It calculates the overall time required to fulfill a submitted task by the system.
- **Throughput:** This measurement is utilized to quantify the number of processes per unit time performed.
- **Scalability:** An algorithm is capable of uniform LB in the system as per demands, as no. of nodes increases. It is a highly scalable preference algorithm.
- **Makespan:** This metric is utilized to measure maximum completion time or the amount of time a user has attributed the tools.
- **Migration time:** time to switch tasks from overloaded node to undercharged node.

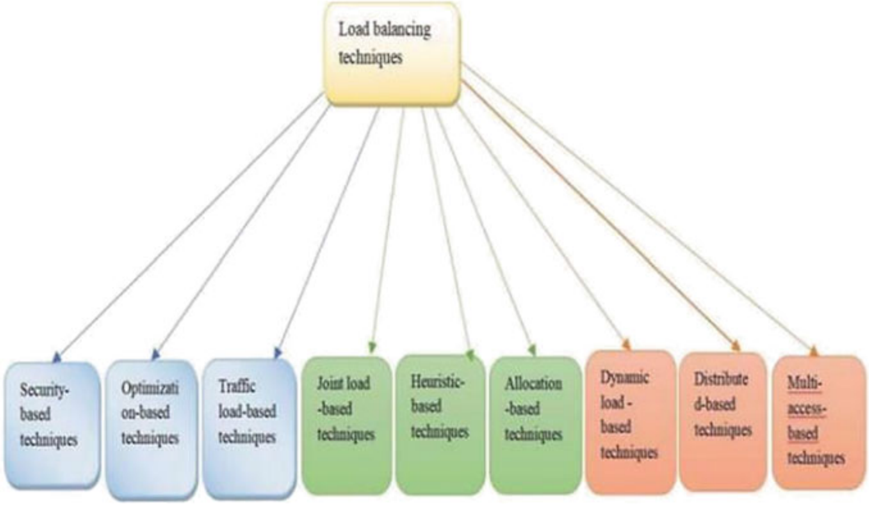

Load balancing techniques
Security-based techniques
Optimizati on-based techniques
Traffic load-based techniques
Joint load -based techniques
Heuristic-based techniques
Allocation -based techniques
Dynamic load - based techniques
Distribute d-based techniques
Multi-access-based techniques

Fig. 6 Categorization of LB techniques in CC

- **Fault tolerance:** It decides the algorithm's capacity to perform LB in event of such nodes or ties breakdown.
- **Performance:** After the LB algorithm, it tests system efficiency.
- **Degree of imbalance:** This measurement the imbalance of VMs.
- **Carbon emission:** The amount of carbon that all materials generated is calculated. Load balancing played an important part in reducing this metric by switching and shutting loads from underloaded nodes.
- **Energy consumption:** It determines how much energy all nodes consume. Load balancing prevents overheating and thereby reduces energy consumption by LB across all nodes.

4 Task Scheduling in Cloud Environment

Cloud computing handles a range of virtualized capabilities that make planning an important component. A customer can use thousands of virtualized properties in the cloud for each task [48]. Manual scheduling is therefore not a workable alternative. The underlying concept behind scheduling consists of distributing tasks (complex and diverse nature) between cloud resources in a manner that reduces time losses and maximizes efficiency by scheduling algorithms. Several study activities have in the past looked at task planning. Cloud operating resources are tracked and loads are computed for any resource before assigning workload or tasks on VMs (resources). If any VM is in over-used mode, a task for those types of resources is not assigned.

The task is a workpiece to be carried out within a given time frame. Inside the cloud, the task is in two forms [49]: a task that is independent and dependent. Task scheduling is the mechanism by which the services are assigned to this task for a certain amount of time. TS is a very common theme in the field of CC. It is used to plan tasks to improve the usage of resources by the allocation of such tasks to definite resources in particular. The main objective of the task algorithm is to increase reliability, enhance service quality, maintain productivity across tasks, and reduce costs. In the TS process, virtual tools are used to their maximum potential. For high performance, effective resource scheduling is essential. The completion period and the cost of completing the task are two different criteria.

4.1 Types of Task Scheduling

Categories of TS are as follows [50]:

- User-Level Scheduling
- Cloud Service Scheduling
- Heuristic Scheduling
- Static and Dynamic Scheduling
- Workflow Scheduling
- Real-Time Scheduling.

4.2 Framework of TS in CC Environment

There is a framework [51] for the TS system in CC. In Fig. 7, users are accessing the cloud environment through the internet. The cloud part shows how the cloud is managing to serve various requests given by the consumers.

(1) **SLA Monitor:** First, a client sends the service request. SLA monitor then reviews the requested QoS application before deciding whether the submission should be accepted or denied. SLA Monitor is responsible for monitoring the success of the job submitted and disciplinary measures must be taken promptly if there are any violations.
(2) **Resource Discovery and Monitoring:** Resource search can essentially be defined as the job of the manufacturer to locate the right resources to satisfy incoming customer demands. Cloud computing's main advantages are the ability to receive and distribute data on demand, so the management of resources must be continuous.
(3) **Task Scheduling**: The input of the algorithm for task scheduling is usually an abstract model. This abstract model describes tasks without defining were tools to perform the tasks are physically placed.
(4) **Reschedule:** If a task cannot be done due to a processor malfunction or other problem, the incomplete task in the next calculation will be rescheduled.

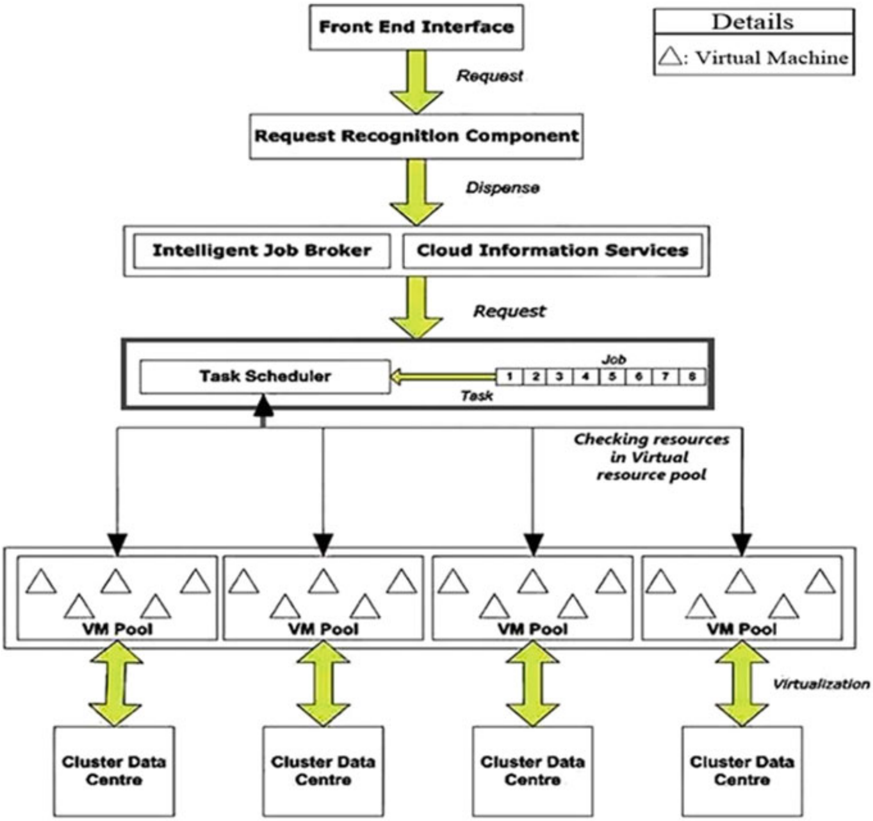


Fig. 7 Framework of TS in CC environment

(5) **Scheduling Optimizer**: Several relevant candidates are illuminated after the information is obtained about the available resources in the cloud. The framework for resource selection chooses a solution that meets all demands and optimizes infrastructure usage. The allocation of resources can be made using an algorithm for optimization. Various optimization methods can be used by common and well-known techniques, including simple metaheuristic algorithms. Examples include GA, ACO and PSO for the cloud.

(6) **Advanced Resource Reservation Monitor**: The Advanced Resource Reservation Control ensures QoS connections in the data center to different services. In the future, users will protect the necessary services. This is necessary if some processors need to be run to complete critical time applications such as real-time and Workflow applications for parallel applications. Providers can more reliably forecast potential demand and use.

(7) **Data Centers and VMs**: Data centers are the infrastructure or hardware part where all the physical servers present. Physical servers are not used as viable

users. Through virtualization, they are converted into different VMs and users' jobs are going to run on these VMs in scheduling tasks.VM manager manages the VMs.

5 Meta-Heuristic Optimization Techniques

Due to the usefulness of these metaheuristic algorithms in solving massive computational and complex problems [52] has been enormously common in the past 20 years. There are some useful metaheuristic characteristics such as:

(1) These algorithms are not unique to problems.
(2) Effectively explore a meta-heuristic algorithm in search areas to find (near) optimal solution or sub-optimal solution of NP-complete problems.
(3) Approximate and typically non-deterministic meta-heuristic algorithms.

Metaheuristic algorithms are problem-independent and can be used to solve issues in a variety of domains. Metaheuristic approaches are standard techniques for solving NP-hard optimization problems.

$$\text{Meta-heuristic} = \text{Heuristic} + \text{Randomization.}$$

Many meta-heuristic algorithms are available in the cloud environment to find an approximate (suboptimal) solution to an NP-Complete problem in a short amount of time. TS is one of the problems due to the vast space for a solution which takes a long time to find an optimal solution. Diverse choice of meta-heuristic algorithms like PSO, ACO, GA, Artificial Bee Colony (ABC), Simulated Annealing (SA), Differential evolution algorithm (DEA), BAT optimization, Bacteria Foraging Optimization, Firefly optimization, Cat swarm optimization, Lion optimization, Cuckoo search, etc.

We also addressed and analyzed numerous task scheduling algorithms in the cloud environment with a meta-heuristic method and their benefits and drawbacks (demonstrated in Table 2). In evaluating precision solutions to a specific problem, the right choice of the optimization algorithm is greatly helpful. At first, we will analyze the PSO algorithm since it is easier to converge the PSO algorithm and the complexity than other metaheuristic algorithms.

6 Literature Survey

1. Oliveira et al. [3] Recently, a great many projects have carried out extensive cloud research, including load balancing and task scheduling. These works illustrate some of the core problems facing state-of-the-art load balancing

Table 2 Summary of the tested meta-heuristic algorithms, including their benefits and drawbacks

S. No.	Author (Year)	Algorithm	Nature of tasks	Techniques	Benefits	Drawbacks
1	Elaziz et al.	Hybrid algorithm	Independent	MSDE algorithm (MSA with DE)	Enhance makespan system time and throughput	A single objective (focused only time) is the algorithm performed and does not take account of other QoS parameters such as reliability, cost, energy
2	Adhikari et al.	BAT	Independent	LB-RC algo via BAT	Enhance cost, makespan time, resource utilization, reliability, etc	Algorithm doesn't discuss compromise among time and cost QoS parameter
3	Sharma and Kumar PSO-COGEN	PSO-COGENT algorithm	Independent	Modified PSO depends on APSO-VI	Enhance time, throughput, energy consumption, QoS, cost parameters, etc	QoS parameters such as availability, reliability, etc. cannot be taken into consideration. VMs can be overloaded when applications of VMs have been improperly described
4	Alaaeldin and Almezeini	Lion algorithm	Independent	TS by lion	Enhance cost, makespan time, resource usage and imbalance degree	Algorithm considers no limit and did not address the compromise solution

(continued)

Table 2 (continued)

S. No.	Author (Year)	Algorithm	Nature of tasks	Techniques	Benefits	Drawbacks
5	Addya et al.	SA	Independent	MVMP algorithm via SA	enhances, energy consumed, profit and execution time	The solution to static VM migration takes time
6	Zamanifar and Rastkhadiv	ABC	Independent	Agile task handling method via ABC	Enhance makespan time and degree of imbalance	Algorithms are only used to carry out independent tasks and main performance parameters such as reliability, energy, and cost, etc. are not considered, where the deadline is a restraint
7	Lin et al.	Binary PSO	NA	Modified sigmoid transfer function	Enhance time of High utility itemset mining	The sigmoid function proposed would not account for the last step of the operation
8	Pacini et al.	ACO	Independent	2 Level cloud scheduler method by ACO algorithm	Enhance response time, throughput and makespan time	Algorithms do not take into account the time and tasks priority; indirect communication mechanisms are utilized to communicate information (pheromone)

(continued)

Table 2 (continued)

S. No.	Author (Year)	Algorithm	Nature of tasks	Techniques	Benefits	Drawbacks
9	Kaushal and Verma	PSO based scheduling	Workflow applications	Bi-objective priority-based PSO	Reduce execution costs and time at the same time	Energy usage and other effective QoS are not taken into account
10	Ramezani and Khadeerhussain	PSO based scheduling algorithm	Independent	task migration, Cloud resource broker method	Enhance rejection ratio time, expense and task	Gbest is affected by local minima and energy utilization is not taken into account
11	Jana and Kulia	TBSLB- PSO	Independent	Task migration from overloaded VMs to under-loaded VMs	Reduce task transfer time and makespan time	Performance of algorithm suggested is not related by other state-of-the-art algorithms
12	Krishna and Babu	ABC	Independent	Task migration based LB using ABC	Time of response, runtime and performance, number of migrated tasks, throughput	Only non-preemptive tasks and deadline of tasks are not considered for proposed algorithm
13	Huo and Zhan	Hybrid PSO	Independent	SA with PSO is added	Enhance algorithm search capabilities with SA	After adding SA and PSO the complexity of the algorithm is improved

applications. Several task scheduling mechanisms have been defined by many researchers for load balancing.

2. Jena [53], focused on TSPSO (task scheduling using multi-objective nested particle swarm optimization) to enhance energy and processing time. Outcome from TSPSO was simulated through cloud platform for open source applications (CloudSim). The findings were finally related to the existing scheduling algorithms and found an optimum balance outcome for different goals with the proposed algorithm (TSPSO) [54].
3. Wang and Zhou [55], Proposed a technique for solving the MapReduce load imbalance issue created by the use of the default partition algorithm of the Hadoop platform. This suggested approach will optimize the activities and balance inputs of a decrease process in the map phase by using multiple partition techniques. Moreover, this suggested technique will use idle nodes to completely offset high load nodes, to attain optimum work scheduling throughout the reduction phase execution method [56].
4. Velde and Rama [57], In this cloud partitioning strategy, LB and resource optimization are effectively extended. Besides the optimal time of refresh that sets the state adjustments of the data center partitions for efficient use of resources. CloudSim is a simulation platform utilized to create prototype applications to prove hypothesis proof. Results have shown that this approach increases cloud resource efficiency, performance, and optimization [53].
5. Basha and Padmavathi [58], proposed an innovative, dynamic, and elasticity algorithm perform LB by ACO to perform load Balance between Systems existing in DCs [55].
6. Vijayakumar and Kanthimathi [59], The suggested scheme used additional virtual machines using genetic methods to address the requests of the highest virtual machines. The allocation of the best virtual machines will deal with demands very efficiently and quickly. During execution, the load might be balanced via the ACO technique if those VMs were overloaded by requirement. The above strategy will share the extra charge with other virtual machines, gently loaded and idle. On the other hand, after their work completion or in idleness total energy consumption is enhanced by turning away VMs [57].
7. (MPSO) hybridization and an improved Q-learning algorithm called QMPSO. Hybridization is done to change MPSO velocity by pbest and gbest depends upon the best action provided by improved Q-learning. The objective of hybridization is to improve machine performance by balancing the charge among VMs, optimize VM's output and maintain the balance between task priorities by maximizing the time of work. By comparing QMPSO results obtained during the simulation process with the existing LB and scheduling algorithm, the robustness of this algorithm is evaluated. The simulation compared with real results of the platform reveals that the proposed algorithm outperforms its competitor [60].
8. Malik and Bansal [61], An optimized approach to scheduling model and resource cost scheduling model labeled MultiFaceted Optimization Scheduling Framework (MFOSF) was timely in this study to resolve the restrictions and

change solution quality. The resource Cost model shows the relation between customer budget and production costs during scheduling. PSO can be used to achieve a model focused on optimizing efficiency and cost. There have been several simulations to test this approach using 4 different metrics (a) Makespan (b) Cost (c) Resource Utilization (d) Deadline. Based on the above parameters, experimental results demonstrate that the MFOSF-PSO approach was better than other models and in the best-case scenario increased by 57.4% [61].

9. Devaraj et al. [62], As a firefly hybrid and Improved Multi-objective Particle Swarm Optimization (IMPSO) technology, abbreviated as FIMPSO, the latest LB algorithm is proposed for use. This technique uses FF (Firefly) algorithm to limit search space to detect the improved response using the IMPSO technique. The proposed FIMPSO algorithm generated and improved effective average load for crucial measures, including proper use of resources and response time for tasks. The simulation result has shown the efficient performance of the FIMPSO model suggested compared to other methods. From the resulting simulation, the FIMPSO algorithm is understood to have generated a successful outcome with an average response time of 13.58 ms maximum, overall CPU usage of 98%, 93% memory utilization, 67% reliability and 72% throughput, along with 148, which were superior to all other techniques compared [62].
10. Miao et al. [63], The new algorithm called adaptive Pbest Discrete PSO (APDPSO) for PSO-based static load balancing was proposed to combat this problem. Good solution held in the external archive is used to upgrade the particles' personal best positions and to update velocity and direction vectors of particles by proposing a probability and comparability approach for PSO. MATLAB and CloudSim systems perform simulation experiments with random synthetic tasks. The findings revealed that our proposed algorithm dramatically increased swarm convergence and diversity and reduced the level of load imbalance relative to state art in the field [63].

6.1 Research Gaps and Issues

This section addresses the study deficiencies of cloud computing that need to be addressed in terms of existing load-balancing techniques.

- Lightweight security solutions were not considered for increased LB efficiency [64].
- Method [65] failed to implement the paradigm built-in CC and real field of joint edge computing. The procedure failed to increase QoE and reduce incoming traffic based on the adaptive bitrate approach for mobile video streaming and optimized edge caching [66].
- The cross-layer approach on heterogeneous networks was more complicated [67].
- The approach in [68] neglected to take into account MEC server security levels, which is very important to industry data.

- In [69], the complexity was not properly calculated and the reliability was not improved. In [70] the strategy to balance the positive and negative impacts on the migration of services has not been established.
- The method in [71] needed additional major structures for orchestration to improve the system performance.
- Performance has been improved, but the allocation of vehicles was highly imbalanced, leading to inefficient QoS for a vehicle to infrastructure (I2V) communication [72].
- However, the forecast was inaccurate because the task length was smaller and the training set was not included [73]. Large-scale heterogeneous MEC problems for optimizing training speed were not taken into consideration [74].
- In [75], live video and car navigation services could not be used. No further data sets were considered to improve the system's performance [76].
- The load-aware user associations that contain dynamic settings were not taken into account [77]. The transmitting power and rates of SCBSs were not taken into account in [78] to achieve minimum energy consumption.
- Because of many interferences and noises, the communication channel was not perfect [79]. They are normally unable to understand the degradation of QoE [80]. The process of the reorder did not however alter the efficiency of the load balancing [81].

The challenges of these approaches are taken into account as a reason for developing a new load-balancing method.

7 Conclusion

AC is the product of the Virtualization concept-enabled grid computing. LB is a challenge in the cloud environment, as it is a complex task to share the task equally between VMs. LB is the main CC mechanism that prevents overloading nodes. Load balancing stabilizes service quality (QoS), covering response time, cost, throughput, efficiency, and the use of resources. At the height of time, servers cannot manage incoming requests with no. of VMs available, meaning that some additional virtual machines needed to be run without fault or delay. The benefit of switching to a virtual environment is important, but there are significant efficiency and cost optimization issues in cloud computing due to different types of need for resources and multiple tasks. The success standard and requirements of SLA to be retained therefore become very difficult because of these constraints. Various alternatives to the planning model and resource cost timeline model have been applied to solve these limitations and to change solution quality. Different meta-heuristic optimization strategies were applied to conduct TS in the CC environment to manage demand on virtual machines.

This survey will offer an idea of the newest developments in LB in cloud systems and will provide an analysis of the technological advances, company benefits and growing use of data centers, and future trends. Amazon EC2, Microsoft Azure, and

Google App Engine are compared. This article highlights research gaps and load handling problems in CC andsummarizes the superiority of CC. The future analysis is to find effective methods for calculating the cloud results.

References

1. Academic paper http://www.mcs.csueastbay.edu/~lertaul/Cloud%20Security%20CamREADY.pdf
2. Chen G, Lu J, Huang J, Wu Z (2010) SaaAS—the mobile agent-based service for cloud computing in internet environment. In: Sixth international conference on natural computation, ICNC 2010. IEEE, Yantai, Shandong, China, 2010, pp 2935–2939. ISBN: 978-1-4244-5958-2
3. Oliveira D, Baião F, Mattoso M (2010) Towards taxonomy for cloud computing from an e-science perspective. In: Cloud computing: principles, systems, and applications (to be published). Springer, Heidelberg
4. Singh G, Sood S, Sharma A (2011) CM-Measurement facets for cloud performance. IJCA 23(3). Lecturer, Computer Science and Engineering, Eternal University, Baru Sahib (India)
5. Ertaul L, Singhal S (2009) Security challenges in cloud computing. California State University, East Bay
6. Global Netoptex Incorporated.—Demystifying the cloud. Important opportunities, crucial choices, pp 4–14. Available: http://www.gni.com. 13 Dec 2009
7. Gulshan S, Kalra M (2014) A novel approach for load balancing in cloud data center. 978-1-4799-2572-8/14/$31.00 c_2014. IEEE
8. Wu H, Ding Y, Winer C, Yao L (2010) network security for virtual machines in cloud computing. In: 5th International conference on computer sciences and convergence information technology, pp 18–21, Seoul, Nov 30–Dec 2, 2010. ISBN: 978-1-4244-8567-3
9. Kliazovich D, Arzo ST, Granelli F, Bouvry P, Khan SU (2013) eSTAB: energy-efficient scheduling for cloud computing applications with traffic load balancing. In: Green computing and communications (GreenCom). IEEE, pp 7–13
10. Greenberg A, Hamilton J, Maltz DA, Patel P (2008) The cost of a cloud: research problems in data center networks. ACM SIGCOMM Comput Commun Rev 39(1):68–73
11. Lu X, Kong F, Yin J, Liu X, Yu H, Fan G (2015) Geographical job scheduling in data centers with heterogeneous demands and servers. In: 2015 IEEE 8th international conference on cloud computing, pp 413–420. https://doi.org/10.1109/CLOUD.2015.62
12. Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I et al (2010) A view of cloud computing. Commun ACM 53(4):50–58
13. Wang W, Li B, Liang B (2013) Dominant resource fairness in cloud computing systems with heterogeneous servers. arXiv preprint arXiv:1308.0083
14. Shafiq DA, Jhanjhi NZ, Azween Abdullah (2021) Load balancing techniques in cloud computing environment: a review. J King Saud Univ Comput Inf Sci. https://doi.org/10.1016/j.jksuci.2021.02.007
15. Khan S, Sharma N (2014) Effective scheduling algorithm for load balancing (SALB) using Ant Colony
16. Optimization in cloud computing. Int J Adv Res Comput Sci Softw Eng 4(2)
17. Kushwah P (2014) A survey on load balancing techniques using ACO algorithm. (IJCSIT) Int J Comput Sci Inf Technol 5(5):6310–6314
18. Farrag AAS, Mahmoud SA, EI Sayed M, EI- Horbaty (2015) Intelligent Cloud Algorithms for Load Balancing problems: a survey. In: 2015 IEEE seventh international conference on intelligent computing and information systems (ICICIS ‘2015)
19. Houssein EH, Gad AG, Wazery YM, Suganthan PM (2021) Task scheduling in cloud computing based on meta-heuristics: review, taxonomy, open challenges, and future trends. Swarm Evol Comput 62:100841. https://doi.org/10.1016/j.swevo.2021.100841

20. Chen Y, Paxson V, Katz RH (2010) What's new about cloud computing security. In: University of California, Berkeley Report No. UCB/EECS-2010-5, January 2010
21. George SS, Suji Pramila R (2021) A review of different techniques in cloud computing. Mater Today Proc. https://doi.org/10.1016/j.matpr.2021.02.748
22. Radarnetworks and Novaspivak. http://radarnetworks.com
23. Odun-Ayo I, Ananya M, Agono F, Goddy-Worlu R (2018) Cloud computing architecture: a critical analysis. In: 2018 18th international conference on computational science and applications (ICCSA), 2018, pp 1–7. https://doi.org/10.1109/ICCSA.2018.8439638
24. Prasad MR, Lakshman Naik R, Bapuji V (2013) Cloud computing: research issues and implications. Int J Cloud Comput Serv Sci (IJ-CLOSER) 2(2):134–140
25. Verma, Kaushal S (2011) Cloud computing security issues and challenges: a survey. (July):445–454
26. Tiwari K, Chaudhary S, Shanu K (2015) Survey paper on cloud Computing. In: International conference on emerging trends in technology, science and upcoming research in computer science, Apr 2015, pp 1777–1782
27. Zhang Q, Cheng L, Boutaba R (2010) Cloud computing: state-of-the-art and research challenges. J Internet Serv Appl I:7–18
28. Joshi S, Kumari U (2016) Cloud computing: architecture and challenges. Mody Univ Int J Comput Eng Res 1(1):56–60
29. Birke R, Chen LY, Smirni E (2012) Data centers in the cloud: a large-scale performance study. In: 2012 IEEE Fifth international conference on cloud computing, pp 336–343. https://doi.org/10.1109/CLOUD.2012.87
30. Abts D, Felderman B (2012) A guided tour through data-center networking. Queue 10(5):10:10–10:23
31. Wood T, Tarasuk-Levin G, Shenoy PJ, Desnoyers P, Cecchet E, Corner MD (2009) Memory buddies: exploiting page sharing for smart colocation in virtualized data centers. In: VEE, 2009, pp 31–40
32. Mehta S, Neogi A (2008) ReCon: a tool to recommend dynamic server consolidation in multi-cluster data centers, In: NOMS, 2008, pp 363–370
33. Chen Y, Das A, Qin W, Sivasubramaniam A, Wang Q, Gautam N (2005) Managing server energy and operational costs in hosting centers. In: SIGMETRICS, 2005, pp 303–314
34. Wanga B, Qi Z, Maa R, Guana H, Vasilakos AV (2015) A survey on data center networking for cloud computing. Comput Netw 91:528–547
35. Wang A, Iyer M, Dutta R, Rouskas GN, Baldine I (2013) Network virtualization: technologies, perspectives, and frontiers. J Lightwave Technol 31(4):523–537
36. Sahu Y, Agrawal N (2015) A survey paper: cloud computing and virtual machine migration. IJCSN J 4(4):577–581
37. Liu L, Qiu Z (2016) A survey on virtual machine scheduling in cloud computing. In: 2016 2nd IEEE international conference on computer and communications (ICCC), 2016, pp 2717–2721. https://doi.org/10.1109/CompComm.2016.7925192
38. Zolfaghari R, Sahafi A, Rahmani AM, Rezaei R (2021) Application of virtual machine consolidation in cloud computing systems. Sustaine Comput Inf Syst 30:100524. https://doi.org/10.1016/j.suscom.2021.100524
39. Mills E (2009) Cloud computing security forecast: clear skies
40. Jiang J, Wen W (2010) Information security issues in cloud computing environment. Netinfo Secur. https://doi.org/10.3969/j.issn.1671-1122.2010.02.026
41. Clark C, Fraser K, Hand S, Hansen JG, Jul E, Limpach C, Pratt I, Warfield A (2005) Live migration of virtual machines. In: Proceedings of NSDI'05. Berkeley CA, USA, 2005. USENIX Association, pp 273–286
42. Gartner, Seven cloud-computing security risks. http://www.infoworld.com. 02 July 2008
43. Prasad MR, Gyani J, Murti PRK (2012) Mobile cloud computing implications and challenges. IISTE J Inf Eng Appl (JIEA) 2(7):7–15. http://iiste.org
44. Jadeja Y, Modi K (2012) Cloud computing—concepts, architecture and challenges. In: International conference on computing, electronics and electrical technologies [ICCEET]

45. Gupta H, Sahu K (2014) Honey bee behavior based load balancing of tasks in cloud computing. Int J Sci Res 3(6)
46. Hwang K, Dongarra J, Fox GC (2013) Distributed and cloud computing: from parallel processing to the Internet of Things
47. Ivanisenko IN, Radivilova TA (2015) Survey of major load- balancing algorithms in distributed system. In: Information technologies in innovation business conference (ITIB)
48. Pydi H, Iyer GN (2020) Analytical review and study on load balancing in edge computing platform. 2020 Fourth international conference on computing methodologies and communication (ICCMC), 2020, pp 180–187. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00036
49. Jafarnejad Ghomi E, Masoud Rahmani A, Nasih Qader N (2017) Load-balancing algorithms in cloud computing: a survey. J Netw Comput Appl 88:50–71. https://doi.org/10.1016/j.jnca.2017.04.007
50. Arunarani AR, Manjula D, Sugumaran V (2018) Task scheduling techniques in cloud computing: a literature survey. Future Gener Comput Syst. https://doi.org/10.1016/j.future.2018.09.014
51. Abdul Qadir OS, Ravi G (2020) A survey on task scheduling algorithms in cloud computing. Int J Innovations Eng Technol 15(4):29–35. https://doi.org/10.21172/ijiet.154.06
52. Sharma P, Shilakari S, Chourasia U, Dixit P, Pandey A (2020) A survey on various types of task scheduling algorithm in cloud computing environment. Int J Sci Technol Res 9(01):1513–1521
53. Jena RK (2015) Multi objective task scheduling in cloud environment using nested PSO framework. Procedia Comput Sci 57:1219–1227. https://doi.org/10.1016/j.procs.2015.07.419
54. Venu G, Vijayanand KS (2020) Task scheduling in cloud computing: a survey. Int J Res Appl Sci Eng Technol (IJRASET) 8(V):2258–2266
55. Wang S, Zhou H (2016) The research of MapReduce load balancing based on multiple partition algorithm. In: 2016 IEEE/ACM 9th international conference on utility and cloud computing (UCC), pp 339–342
56. Kumar M, Sharma SC, Goel A, Singh SP (2019) A comprehensive survey for scheduling techniques in cloud computing. J Netw Comput Appl. https://doi.org/10.1016/j.jnca.2019.06.006
57. Velde V, Rama B (2017) Simulation of optimized load balancing and user job scheduling using CloudSim. In: 2017 2nd IEEE international conference on recent trends in electronics, information and communication technology (RTEICT), pp 1379–1384. https://doi.org/10.1109/RTEICT.2017.8256824
58. Padmavathi M, Basha SM (2017) Dynamic and elasticity ACO load balancing algorithm for cloud computing. In: 2017 International conference on intelligent computing and control systems (ICICCS), pp 77–81. https://doi.org/10.1109/ICCONS.2017.8250571
59. Kanthimathi M, Vijayakumar D (2018) An enhanced approach of genetic and ant colony based load balancing in cloud environment. In: 2018 International conference on soft-computing and network security (ICSNS), pp 1–5. https://doi.org/10.1109/ICSNS.2018.8573608
60. Kumar KP (2018) gravitational emulation-grey wolf optimization technique for load balancing in cloud computing. In: 2018 Second international conference on green computing and Internet of Things (ICGCIoT), pp 177–184. https://doi.org/10.1109/ICGCIoT.2018.8753108
61. Bansal M, Malik SK (2020) A multi-faceted optimization scheduling framework based on the particle swarm optimization algorithm in cloud computing. Sustain Comput Inf Syst 28:100429. https://doi.org/10.1016/j.suscom.2020.100429
62. Devaraj AFS, Elhoseny M, Dhanasekaran S, Laxmi Lydia E, Shankar K (2020) Hybridization of firefly and improved multi-objective particle swarm optimization algorithm for energy efficient load balancing in cloud computing environments. J Parallel Distrib Comput 142:36–45. https://doi.org/10.1016/j.jpdc.2020.03.022
63. Miao Z, Yong P, Mei Y, Quanjun Y, Xu X (2021) A discrete PSO-based static load balancing algorithm for distributed simulations in a cloud environment. Future Gener Comput Syst 115:497–516. https://doi.org/10.1016/j.future.2020.09.016
64. Li J, Luo G, Cheng N, Yuan Q, Wu Z, Gao S, Liu Z (2018) An end-to-end load balancer based on deep learning for vehicular network traffic control. IEEE Internet of Things J 6(1):953–966

65. Dong Y, Xu G, Ding Y, Meng X, Zhao J (2019) A ‘joint-me’task deployment strategy for load balancing in edge computing. IEEE Access 7:99658–99669
66. Liu J, Shou G, Liu Y, Hu Y, Guo Z (2018) Performance evaluation of integrated multi-access edge computing and fiber-wireless access networks. IEEE Access 6:30269–30279
67. Liu L, Chan S, Han G, Guizani M, Bandai M (2018) Performance modeling of representative load sharing schemes for clustered servers in multiaccess edge computing. IEEE Internet of Things J 6(3):4880–4888
68. Niu X, Shao S, Xin C, Zhou J, Guo S, Chen X, Qi F (2019) Workload allocation mechanism for minimum service delay in edge computing-based power Internet of Things. IEEE Access 7:83771–83784
69. Xu X, Fu S, Cai Q, Tian W, Liu W, Dou W, Sun X, Liu AX (2018) Dynamic resource allocation for load balancing in fog environment. Wirel Commun Mobile Comput 2018
70. Fahs A, Pierre G (2019) Proximity-aware traffic routing in distributed fog computing platforms
71. Lee H, Kwon B, Kim S, Lee I, Lee S (2015) Theoretical-analysis-based distributed load balancing over dynamic overlay clustering. IEEE Trans Veh Technol 65(8):6532–6546
72. Fernando N, Loke SW, Rahayu W (2016) Computing with nearby mobile devices: a work sharing algorithm for mobile edge-clouds. IEEE Trans Cloud Comput (2016)
73. Lu H, Gu C, Luo F, Ding W, Liu X (2020) Optimization of lightweight task offloading strategy for mobile edge computing based on deep reinforcement learning. Future Gener Comput Syst 102:847–861
74. Fan Q, Ansari N (2018) Towards traffic load balancing in drone-assisted communications for IoT. IEEE Internet Things J 6(2):3633–3640
75. Han T, Li S, Zhong Y, Bai Z, Kwak K-S (2019) 5G software-defined heterogeneous networks with cooperation and partial connectivity. IEEE Access 7:72577–72590
76. Li C, Wang YaPing, Tang H, Zhang Y, Xin Y, Luo Y (2019) Flexible replica placement for enhancing the availability in edge computing environment. Comput Commun 146:1–14
77. Li C, Sun H, Chen Y, Luo Y (2019) Edge cloud resource expansion and shrinkage based on workload for minimizing the cost. Future Gener Comput Syst 101:327–340
78. Asif-Ur-Rahman M, Afsana F, Mahmud M, Shamim Kaiser M, Ahmed MR, Kaiwartya O, James-Taylor A (2018) Toward a heterogeneous mist, fog, and cloud-based framework for the internet of healthcare things. IEEE Internet of Things J 6(3):4049–4062
79. Bulkan U, Dagiuklas T, Iqbal M, Huq KMS, Al-Dulaimi A, Rodriguez J (2018) On the load balancing of edge computing resources for on-line video delivery. IEEE Access 6:73916–73927
80. Ramaswamy L, Liu L, Iyengar A (2007) Scalable delivery of dynamic content using a cooperative edge cache grid. IEEE Trans Knowl Data Eng 19(5):614–630
81. Wan J, Chen B, Wang S, Xia M, Li D, Liu C (2018) Fog computing for energy-aware load balancing and scheduling in smart factory. IEEE Trans Ind Inf 14(10):4548–4556

Balancing Exploration and Exploitation in Nature Inspired Computing Algorithm

K. Praveen Kumar, Sangeetha Singarapu, Mounika Singarapu, and Swaroop Rakesh Karra

Abstract Nature is the one of the best inspiration for solving problems around us. Nature inspired computing algorithms are inspired from nature. These algorithms have inbuilt features of self learning, self motivation, co-ordination and collective behavior for solving a particular task. These nature inspired computing algorithms are specified by local search (exploitation) and global search (exploration). As per the effects of nature inspired computing algorithms, depends the searching capability. This paper focuses balance of local and global searching attribute of nature inspired computing algorithms. The people in a group are managed by leaders. By applying PSO algorithm, the best leader in the group is selected using efficient fitness functions. And this proposed fitness function is compared with the benchmark fitness function for providing better optimum results in the balancing of exploration and exploitation.

Keywords Nature inspired computing algorithms · Swarm intelligence · Particle swarm optimization · Local search · Global search

K. P. Kumar (✉) · S. Singarapu
Department of IT, KITSW, Warangal, India
e-mail: kpk.it@kitsw.ac.in

S. Singarapu
e-mail: ss.it@kitsw.ac.in

M. Singarapu
Dept. of MS ITM, Campbellsville University, Louisville, KY, USA

S. R. Karra
MS IT, New Castle Wilmington University, New Castle, DE, USA

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_12

1 Introduction

Nature is the great inspiration for solving complex problems in the universe. Nature İnspired Computing (NIC) algorithms are inspired from nature for solving complex problems and produce nearest results at reasonable amount of time [1, 2]. The NIC algorithms having features of collective behaviour, self learning, self adapt nature are focused to solve NP hard problems. The nature inspired computing algorithms have feature of random probability i.e., algorithms randomly pick solutions for better optimum results [3–6].

Particle Swarm Optimization (PSO) is a popular nature inspired computing algorithm proposed by Dr. Eberhart and Dr. Kennedy in 1995. PSO is influenced by social behavior of birds and animals [7–10]. PSO is a well known nature inspired computing algorithm for the optimization of initializing random population of solutions. PSO updates component generations for searching best positions. Their particles manage their velocity of own positions and also positions of their neighbor particles to get optimum solutions.

2 Existing Work

Paper [11] focused on particle swarm optimization about local search (exploitation) and global search (exploration) by applying master and slave method. The leader swarm acts, master and other swarms act and then the slaves. The slave swarms provide new positions along best fitness values to master swarms. The position of new particles is updated by master swarm for best position and manage updated positions of themselves (master swarms) and update slave swarms' position.

Article [12] proposed reduced energy conservation model of power system for socio-economic development by applying modified particle swarm optimization. Balancing of local search as well as global search using a new technique of multi-swarm was proposed. In the article, 'clan' was introduced in place of swarms, and every group had a leader. Representing local best is the best solution that is represented by the leader of clan [13, 14]. To select global best, all these leaders can represent themselves for selection procedure. The clans update their positions, and manage their population risks for selection of local and best positions accordingly.

2.1 *Particle Swarm Optimization*

PSO is initialized with randomly generated population of particles (initial swarm) and a random velocity is assigned to each particle that propagates the particle in search space toward optima over a number of iterations. Each particle has a memory, remembering best position attained by it in the past, which is called personal best position (Pbest). Each particle has its Pbest and the particle with the best value of fitness is called global best particle (Gbest). Suppose that the search space is D dimensional, the ith particle of the population can be represented by a D-dimensional vector (xi1, xi2, …, xiD)T. The velocity of this particle can be represented by another D-dimensional vector (Vi1, Vi2, …, ViD)T.

The previously best visited position of ith particle is denoted by Pi and the best particle in the swarm is denoted by Pg. The update of the particle's position is accomplished by the following two equations. Eq. (1) calculates a new velocity for each particle based on its previous velocity and Eq. (2) updates each particle's position in search space.

$$V_{id}^{k+1} = wV_{id}^{k} + c_1 r_1\left[p_{id}^{k}(t) - x_{id}(t)\right] + c_2 r_2\left[p_{g}^{k}(t) - x_{id}^{k}(t)\right] \tag{1}$$

$$x_{id}^{k+1}(t+1) = x_{id}^{k}(t) + v_{id}^{k+1}(t+1) \tag{2}$$

where k = iteration number, d = 1,2,3,…, D, i = 1,2,3,…, N, N = swarm size, w = inertia weight which controls the momentum of particle by weighing the contribution of previous velocity, c1 and c2 are positive constants called acceleration coefficients and r1 and r2 are random numbers uniformly distributed between [0,1].

PSO particle represents D-dimensional area as particle i (particle i1, particle i2, particle i3… particle iD), where D represents dimension and particle iD minimum dimension, maximum dimension is mini and max limitations of Dth dimension. Velocity of particle i is velocity i = (velocity i1, velocity i2, velocity i3…. Velocity iD). Particles and time t made manipulations as per the following relations.

$$\begin{aligned} &Velocity\, i\ (time\, t+1) = velocity\, i\ (time\, t) \\ &+ r\ value\, 1\, con\, 1\ (Ppbp\ -\ particle\, i\ (time\, t)) \\ &+ r\ value\, 2\, con\, 2\ (Pgbp\ -\ particle\, i\ (time\, t)) \end{aligned} \tag{3}$$

$$Particle\, i\ (time\, t+1) = particle\, i\ (time\, t) + velocity\, i\ (time\, t) \tag{4}$$

In above equation, r value1 and r value2 are random values in the range 0 and 1. The con 1 and con 2 are the velocity constants that improve the movement of particles for a given iteration. The last best position of ith particle is noted as Ppbp and the best position of ith· particle is noted as Pgbp.

$$\begin{aligned} &Velocity\, i\ (time\, t+1) = inertia\ w\ x\ velocity\, i\ (time\, t) \\ &+ rvalue1con1\ (Ppbp\ -\ particle\, i\ (time\, t)) \\ &= rvalue2con2\ (Pgbp\ -\ particle\, i\ (time\, t)) \end{aligned} \tag{5}$$

In the above equation, inertia is added [15]. For a perfect balancing between the global and local explorations through inertia, minimum iterations are considered to achieve optimal solution.

$$inertia = inertia\,max = \frac{inertia\,max - inertia\,min}{iterations\,max} \times iterations \quad (6)$$

In above equation, inertiamax is denoted by initial weight, inertiamin is denoted by final weights, iterationsmax is max number of iterations considered in each iteration which is denoted as the number of iterations at present.

In addition, Linearly decreasing inertia weight method in PSO (LPSO) is a random inertia weight factor for dynamic systems tracking [16]. Inertia weight factor noted for random vary as per the following equation.

$$inertia\ = 0.5 - \frac{rand()}{2} \quad (7)$$

In above equation, rand() denotes equally allocated random numbers in between values 0 to 1 range. Value 1.494 must be maintained by acceleration coefficients. This equation refers to Random weight method in PSO (RPSO).

3 Proposed Work

3.1 Improved Multiple Particle Swarm Optimization (IMPSO)

The main objective of Extensible Multiple Swarm Particle Swarm Optimization is the co-ordination behaviour of communication among the people in the group ‘clan’ (group of people) as in Fig. 1. The leaders of the groups search for optimum solutions. The swarm (group birds) is called as clan (group of people), and every clan (group of members) have solutions produced by the members of group. The best member of every clan is treated as the leader of that clan. A new leader of the clan is selected every generation due to the changing of positions of the members for meeting participation. The inertia function control exploration (global best) presents the algorithm. The main task of clan is to manage and control other members of the clan for movement and time control. Based on the new available position, every leader meet at one room (location) to find the best leader.

Fig. 1 Swarm optimization of clan with individual members

PSO algorithm is used for balancing the position for exploration as well as to find exploitation. Based on PSO, the Optimization of Meeting Room Method (OMRM) is performed. İn this, every member in the clan is equal to every particle in swarm. The position as well as velocity is updated by PSO. The proposed algorithm's step-wise performance, at first step shows to select the best leader (in a group of clan) who is treated as local best. In second step, overall best leader (in clan groups) is selected and is treated as global best. The optimization is determined by OMRM in the meeting room as in Fig. 2. By executing this algorithm, those who are selected as the overall best leaders, share their position with other leaders.

$$iWeight^{nl} + \left(\frac{iWeight^{gl} - iWeight^{nl}}{iteration_{\max}}\right) x\, rand() = \tag{8}$$

$$\begin{aligned}(time_t + 1) &= iweight^{nl}\, x\, Velocity_i^{nl}(t)\\ &+ rvaluecon(Particle_{global}^{Leader} - Particle_{local}^{Leader}(time_t))\end{aligned} \tag{9}$$

$$particle_i^{nl}(\mathrm{time_t} + 1) = particle_i^{nl}(\mathrm{time_t} + 1) + Velocity_i^{nl}(time_t) \tag{10}$$

In above equation, nl denotes the normal_leaders, gl represents the overall global leader, partile i states the position of the normal_leaders, $Velocity_i^{nl}$ states the velocity of normal leaders, $iWeight^{gl}$ and $iWeight^{nl}$ denote inertia weight of global leader and normal leader respectively.

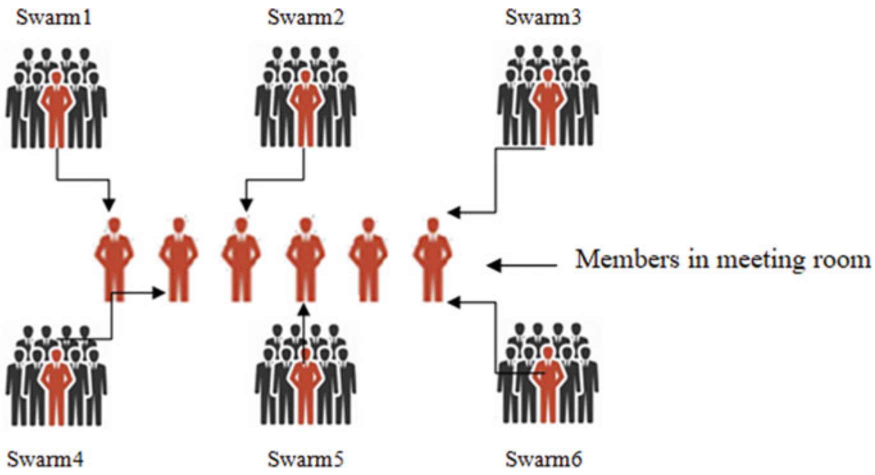


Fig. 2 Meeting room approach

Algorithm for Improved Multi-Particle Swarm Optimization

See Fig. 3**.**

4 Results and Discusssion

The performance of Improved Multiple Particle Swarm Optimization (IMPSO) is evaluated using benchmark functions compared with MPSO, Master–Slave SPO (MSPO) [11], and Multi-swarm Cooperative Particle Swarm optimizer (MCPSO) [17]. Parameters used are same for all these algorithms. The number of iterations used is 500 with 50 dimensions. For every experiment, 30 runs are performed. Testing with function is shown in Table 1 and parameter values in Table 2.

Table 3 indicates the performance optimum results of particles for benchmark functions. EMPSO perform best results when compare to other algorithms. IMPSO has 6 swarms, every group consists of 10 particles and 6 particles in the room. IMPSO has less computation and shows better performance for optimum solutions.

In Fig. 4a, by applying Sphere function, simulation results of IMPSO shows better performance when compared with other algorithms.

In Fig. 4b, by applying Griewank function, simulation results of IMPSO shows better performance than that of MCPSOS, SPSO, MPSO, IMPSO [12].

```
Input
#swarm, $\#particle_i$, con1, con2,#dimension,#maxgen
Output
Best solution(leader)
Procedure
Start
   Swarm initialization
   For each particle evaluate fitness
   while(iteration$\leq$ #maxgen)
      For every s in swarm
           for every swarm member s
            For updation use equation3 for velocity of member in swarm
                       For updation use equation2 for position of member in swarm
           next
           select local best as leaders
Next
Among all leaders select the best leader
For updation use equation6 for inertia weight
For updation use equation7 for leaders velocity
For updattion use equation8 for leaders position
Selection of Best_leader for global_best
Loop
Return to best_leader
Stop
```

Fig. 3 Improved multi particle swarm optimization algorithm

Table 1 Testing with function

Function name	Function
Sphere unimodal	$f_1 = \sum_{i=1}^{D} X^2$
Griewank unimodal	$f_2 = \sum_{i-1}^{D} -x^2_{1/4000} - \prod_{i=1}^{D} \cos \frac{x_i}{\sqrt{i}}$

5 Conclusion

The paper shows the importance of Swarm Intelligence algorithms for solving real-time problems. Swarm Intelligence algorithms are related to Nature Inspired Computing algorithms. The social group behavior of human being is proposed by applying swarm intelligence algorithm. The proposed algorithm Improved Multi-purpose particle swarm optimization manages the local search (exploitation) as well as the global search (exploration) techniques. By applying benchmark functions, the results of the extensible multi purpose swarm optimization shows better performance

Table 2 Parameter values

Algorithm	Parameter	Value
SPSO	W (inertial weight)	0.9–0.4
	Number of swarms	1
	Swarm size (population)	50
	c1, c2 (velocity constants)	1.5
MCPSO	W (inertial weight)	0.9–0.4
	Number of swarms	1
	Swarm size	50
	c1, c2(velocity constants)	1.5
MSPO	Wln	0.8–0.5
	Wlg	0.9–0.7
	c1,c2(velocity constants)	1.5
	Number of clans	5
	Clan size	10
IMPSO	Inertianl	0.8–0.9
	Inertiagl	0.9–0.9
	Constant1, constant2	1.5
	Number of clans	6
	Clan size	10

Table 3 Benchmark test functions' results

Fn	Swarm	Algorithm	Best
F1	50	SPSO	2.5457521
		MCPSO	0.9854126
		MPSO	0.0007845
		EMPSO	0.0006555
F2	50	SPSO	0.0884741
		MCPSO	0.0078414
		MPSO	0.0000897
		IMPSO	0.0000500

than the existing swarm optimization algorithms. The future work plans to implement hybrid nature inspired computing algorithms for optimum results.

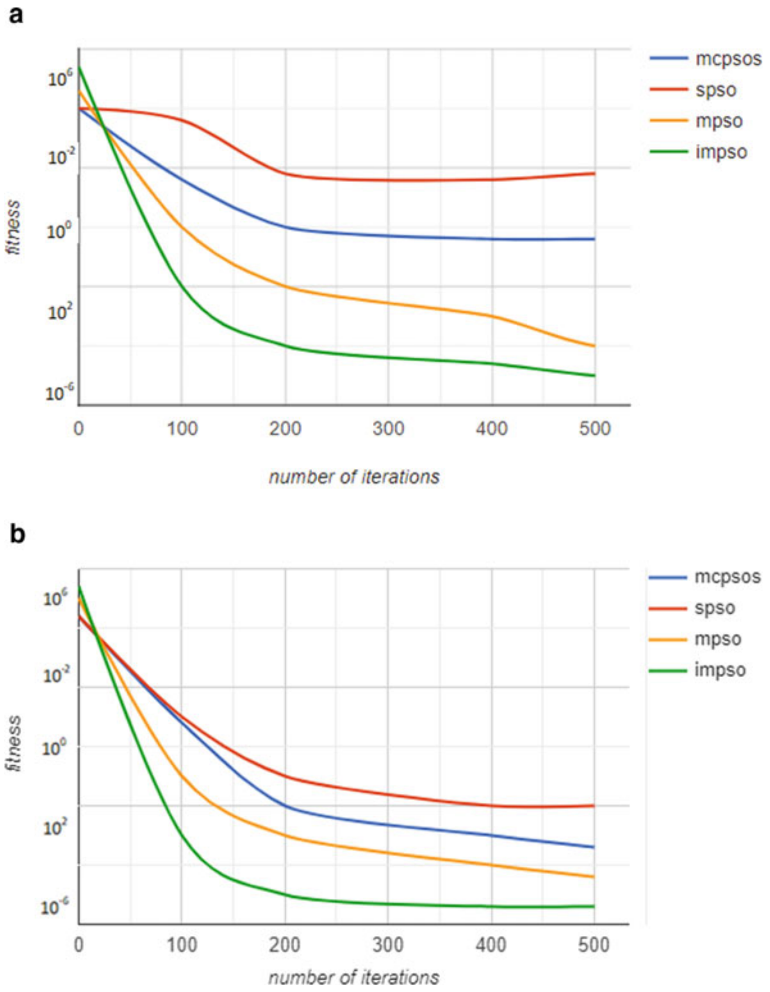


Fig. 4 **a** Convergence curve using sphere function, **b** convergence curve using Griewank unimodal function

References

1. Binitha S, Sathya SS (2012) A survey of bio inspired optimization algorithms. Int J Soft comput Eng 2(2):137–151
2. Dixit M, Upadhyay N, Silakari S (2015) An exhaustive survey on nature inspired optimization algorithms. Int J Softw Eng Appl 9(4):91–104
3. Slowik A, Kwasnicka H (2018) Nature inspired methods and their industry applications—swarm intelligence algorithms. IEEE Trans Ind Inform 14:1004–1015
4. Diao R, Shen Q (2015) Nature inspired feature selection meta-heuristics. Artif Intell Rev 44:311–340

5. Azrag MAK, Kadir TAA, Odili JB, Essam MHA (2017) A global African Buffalo optimization. Int J Softw Eng Comput Syst 3:138–145
6. Odili JB, Kahar MNM, Anwar S (2015) African Buffalo optimization: a swarm-intelligence technique. Procedia Comput Sci 76:443–448
7. Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In: Proceedings of the 6th ınternational symposium on micro machine and human science 1995, MHS 1995, pp 39–43
8. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of IEEE ınternational conference on neural networks 1995, vol 4, pp 1942–1948
9. Zhang Y, Wang S, Ji G (2015) A comprehensive survey on particle swarm optimization algorithm and its applications. Math Probl Eng
10. Kameyama K (2009) Particle swarm optimization-a survey. IEICE Trans Inf Syst 92(7):1354–1361
11. Niu B, Zhu Y, He X, Wu H (2007) MCPSO: a multi-swarm cooperative particle swarm optimizer. Appl Math Comput 185:1050–1062
12. Vijayakumar T, Vinothkanna R (2020) Efficient energy load distribution model using modified particle swarm optimization algorithm. J Artif Intell 2(4):226–231
13. Korani W, Mouhoub M (2021) Review on nature-inspired algorithms. In: Operations research forum, vol 2, no 3. Springer International Publishing
14. Kumar A, Nadeem M, Banka H (2022) Nature inspired optimization algorithms: a comprehensive overview. Evolving Syst 1–16
15. Shi Y, Eberhart R (1998) A modified particle swarm optimizer. In: 1998 IEEE ınternational conference on evolutionary computation proceedings. IEEE World Congress on Computational Intelligence, pp 69–73
16. Eberhart RC, Shi Y (2001) Tracking and optimizing dynamic systems with particle swarms. In: Proceedings of IEEE congress on evolutionary computation, pp 94–97. IEEE, Seoul
17. Shi Y, Eberhart RC (1999) Empirical study of particle swarm optimization. In: 1999 IEEE ınternational conference on evolutionary computation proceedings. IEEE World Congress on Computational Intelligence, vol 3, pp 1945–1950

Blockchain Based Secure, Efficient, and Scalable Platform for the Organ Donation Process of Healthcare Industry

Keyur Parmar, Vadlapudi Devanand Kumar, Neduri Leela Prasanth, Pranoppal, Kasa Charan Teja, Shriniwas Patil, and Kaushal A. Shah

Abstract Organ transplantation is one of the most effective medical procedures to save lives. An individual's organs can save up to nine lives. However, individuals refuse to donate organs because of lack of awareness and trust in the procedure, leading to the reduction in the number of organ donors. Individuals who wish to donate organs have to go through a complex administrative process, and sometimes these donated organs are managed by unauthorised individuals. To encourage individuals who wish to donate organs, we need a secure, efficient, and scalable platform. In this article, we present our perspective on the blockchain based organ donation management, in particular, for organ donation between organ donors and patients. The proposed platform uses the smart contract to automate the organ donation process and reduces the overall time of organ donation process. The proposed blockchain-based Organ Donation Platform (ODP) help patients in finding a matching donor efficiently. The ODP facilitates the process of organ donation by a decentralized network ensuring security, integrity, and transparency that eliminates the intermediaries. We comparatively evaluate the performance of the proposed ODP with the state-of-the-art literature. The proposed ODP is not only secure and scalable, but also efficient and reliable to find matching donor without revealing their identities.

Keywords Blockchain · Smart contract · Ethereum · Security · Public key cryptography · Organ donation · Healthcare · Transparency

1 Introduction

Organ transplantation is a medical process to remove a non-functional organ from the patient and then surgically replace it with a new organ taken from a healthy person.

K. Parmar (✉) · V. D. Kumar · N. L. Prasanth · Pranoppal · K. C. Teja · S. Patil
S. V. National Institute of Technology, Surat, India
e-mail: keyur@coed.svnit.ac.in

K. A. Shah
Pandit Deendayal Energy University, Gandhinagar, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_13

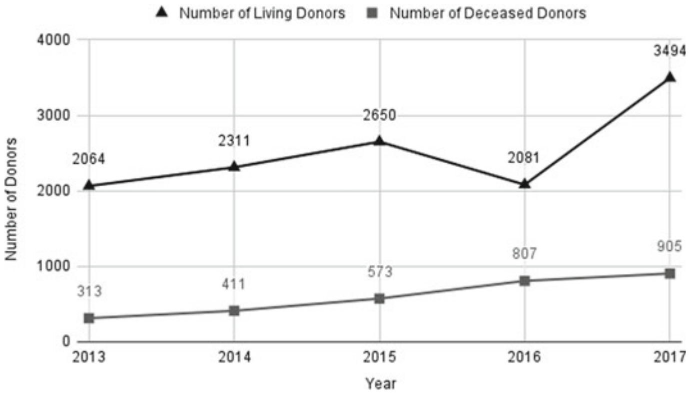


Fig. 1 Number of living and deceased organ donations from 2013–2017 [1]

Organ transplantation is one of the successful treatments for patients. Patients in need of organs often die due to a lack of information about donors in time.

Organ donation process is categorised into living organ donation and deceased organ donation. In living organ donation, organs are retrieved from a healthy living person and transplanted into patients. In deceased organ donation, the functioning organs are retrieved from a deceased person and transplanted into patients. The donation trend for living donors and deceased donors between the years 2013–2017 is shown in Fig. 1. The trend of donating organs is more in the case of living donors. Nevertheless, due to a lack of public awareness, organ donations in the case of deceased donors is less compared to living donors. The lack of public awareness results in the acute shortage of organs that otherwise can be used to save lives. Therefore, there is need to create awareness about the organ donation.

The Government of India organises organ donation programs such as National Organ Tissue Transplant Organization (NOTTO) [2] and Regional Organ Tissue Transplant Organization (ROTTO) [2]. Although the Government of India organises organ donation programs, individuals hesitate to donate organs due to the lack awareness about the of current organ donation programmes. Therefore, there is a need of transparent, secure, and a scalable platform for the organ donation process.

In organ donation, the challenge is to find a matching donor. The number of patients die because of an unavailability of a donor [1, 3]. Another challenge is the lengthy paperwork that delays the whole process of organ transplant. A non-relative organ donor has to go through a number of procedures, such as, getting consent from the state government or local hospital. Moreover, there is a lack of infrastructure for organ donation which makes the whole process complex even if the donors are available.

The above challenges are tackled by creating a decentralised system for the organ donation process. The proposed ODP will help individuals who wish to donate their organs and patients who are seeking organs.

In this article, we use the blockchain, a distributed ledger to store the data and to provide tamper resistance. Blockchain technology also guarantees security and transparency [4]. Therefore, a number of individuals will join the ODP who wish to donate organs. The proposed work uses the smart contract [5–7] with Ethereum Network [8, 9] to make the organ donation process automatic and fast. The smart contract reduces the time required to complete due formalities.

The proposed work ensures transparency between donors and recipients (patients) and helps patients in finding the vital organ. The proposed work solves the challenge of finding organ donors and provides a choice for patients to choose organ donors. The ODP is trustworthy because all records are maintained using blockchain. The ODP eliminates the illegal smuggling of organs by restricting illegal access to ODP. The ODP will help to save lives by connecting the suitable donor to the right patient at the right time.

In this article, our contributions are as follows:

1. We present a comprehensive literature review to identify the challenges in the conventional organ donation process.
2. We comparatively evaluate the strengths and weaknesses of the state-of-the-art proposals in the domain of organ donation.
3. We propose a blockchain based platform to facilitate patients to find a matching donor. The proposed blockchain based platform helps in mitigating the issues found in conventional organ donation process.
4. We implement the proposed blockchain based platform to demonstrate the working of the proposed ODP.
5. We evaluate the security and performance of the proposed blockchain based platform to highlight its strengths and weaknesses.

The rest of the article is organized as follows: In Sect. 2, we discuss the related work. In Sect. 3, we discuss the preliminaries related to the proposed system. Section 4 presents the proposed ODP. Section 5 presents the experimental implementation of the proposed ODP. In Sect. 6, we conclude the article by emphasizing our contribution.

2 Related Work

Organ donation is a well-researched topic that provides patients and organ donors a common platform. Chen et al. [10] identified an issue in a rule-based strategy of prioritising kidney donation in kidney donation programs where exchanges are initiated by Altruistic Donors (ADs). Authors developed a software-based decision support system. In addition, authors considered and compared two graph-based organ allocation algorithms, namely, MEU-Parallel and MEU-Sequential [10], to find the correct match. However, the authors work does not emphasize more on other kidney-allocation algorithms.

Prajapati et al. [11] implemented an online system to manage blood and organ transplant. The online system helps in searching for blood in case of emergency and in maintaining the records of blood and organ donors.

Kadam et al. [12] discuss the drawbacks of manual blood donation and the organ donation process. Authors proposed an electronic donation system that manages the records of blood and organ donors and enables patients to monitor and find the correct donor.

Rastogi and Tiwari [13] implemented an online web-based system for organ donation and management. An administrator manages the system. The administrator holds the rights and privileges to print the organs list. However, the system depends on the administrator which makes the system centralised and vulnerable to insider threats. Similarly, Ali et al. [14] implemented a web application with the support of a mobile application for a blood donation management system.

The existing systems manage organ donations, but these systems require a connection between organ donors, patients, and central trusted third-parties. The system that uses the central trusted third-party such as hospital agents or government agents is vulnerable to insider and outsider threats and discloses the organ donor's identity. If anonymity is not provided to organ donors and patients, the central third-party can contact patients to prioritise the patient to receive organs by asking for extra money.

Dajim et al. [15] implemented a decentralised organ donation application. Authors used blockchain technology [16, 17] to distribute the information. The use of blockchain makes the system secure and tamper-proof. However, anonymity is not provided for organ donors and patients. Rajan et al. [18] proposed a system for organ donation and transplantation using blockchain technology. Authors used smart contracts to reduce the cost of transactions. However, the system does not allow anyone to check the availability of organs. The existing systems fail to provide security, transparency, a tamper-resistance, and do not protect patients from insider and outsider adversaries.

Lakshminarayanan et al. [19] identified the dynamic update issue in blood management system. The authors proposed and implemented blockchain-based blood management system using hyperledger fabric framework to track the blood trail. The challenge of tracing thee blood trail is modelled as a supply-chain management issue. The proposed system provides transparency in blood donation process. However, the use of hyperledger fabric framework makes system permissioned. Therefore, no one can access system to get blood donation related information without getting permission from the network. Similarly, Quynh et al. [20] proposed blockchain-based innovative system to manage blood information. Authors proposed the system based on the architecture of hyperledger fabric with support of supply and demand of blood.

Hawashin et al. [21] proposed a private Ethereum blockchain-based solution to enable organ donation and transplantation management. The use of private blockchain makes system decentralized, secure, traceable, auditable, private, and trustworthy. However, the private blockchain makes the organ donation and transplantation not visible to people outside the organization.

3 Preliminaries

In this section, we briefly discuss the Blockchain [16, 17], Smart Contracts [5–7], and Ethereum [22].

3.1 Blockchain

"A blockchain is a continuously growing list of records called blocks, which are linked and secured using cryptography [23]." The idea to timestamp [24] all transactions is to maintain the order of transactions. In a blockchain, the new blocks are appended securely using cryptographic hash functions.

One of the fundamental features of the blockchain network is that it has no central authority. The blockchain network is shared among all the nodes and each node has a ledger or its own database, and the same copy is maintained by all nodes of the network. Therefore, any data present in the blockchain is secure, transparent, and tamper-proof [25].

3.2 Ethereum

Ethereum [22] is a stable platform that can be used to codify and exchanges based on contracts. Ethereum is an open-source, blockchain-based platform for creating and running safe smart contracts and distributed applications (DApps) [26].

3.2.1 Infura

Infura provides the tools and infrastructure with stable access to Ethere-um and InterPlanetary File System (IPFS). Infura enables developers to test and deploy its scaled blockchain applications.

3.2.2 Metamask

Metamask is a browser extension that enables users to create decentralised applications (DApps) [26] without using the Ethereum network as an Ethereum node. Metamask is also serves as wallet that stores the Ethers (cryptocurrency) and allows users to send and receive Ethers via a decentralised application (DApp) [26]. Metamask allows us to link Infura, another Ethereum node, and runs smart contracts on the node.

3.2.3 Web3.js

Web3.js is a collection of libraries that enables an HTTP or IPC connection to communicate with a local or remote Ethereum node. The Ethereum blockchain is accessed via the web3 JavaScript library. Web3.js can perform operations such as accessing user accounts, submitting transactions, and communicating with smart contracts. We use Web3 Truffle-HDWallet-provider. The Truffle-HDWallet-provider provides a convenient network to link Ethereum thro-ugh infura.io. For example, the HDWallet provider adds some truffle-required features like event filtering and transaction signing that are unavailable with Infura.

3.3 Smart Contracts

Smart contracts [5–7] are self-contained programmes that run on the blockchain network. Smart contracts are programmed to perform actions on blockchain as per the business logic and requirements [27]. The smart contract executes functions if certain code and business logic conditions are met or not met.

4 The Proposed Organ Donation Platform Using Blockchain Technology

We propose an efficient, secure and scalable platform to facilitate donors and patients to manage organ donation. The platform using which both donor and patient will be able to manage all the operations such as login and signup. Donors can provide the required information of the organs which he or she wish to donate, patients can also find the donors by searching for the required organ.

We present the whole process of the ODP as shown in Fig. 2. There are two entities in the system, namely, a donor and a recipient (patient). The entities are connected to

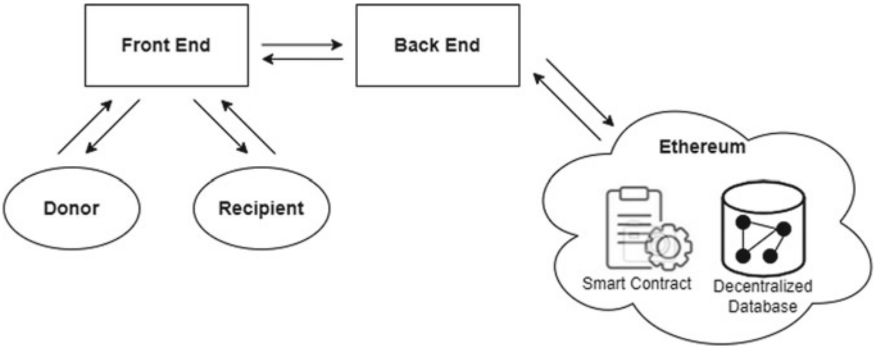


Fig. 2 Proposed organ donation platform using smart contracts with ethereum

the platform through front-end service where they can easily perform the operations such as login, signup, and search for organs. The back-end service executes the required operations requested through the front end, and all records are appended to the blockchain. The use of smart contracts [5–7] help in automating the organ donation process.

The algorithm is used to find the available donors for the required organs. The proposed algorithm to retrieve all donors with the same organ as recipient is as shown in Algorithm 1.

Algorithm 1 Algorithm to retrieve all donors with the same organ as recipient

```
Result: Donor details
Query database for all donors where recipient.organ= donor.organ;
if results.length >0 then
    Show available donors;
    From donors list select a donor based on city;
    View donor details to contact;
else
    Print("No donors available");
end
```

Next, we present another algorithm to find the correct donor for the patient to receive an organ and make the process more efficient so the patient can perform the transplantation in the available time. The algorithm to match donor and recipient based on location is presented in Algorithm 2.

Algorithm 2 Algorithm to match donor and recipient based on location

```
Result: Match donor and recipient so that they can contact each other
Function matchDonorRecipient(address donorId, address recipientId):
    public payable
    if match found for both donor and recipient then
        retrieve donor info using donor address;
          retrieve recipient info using recipient address;
    end
End Function
```

The proposed ODP provides transparency by adding donors on the blockchain that makes all the donors visible to anyone on the blockchain network with cryptoghraphically generated keys. The cryptographically generated keys are used to protect the privacy of the donors. The algorithm to add a donor to the blockchain is shown in Algorithm 3.

Algorithm 3 Algorithm to add donors to the blockchain

```
Result: Add Donor to the Blockchain
Each donor has the following attributes
 $donorId, hash, organ, blood\_group, matchfound, exist$;
Validate donor credentials;
Call function Add_donor;
Function Add_donor ($donorID, memory\ hash, organ\_name, bgroup, factor$):
    public checkdonorexist(donorID):
     if $True$ then
        $Donors[donorID] \longleftarrow donor(donorID, hash, bgroup, organ\_name, false,$
          $true)\ donor\ array.push(donorID)$;
    end
End Function
```

5 Results and Analysis

In this section, we discuss the implementation of the proposed ODP. We also discuss each functionality implemented in detail. The ODP consist of three layers as follows:

1. HTML, CSS, Javascript is used as a front end user interface for the organ donor and recipient to provide all the functionalities.
2. A node.js server is used to interact with the Ethereum blockchain network and provides the required data to Javascript Frontend.
3. All the respective organ donor and recipient data are stored in a decentralised network using Ethereum.

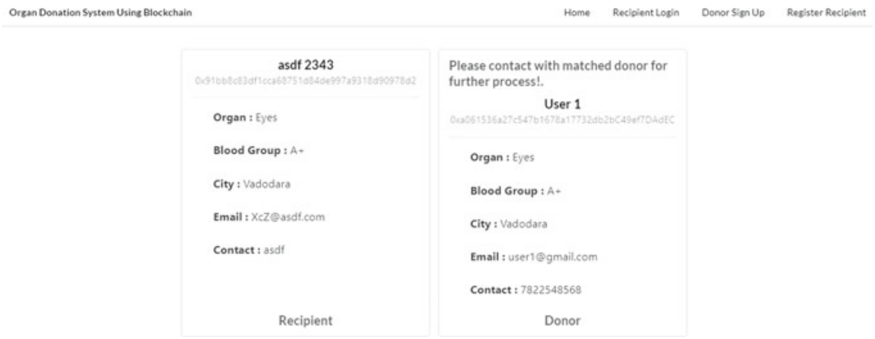
The recipient logs in with the public key provided after registering. The credentials are to be provided by the user while signing up. The details of both donors and recipients are available in the blockchain ledger. Therefore, the proposed ODP can verify whether it is an existing user or not, if yes, then the system redirects the same to the login page otherwise, on the signup page.

There are two separate portals for both donors and patients while registering on the proposed platform. Organ donors will be asked for his or her basic information and also the list of organs donors wish to donate after filling out a consent form.

A patient will be able to register with his or her basic information and the organ that is required by the patient along with his contact information as a consent form.

Patients verifies the matched donor in the dashboard provided when patients log in, and when there are no matches, the dashboard shows the message, “no matches found” and suggests the active donor’s list as shown in Figs. 3 and 4, respectively.

In order to verify the security strength of the proposed platform, we compare the proposed platform with similar proposals in the domain of organ donation. The comparison is given in Table 1. The comparison is based on the well-known security services provided by existing proposals for organ donation. Based on the comparison of different organ donation proposals, in our opinion, the existing proposals only provides a few security services. However, in the proposed ODP, the confidentiality is achieved using PKC [28], the integrity of an information is verified by



asdf 2343
0x91bb8c83df1cca68751d84de997a9318d90978d2

Organ : Eyes

Blood Group : A+

City : Vadodara

Email : XcZ@asdf.com

Contact : asdf

Recipient

Please contact with matched donor for further process!.

User 1
0xa061536a27c547b1678a17732db2bC49ef7DAdEC

Organ : Eyes

Blood Group : A+

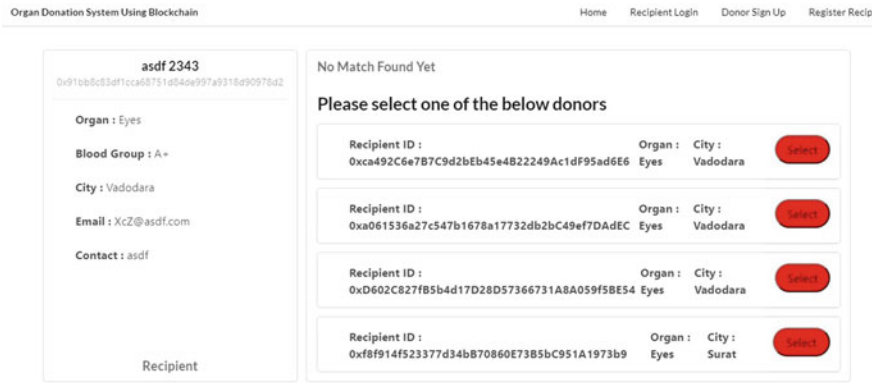
City : Vadodara

Email : user1@gmail.com

Contact : 7822548568

Donor

Fig. 3 Dashboard when donor is matched with recipient



asdf 2343
0x91bb8c83df1cca68751d84de997a9318d90978d2

Organ : Eyes

Blood Group : A+

City : Vadodara

Email : XcZ@asdf.com

Contact : asdf

Recipient

No Match Found Yet

Please select one of the below donors

Recipient ID :	Organ :	City :
0xca492C6e7B7C9d2bEb45e4B22249Ac1dF95ad6E6	Eyes	Vadodara
0xa061536a27c547b1678a17732db2bC49ef7DAdEC	Eyes	Vadodara
0xD602C827fB5b4d17D28D57366731A8A059f5BE54	Eyes	Vadodara
0xf8f914f523377d34bB70860E73B5bC951A1973b9	Eyes	Surat

Fig. 4 Dashboard when no donor is matched with recipient

Table 1 Comparison of different organ donation systems

Authors	A	B	C	D	E	F	G	H
Chen et al. [10]	×	×	×	×	×	×	✓	✓
Prajapati et al. [11]	✓	×	×	×	×	×	×	×
Kadam et al. [12]	×	×	×	×	✓	×	×	✓
Rastogi and Tiwari [13]	×	×	×	×	✓	×	×	✓
Ali et al. [14]	×	×	×	×	✓	×	×	×
Dajim et al. [15]	×	×	✓	✓	×	×	×	✓
Ranjan et al. [18]	×	✓	✓	✓	×	✓	×	✓
Proposed Scheme	✓	✓	✓	✓	✓	✓	✓	✓

A awareness, *B* confidentiality and privacy, *C* decentralization, *D* tamper-resistance, *E* transparent, *F* automation, *G* algorithms to find match, *H* eliminate trusted third-party such as organ banks comparing message digest of the information received along with encrypted information and with message digest of decrypted information. Privacy is achieved using an unique cryptographically generated key and it is unrelated to the identity of the users. Decentralization, tamper-resistance, and transparency are provided using blockchain technology [16, 17]. The protection from insider and outsider adversaries is achieved using timestamp. The platform is automated using smart contracts [5–7].

6 Conclusions and Future Works

In this article, we propose a platform for organ donation using blockchain technology. To protect organ donors and patients from adversaries, we need a platform that provides transparency in the organ donation system. The ODP uses blockchain technology to make the central trusted third-party, e.g. hospital agent or government agent, more transparent and accountable while managing organ donation. The ODP is trustworthy and significantly reliable. In addition, the ODP protects organ donors and patients from insider and outsider adversaries. As a future work, there is a possibility to use the optimised algorithm that supports multiple cities and hospitals as an authority to verify the donors and patients.

Acknowledgements This research was a part of the project "Design and Analysis of Secure and Efficient Smart Contracts Using Blockchain Technology". It was partially supported by the SEED Money/Research Grant of the author, Dr. Keyur Parmar, Department of Computer Science and Engineering, S. V. National Institute of Technology (NIT), Surat, India.

References

1. Ahammed Mekkodathil MA, Sathian B, Rajesh E, Kumar RN, Simkhada P, van Teijlingen E (2019) Current scenario of organ donation and transplantation in Kerala India. Nepal J Epidemiol 9(2):759. https://doi.org/10.3126/nje.v9i2.24679
2. National organ tissue transplant organization. https://notto.gov.in/download-forms.htm. Accessed 03 Apr 2022
3. Nallusamy S et al (2018) Organ donation-current Indian scenario. J Pract Cardiovasc Sci 4(3):177–177. https://doi.org/10.4103/jpcs.jpcs_59_18
4. Abhishek B, Panjanathan R, Sarobin VR, Raja BE, Narendra M (2022) Data security in e-health monitoring system. Mater Today Proc 62:4620–4628. http://orcid.org/10.1016/j.matpr.2022.03.079
5. Szabo N (1997) Formalizing and securing relationships on public networks. First Monday 2(9). https://doi.org/10.5210/fm.v2i9.548
6. Szabo N (1994) Smart contracts
7. Wang S, Ouyang L, Yuan Y, Ni X, Han X, Wang FY (2019) Blockchain-enabled smart contracts: architecture, applications, and future trends. IEEE Trans Syst Man Cybern Syst 49(11):2266–2277. https://doi.org/10.1109/TSMC.2019.2895123
8. Buterin V (2022) Ethereum: a next-generation smart contract and decentralized application platform. https://ethereum.github.io/yellowpaper/paper.pdf. Accessed 30 Apr 2022

9. Wood G et al (2014) Ethereum: a secure decentralised generalised transaction ledger. Ethereum Proj Yellow Pap 151:1–34
10. Chen Y, Li Y, Kalbfleisch JD, Zhou Y, Leichtman A, Song PXK (2012) Graph-based optimization algorithm and software on kidney exchanges. Trans Biomed Eng 59(7):1985–1991. https://doi.org/10.1109/TBME.2012.2195663
11. Prajapati Mayur, Dungar Bhati YSHJ (2017) Online blood and organ transplant management system. Int J Eng Sci Res Technol 6(3):17–21. https://doi.org/10.5281/zenodo.266682
12. Nikhil Kadam N, Akshay Kaware VG (2020) Blood and organ donation system. Int J Sci Res Eng Dev 3(2):44–48
13. Rastogi S, Tiwari S (2019) Online organ donation management system. Int J Sci Res Eng Dev 5(10):445–448
14. Ali A, Jahan I, Islam A, Parvez S (2015) Blood donation management system. Am J Eng Res 4(6):123–136
15. Dajim LA, Al-Farras SA, Al-Shahrani BS, Al-Zuraib, AA, Merlin Mathew R (2019) Organ donation decentralized application using blockchain technology. In: Proceedings of the international conference on computer applications information security (ICCAIS), Riyadh, Saudi Arabia. IEEE, pp 1–4. https://doi.org/10.1109/CAIS.2019.8769459
16. Di Pierro M (2017) What is the blockchain? Comput Sci Eng 19(5):92–95. https://doi.org/10.1109/MCSE.2017.3421554
17. Nofer M, Gomber P, Hinz O, Schiereck D (2017) Blockchain. Bus Inf Syst Eng 59(3):183–187. https://doi.org/10.1007/s12599-017-0467-3
18. Ranjan P, Srivastava S, Gupta V, Tapaswi S, Kumar N (2019) Decentralised and distributed system for organ/tissue donation and transplantation. In: Proceedings of the conference on information and communication technology, Allahabad, India. IEEE, pp 1–6. https://doi.org/10.1109/CICT48419.2019.9066225
19. Lakshminarayanan S, Kumar P, Dhanya N (2020) Implementation of blockchain-based blood donation framework. In: Proceedings of ICCIDS: international conference on computational intelligence in data science, Chennai, India. Computational intelligence in data science. Springer, pp 276–290. https://doi.org/10.1007/978-3-030-63467-4_22
20. Quynh NTT, Son HX, Le TH, Huy HND, Vo KH, Luong HH, Tuan KNH, Anh TD, Duong-Trung N et al (2021) Toward a design of blood donation management by blockchain technologies. In: Proceedings of ICCSA: international conference on computational science and its applications. Lecture notes in computer science. Springer, Cagliari, Italy, pp 78–90. https://doi.org/10.1007/978-3-030-87010-2_6
21. Hawashin D, Jayaraman R, Salah K, Yaqoob I, Simsekler MCE, Ellahham S (2022) Blockchain-based management for organ donation and transplantation. IEEE Access 10:59013–59025. https://doi.org/10.1109/ACCESS.2022.3180008
22. Buterin V (2022) On public and private blockchains. https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains/. Accessed 28 Apr 2022
23. Vos J (2017) Blockchain and land administration: a happy marriage? Eur Property Law J 6(3):293–295. https://doi.org/10.1515/eplj-2017-0018
24. Haber S, Stornetta WS (1990) How to time-stamp a digital document. In: Proceedings of CRYPTO 90: advances in cryptology, conference on the theory and application of cryptography, Santa Barbara, USA. vol 537. Lecture notes in computer science. Springer, pp 437–455. https://doi.org/10.1007/3-540-38424-3_32
25. Nakamoto S (2008) Bitcoin: a peer-to-peer electronic cash system. Decentralized Bus Rev
26. Metcalfe W et al (2020) Ethereum, smart contracts, dapps. Blockchain Crypt Currency 77. https://doi.org/10.1007/978-981-15-3376-1_5
27. Bashir I (2018) Mastering blockchain: distributed ledger technology, decentralization, and smart contracts explained, 2nd edn. Packt Publishing Ltd, ISBN: 978-1-78883-904-4
28. Al-Riyami SS, Paterson KG (2003) Certificateless public key cryptography. In: Proceddings of ASIACRYPT: advances in cryptology, international conference on the theory and application of cryptology and information security, Taipei, Taiwan. vol 2894. Springer, pp 452–473. https://doi.org/10.1007/978-3-540-40061-5_29

Image Enhancement in Frequency Domain Fingerprint Detection and Matching Approach

Suhasini S. Goilkar and Shashikant S. Goilkar

Abstract In digital image processing image enhancement techniques are used to improve the observation, perception and interpretability of the image information through human visual system. Image enhancement is implemented in spatial domain approach that operates directly to the pixels and frequency domain approach that operates through Fourier transform of an image. Image enhancement techniques in frequency domain are useful in different fields like early detection of physiological disorder, remote sensing, forensic science and biometric science. This implementation is done in frequency domain approach that means sharpening, smoothing and homomorphic filters are designed, implemented and image is enhanced to give the better input to the image processing automated techniques like biomedical, recognition and matching applications. In this research work, frequency domain Fourier transform techniques are designed and implemented for fingerprint detection and matching social applications. The frequency domain analysis performance is measured with performance measures which are peak signal to noise ratio and contrast to noise ratio with mean and variance. The fingerprint detection and matching progression technique is usually disintegrated into image pre-processing, matching and extraction. The designed, implemented and presented results of this research work will be very useful in social forensic science fingerprint matching different applications to improve the accuracy before applying to matching process.

Keywords Fingerprint image · Frequency domain transforms · Image enhancement · Peak signal to noise ratio · Contrast to noise ratio

1 Introduction

Now a days everywhere we are using digital techniques and to prefer digital techniques so many advantages are there as compared to analog techniques. This is clear and understandable that to prefer the image signal processing digitally for achieve the

S. S. Goilkar (✉) · S. S. Goilkar
Finolex Academy of Management and Technology, Ratnagiri, Maharashtra, India
e-mail: spkashid@gmail.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_14

computer processing capabilities. Digital image processing is a very important tool to know and understand finer and finer details of any image. To process on the image digital signal with computer algorithms is nothing but the digital image processing. In digital image processing different techniques are used to perform different operations related to different applications. Image enhancement is the popular and important technique used for analysis and sharping of the different image features like contrasts, boundaries, edges of the digital image for the analysis and display purpose. The analysis and design of this approach is possible in time domain and frequency domain also. In frequency domain approach more emphasize of the image features is easily possible. From the input processing data, the intrinsic information contents will not increase in enhancement techniques, but the dynamic range definitely increased and because of that specific feature of the input image will be easily detected. In spatial domain image enhancement techniques from simple point processing to bit plane slicing total fourteen techniques are possible to use in the process of image enhancement, using these techniques possible to improve better and better human visual appearance of the original image for pre-processing analysis [1–3].

The highlighted image processing operations are gray level and contrast manipulation, edge sharpening and crispening, noise removal, magnification and interpolation etc. The image enhancement techniques are used to emphasize the silent features of the original data and simplify the processing task. In frequency domain image enhancement approach Fourier are used in transform and other frequency domain transforms are used for two dimensional images. Some transforms are used for the process to enhance the image visualization, feature selection, interest area measurements. Some transforms are used for image compression to reduce the data from original image with increased efficiency for the purpose of transmission and storage [4, 5].

In this paper, Image Enhancement in Frequency Domain Fingerprint Detection and Matching Approach is proposed and the comparative analysis is done with different cut off distance and filter order using PSNR and CNR parameters. The Section two highlights the important aspects of enhancement approach in frequency domain. In Section three adopted methodology is elaborated in detail. In section four designed and implemented filtering process steps are given. Finally, the simulated results are discussed in Section five and the entire work is concluded in Section six in Conclusion.

2 Frequency Domain Approach in Enhancement

In frequency domain approach is the method in which it modifies the Fourier transform of an image. That means first required to compute the Fourier transform of the image. After that with this Fourier transform of the image required to multiply with filter transfer function. To get the modified image now need to compute inverse Fourier transform [6, 7]. In this operation the significant key parameters are the filter transfer functions like lowpass filters, high pass filters and Butterworth filters.

In image enhancement in frequency domain first required to compute the Fourier transform of a particular image which we want to enhance and just multiply it with filter function, convolution is not required and finally take the inverse transform for modified image. In frequency domain approach, it is instinctively simple to understood for image enhancement or idea of image blurring and deblurring to reduce high frequency components and for image sharpening to increase the magnitude of high frequency components. In spatial domain approach these operations are also effective with convolution using spatial filters. [8–10]. In the process of lowpass filtering it eliminates the image high frequency components to reduce the noise and blurring. In ideal low pass filtering possible to recall all low frequency components and to eliminate all high frequency components. But here in this computation process ideal low pass filter suffers with blurring and ringing effects and introduced large number if ripples or waves in the output image. In this the Butterworth filter in frequency domain is used for better results and smooth transitions in the computational process. The image captured from camera, or any sensor contains the light which is reflected from the object which we want to capture the image [11]. Generally, in the basic nature of the image is categorized in to two components that is amount of light incident on the object that is illumination to view the object and the amount of light reflected from the object that is reflection. In image enhancement homomorphic filters are used to enhance illumination and reflection components separately [12, 13]. In frequency domain image enhancement analysis low pass or smoothing filter, high pass or sharpening filter, homomorphic filter and colour image. In frequency domain using filters possible to sharp, smooth, deblur and restore the image. Basically, filtering process means the convolution of a function with filter function. Filtering process in frequency domain approach is concise below in Fig. 1.

In the filtering process frequency domain approach, the significant points are to be understood that compute DFT transform of an image in frequency through filter and required to take inverse DFT to obtain filtered image [14, 15]. In frequency domain approach Fourier transform is applied for 2D images with different reasons

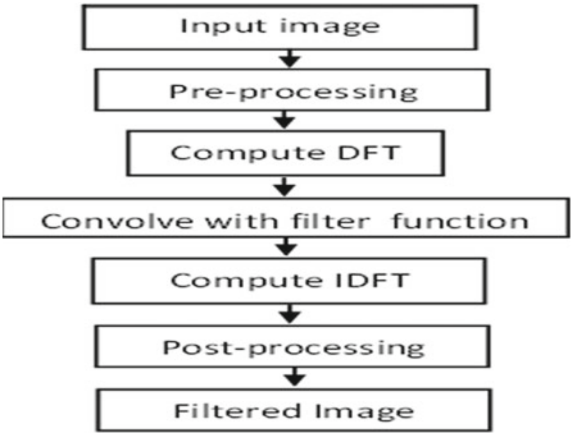


Fig. 1 Filtering process in frequency domain approach

like to enhance the perceptibility of the image, selection of any part of an image, area of interest and measurement image compression for storage and transmission with increased efficiency. For the transmission and storage purpose fast Fourier transform is used for its efficient and powerful algorithm [16–18].

After image enhancement process the quality of the image is evaluated with two techniques or parameters that is qualitative and quantitative. In quantitative technique to check the performance psnr and variance values are calculated [19–21].

3 Adopted Methodology

In this work fingerprint images are captured from fingerprint module and stored for further processing. For this fingerprint input stored image, the enhancement and filtering process in frequency domain is given in the block diagram shown below in Fig. 2.

As shown in the above block diagram first step is required to do the pre-processing on the input fingerprint images like convert in gray image, resize if required and format the image. Now apply the frequency domain image enhancement techniques that are smoothing filtering, sharpening filtering and homomorphic filtering. In this implementation Butterworth and gaussian filtering techniques are implemented given in detail in the next session. At the end after simulation resulted images are stored and PSNR and CNR parameters are calculated on the basis of these parameters comparison is done.

To calculate the peak signal to noise ratio matric alike signal to noise ratio having high values also shows the sign of more precise denoising.

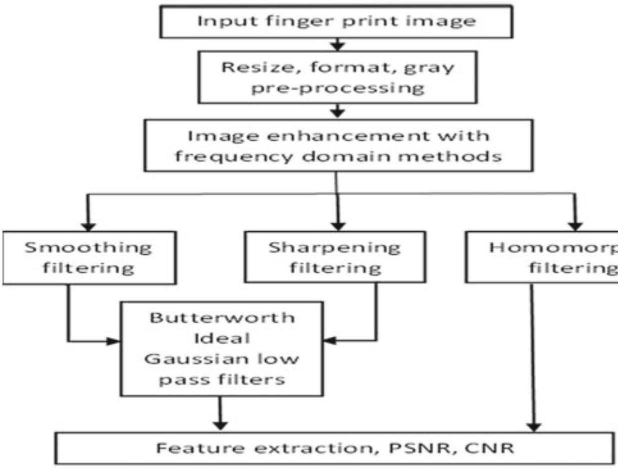

Input finger print image
Resize, format, gray pre-processing
Image enhancement with frequency domain methods
Smoothing filtering
Sharpening filtering
Homomorphic filtering
Butterworth Ideal Gaussian low pass filters
Feature extraction, PSNR, CNR

Fig. 2 Enhancement and filtering process in frequency domain approach

4 Design and Implementation

In the design and implementation of frequency domain algorithms the basic three parameters are considered that is in the Fourier transform edges and sharp changes mainly contribute with high frequency contents or low frequency contents. The smoothing is done with filtering high frequency contents and Butterworth, gaussian and ideal filters are used to enhance the fingerprint image.

A. Smoothing frequency domain algorithm steps

1. Capture the fingerprint image using scanner or module and convert to gray.
2. Resize and format.
3. Obtain FFT and compute filter transfer function for three filters.
4. Apply all three smoothing filters with different cut-off frequencies and filter orders.
5. Apply transfer function and do convolution operation.
6. Compute Inverse FFT and obtain the new enhanced image.
7. Calculate PSNR and CNR.

B. Sharpening frequency domain algorithm steps

1. Capture the fingerprint image using scanner or module and convert to gray.
2. Resize and format.
3. Obtain FFT and compute filter transfer function for three filters.
4. Apply all three sharpening filters with different cut-off frequencies and filter orders.
5. Apply transfer function and do convolution operation.
6. Compute Inverse FFT and obtain the new enhanced image.
7. Calculate PSNR and CNR.

C. Homomorphic filtering frequency domain algorithm steps

1. Capture the fingerprint image using scanner or module and convert to gray.
2. Resize and format.
3. Obtain logarithm function as Fourier transform of low frequencies as illumination and high frequencies as reflection.
4. Determine the filter function which compress the gray level range and enhance the contrast.
5. Apply transfer function and do convolution operation.
6. Compute Inverse logarithm and obtain the new enhanced image.
7. Calculate PSNR and CNR.

5 Simulation Results and Discussion

Finger print frequency domain image enhancement three techniques implemented and effectively completed. The image parameters like PSNR and CNR are extracted and the features are compared. In this implementation Butterworth filter, gaussian filter and ideal lowpass filters are used for fingerprint image enhancement. In these three-filter design filter order n = 4 and n = 8 is used and the cut-off distance D = 24 and D = 36 is used. The fingerprint enhanced images are shown below Fig. 3.

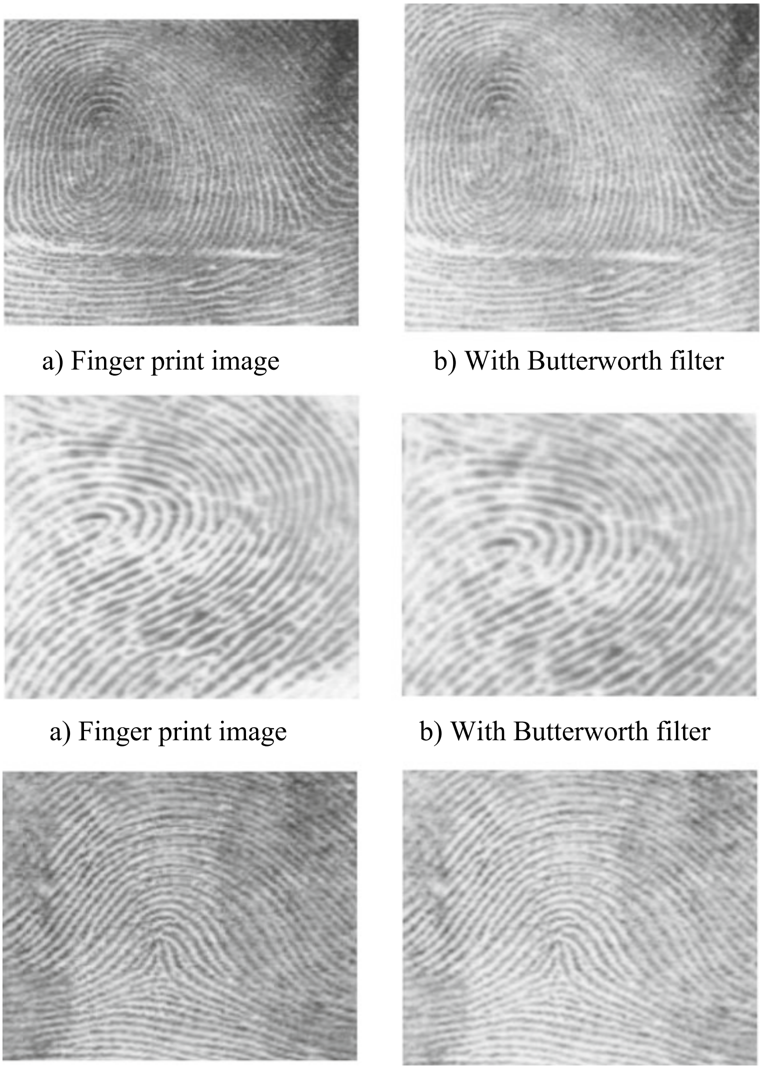
The performance parameters PSNR and CNR are calculated for three filters with different filter orders and cut-off distance as shown in Table 1.

For Parameter comparations the cutoff distance D = 24 and D = 36 is taken and the filter orders are n = 4 and n = 8 are used for all three filters. The Parameter comparation graphs of Butterworth filter, Gaussian filter and Ideal lowpass filter are shown below Fig. 4.

In the comparation graph different cut off distance, filter order, different filters are taken to compare and show the comparation of psnr and cnr parameters.

6 Conclusion

Image enhancement techniques are used to improve the image observation Image enhancement techniques in frequency domain are useful in different fields like early detection of physiological disorder, remote sensing, forensic science and biometric science. This implementation is done in frequency domain approach that means sharpening, smoothing and homomorphic filters are designed, implemented and image is enhanced to give the better input to the image processing automated techniques like biomedical, recognition and matching applications. The fingerprint detection and matching progression technique is usually disintegrated into image pre-processing, matching and extraction. The designed, implemented and presented results of this research work will be very useful in social forensic science fingerprint matching different applications to improve the accuracy before applying to matching process.



a) Finger print image

b) With Butterworth filter

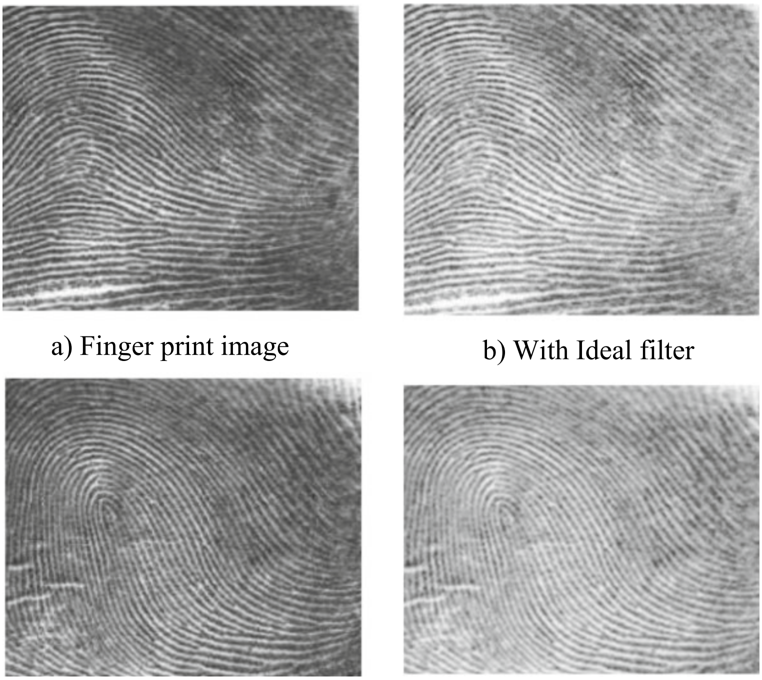
a) Finger print image

b) With Butterworth filter

a) Finger print image

b) With Gaussian filter

Fig. 3 Finger print images and enhanced with three different filters



a) Finger print image

b) With Ideal filter

a) Finger print image

b) With Ideal filter

Fig. 3 (continued)

Table 1 Parameter comparations with different filters

Filter type	Cut off distance	Filter order	PSNR	CNR
Butterworth filter	D = 24	n = 4	43.51	0.0067
		n = 8	43.49	0.0064
	D = 36	n = 4	43.72	0.0066
		n = 8	43.75	0.0063
Gaussian filter	D = 24	n = 4	43.69	0.0067
		n = 8	43.73	0.0065
	D = 36	n = 4	44.01	0.0064
		n = 8	44.09	0.0064
Ideal low pass	D = 24	n = 4	43.61	0.0058
		n = 8	43.62	0.0056
	D = 36	n = 4	43.82	0.0056
		n = 8	43.85	0.0055

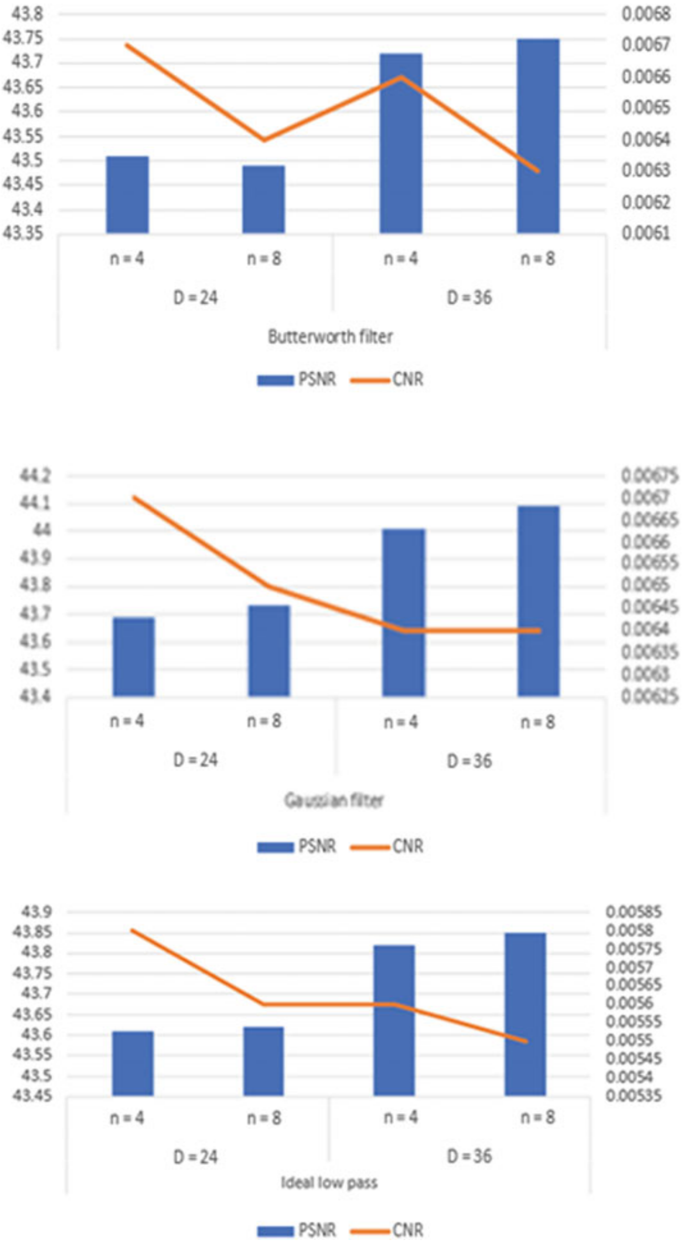


Fig. 4 Parameter comparation graphs of Butterworth filter, Gaussian filter and Ideal lowpass filter

Acknowledgements The authors acknowledge the financial support from University of Mumbai, Maharashtra, India through the Minor Research Grant for this Project.

References

1. Daluz HM (2015) Fundamentals of fingerprint analysis. Taylor & Francis Group
2. Ali MMH, Mahale VH, Yannawar P, Gaikwad AT (2016) Overview of fingerprint recognition system. In: International conference on electrical electronics and optimization techniques ICEEOT, pp 1334–1338
3. Adam EEB (2021) Evaluation of fingerprint liveness detection by machine learning approach-a systematic view. J ISMAC 3(01):16–30
4. Ahmed ZJ, George LE (2017) Fingerprints recognition using the local energy distribution over haar wavelet subbands. Int J Sci Res 6(9)
5. Wenchao W (2012) A Fingerprint identification algorithm based on wavelet transformation characteristic coefficient. ICSAI 2–4
6. Tang T (2012) Fingerprint recognition using wavelet domain features. Int Conf Nat Comput 531–534
7. Dhannoon BN (2017) Fingerprint recognition by using iterative closest point. 7(4)
8. Dautov ÇP (2018) Wavelet transform and signal denoising using wavelet method. Signal Process Commun Appl Conf 1–4
9. Kwaochai A, Pongyupinpanich S, Areekul P, San-Um W (2017) An application program of fingerprint detection using wavelet transform for authentication. In: Management and innovation technology international conference (MITicon), pp MIT-217–MIT-220
10. Tewari K, Kalakoti RL (2014) Fingerprint recognition and feature extraction using transform domain techniques. In: International conference on advances in communication and computing technologies (ICACACT), pp 1–5
11. Krishnasamy P, Kriegman D, Belongie S (2011) Wet fingerprint recognition: challenges and opportunities. In: International joint conference on biometrics (IJCB), pp 1–7
12. Hong L, Wan Y, Jain A (1998) Fingerprint image enhancement: algorithm and performance evaluation. IEEE Trans Pattern Anal Mach Intell 20:777–789
13. Chikkerur S, Cartwright A, Govindaraju V (2007) Fingerprint enhancement using STFT analysis. Pattern Recogn 40:198–211
14. Maltoni D, Maio D, Jain A, Prabhakar S (2009) Handbook of fingerprint recognition, 2nd edn. Springer, London
15. Lin C, Kumar A (2018) Matching contactless and contact-based conventional fingerprint images for biometrics identification. TIP 27(4):2008–2021
16. Dyre S, Sumathi CP (2016) A survey on various approaches to fingerprint matching for personal verification and identification. Int J Comput Sci Eng Surv (IJCSES) 7(4):1–7
17. Dyre S, Sumathi CP (2014) Hybrid approach to enhancing fingerprint images using filters in the frequency domain. In: IEEE international conference on computational intelligence and computing research (ICCIC), pp 1–6
18. He Z, Zhao X, Zhang (2015) Low-quality fingerprint recognition using a limited ellipse-band-based matching method. J Opt Soc Am A 32(6):1171–1179
19. Chandra E, Kanagalakshmi K (2011) Noise elimination in fingerprint images using median filter. Int J Adv Netw Appl 02(06):950–955
20. Thai D, Huckemann S, Gottschlich (2015) Filter design and performance evaluation for fingerprint image segmentation. arXiv:1501.02113 [cs.CV]
21. Turroni F, Maltoni D, Cappelli R et al (2011) Improving fingerprint orientation extraction. IEEE Trans Inf Forensics Sec 6(3):1002–1013

Developing Machine Learning Based Mobile App for Agriculture Application

R. Dhivya and N. Shanmugapriya

Abstract With the help of machine learning algorithms including KNN, SVM and LDA; it is possible to determine which crops are to be grown, when the soil needs more water and fertilizers, and what pests are present in the crops. The proposed system is designed to collect the current soil conditions and to calculate soil nutrients by analyzing the current soil conditions. In the proposed model, SVM offers 100% precision and LDA offers 95% accuracy in soil prediction. IoT camera sensor modules will assist farmers in determining whether their crops have been infected with pests so that they can take appropriate measures. In the application, the farmer can get alert notifications as well as other relevant information about crops such as soil type, crop types, nitrogen, potassium, and phosphorus, based on the soil and weather conditions. Farmers can also determine which crops to plant based on the soil and weather conditions. This empowers the farmer to take appropriate action for minimizing crop loss and maximizing crop yield.

Keywords Machine learning · IOT · Soil condition · Mobile application · Camera sensor module

1 Introduction

In order for a crop to be successful, farmers need to pay attention to a wide range of factors, including the quality of the soil, the amount of water that is in the soil, the temperature of the soil, and the amount of fertilizer that is used. At various points in their development and depending on the type of soil they are growing in,

R. Dhivya (✉)
Department of Computer Science, Dr. SNS Rajalakshmi College of Arts and Science, 49, Coimbatore, Tamil Nadu, India
e-mail: divgopal2@gmail.com

N. Shanmugapriya
Department of Computer Applications (PG), Dr. SNS Rajalakshmi College of Arts and Science, 49, Coimbatore, Tamil Nadu, India
e-mail: spriyanatrajan@gmail.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_15

plants require varying concentrations and proportions of specific nutrients. Before the farmer can start to see any returns on his labor, he has to put in a significant amount of time and energy. Cut down on their workload by coming up with an answer that satisfies their requirements. The amount of water that is required is determined by the current moisture content of the soil as well as the climate outside. The amount of insecticide and fertilizer that should be applied is determined, in part, by the weather that is currently taking place. Our main goal is to provide a system that can detect and recommend the requirements to increase crop earnings. As a result, the farmer avoids losses thanks to our system's earlier pest detection. By analyzing the climate and soil conditions, it also proposes what crop should be produced. in order to increase profits, the crop's market demand. Now, Sect. 2 of the paper below discusses how relevant the suggested idea is. The part III that follows Sect. 2 discusses the literature review that was conducted. The flow of the suggested solution is presented in section IV after which each module is thoroughly described. Our suggested solution's results are finally put to rest.

2 Relevance

To date, farmers have only added fertilizer, sprayed pesticides, and increased soil moisture based on their own experience, expertise, and observations of the existing state of the land. Instead of relying on the present weather and soil conditions, farmers themselves decide how much water, insecticides, and fertilizer should be applied, how often, and how much of each. To effectively identify this, farmers generally lack the experience and knowledge necessary. A lot of time passes before soil testing is done. As a result of the fertilizers being added in accordance with the results of the soil testing, and because the present soil conditions may differ greatly, this may result in soil degradation. Additionally, pests are not found in agricultural produce at the proper period, which has a negative and destructive effect. Pests should be watched out for in crops, and appropriate action should be taken to address problems. Which crops should be grown cannot be determined with accuracy by farmers. The type of crop that has to be grown is chosen without taking present soil conditions, weather, or market demand into account. The issues highlighted above will be resolved by our suggested solution, which will also assist farmers in ensuring a better crop.

3 Related Work

There is a list of suitable crops that has been provided by Jain [1], and it takes into account both the yield and the market prices. In addition to that, there is an algorithm for crop sequencing that can be used to rank the crops that are chosen. The CSM crop selection algorithm [2] is based on a variety of crop classifications, such as seasonal, year-round, short-term, and long-term plantations. These crop classifications are used to determine which crops should be planted. The CSM algorithm takes into account a variety of factors, including the climate, the composition of the soil, the amount of available water, and the type of crop being grown. The author of [3] recommends harvesting crops through the use of a random forest methodology. The following are some of the criteria that are taken into consideration: weather, soil quality, moisture content, the due factor, irrigation, and yield form. An ARIMA model, a neural network model, and a multilinear regression model are incorporated into a proposed multi-model for price prediction that is presented in [4]. This model is intended to estimate crop market prices. The author of reference [5] presents a model for predicting prices. Methods of image processing such as masking and segmentation are described in [6], which was written by its author. The clustering method uses the relevant parts of the selected images that were left over after segmentation. In this instance, the image of the pest is obtained by first obtaining an image of the grouped plants and then subtracting the image of the pest from that image. In [7], the identification of pests is accomplished through the use of a few different image processing techniques. The SVM's training was done with the help of color features. When using an SVM, it is much simpler to differentiate between the pixels representing the leaves and the insects.

Agriculture continues to be the primary source of income for more than half of the Indian population in recent years. About 18% of the nation's GDP is derived from agriculture, making it one of the most important sectors of the economy. All emerging technologies can boost crop yields by maintaining ideal growing conditions. Through an Android application, the internet of things is used to automate agricultural machinery in accordance with the article's suggested approach. Through machine automation, all work can be performed without human intervention. It utilizes a beagle bone air processor, wasp mote sensors, and numerous other IoT-enabled hardware components to automate agricultural fields. By using this method, agricultural fields could become more productive, and farmers' fatigue could be reduced significantly [8].

It will be necessary for the world population to increase output in each industry, especially agriculture, in order to meet the growing demand for food. There will, however, be times when demand and supply diverge. It is challenging to increase agricultural productivity through management of capital, people, and resources. Smart agriculture is a better solution for maximizing food production, resource management, and labor. This research presents forecasting analysis, Internet of Things (IoT) devices with cloud management, and security units for multi-culture agriculture while taking previous farming experience into consideration. Statistical and quantitative methods can be used to revolutionize the agriculture system more effectively by integrating contemporary technologies and traditional farming practices. A green field also includes irrigation, plant diseases, crop phases, and drone activation via IoT. Sensors are triggered for a variety of purposes in the Internet of Things. Modern agriculture will be developed using cutting-edge IoT tools and ideas in this research. A systematic evaluation provides an overview of current and upcoming trends in the agriculture sector. There are numerous analysis sections in this study that provide a broad framework for the model presented. We have thus been able to provide a number of advantages to our integrated units, as we have previously stated in our article regarding smart agriculture units with IoT modules. There are also some restrictions associated with this restricted paradigm for platforms and security. Most IoT-based agricultural devices consider the number of difficulties and constraints. Devices that are connected to the Internet of Things aim for cost-effectiveness by reducing hardware and software costs without compromising their output precision. Imported gadgets ignore the compromise of reduced component costs. Additionally, standardizing the data format for the process will improve device consistency and execution speed. Initially, there is a process barrier for engaged farmers to buy products or services when the system is being enhanced. As a result of their interconnection via a web server, the proposed integrated system will be complex [9]. An important technique for improving accuracy and performance in the IoT industry is the heterogeneity property. Deep learning analysis can be used to increase the productivity of IoT-enabled smart agriculture by using a large number of characteristics or data.

4 Proposed Work

The use of an Android application comes highly recommended as a method of operation. The capabilities of the application can be put to use in order to assist farmers in making decisions regarding which crops to cultivate, how much of each crop can be expected from a given yield, and how much money each crop will fetch on the open market. The farmer is informed, for one thing, of the current moisture content of the soil and is made aware of the presence of a pest as a result of using this tool. The system consists of NPK pills, a Moisture module, and an IoT camera module in order to ascertain the NPK values of the soil, ascertain the moisture content of the soil, and take photographs of pests. This strategy seeks to better the lives of farmers by addressing the problems they face and offering solutions to those problems.

5 Block Diagram

The represented diagram depicts the block diagram of our system (Fig. 1).

6 Architecture Description

The architecture of our solution is depicted in Fig. 1, which is a diagram. After a certain amount of time has passed, the IoT camera module and the moisture sensor will both send their data to Firebase. The farmer is responsible for inputting the NPK values into the application based on the readings from the pill chart. The hosted API

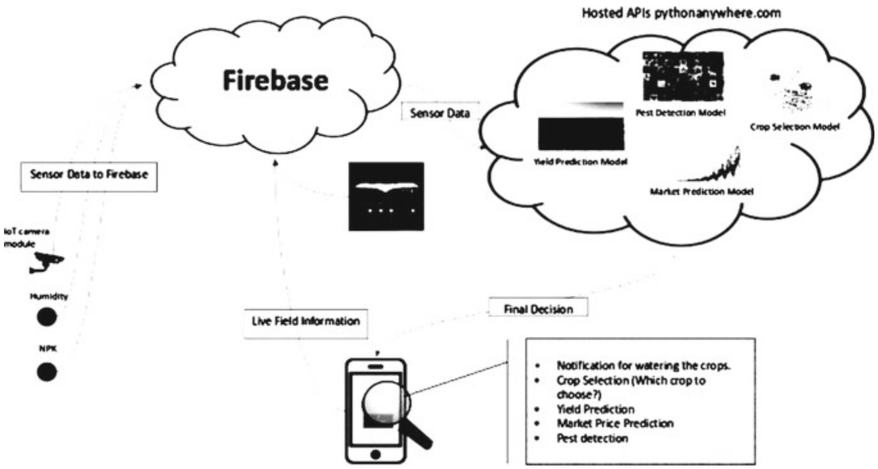


Fig. 1 Block diagram

for pest detection is called by the Raspberry Pi module that is connected to the Internet of Things camera. Notifications are dispatched in the event that a pest is identified through the mobile application used by farmers. The farmer initiates the API request for the other modules of the application by selecting one of the features available in the application. These modules include crop selection, yield prediction, and market price prediction. In order to acquire the required data, hosted APIs make use of two APIs: the weather API and firebase (weather data required for yield prediction and crop selection module). The desired outcomes are achieved by training models on the available input data. The results will be sent back to the application by the API. In the steps that follow, the results will be shown to the farmer on the screen of the application.

7 Live Field Status

Real-time information on soil moisture is fetched by this module, which also alerts the farmer when to water the plants. A predetermined threshold value for soil moisture was used to compare the soil's moisture level to. A notification detailing the situation will be sent to the farmer's cell phone if the scale falls below the threshold. To do this, a soil moisture sensor called NodeMCU (ESP8266) is employed, continuously recording the field's real-time status. The processing of this data then proceeds once it is pushed to the database. It is Firebase's Realtime Database that is used for this purpose. You may create top-notch apps with the help of Google's mobile platform, Firebase. A No-SQL database housed in the cloud called the Firebase

Dataset
Training Set
Test Set

Fig. 2 Dataset repesentation

Realtime Database enables real-time data syncing and storing between users. FirebaseESP8266.h is the ESP8266 library that is utilized for the transmission in NodeMCU.

8 Crop Detection and Market Prediction

As part of this session, you will advise the farmer on the best crops to plant in order to increase their yield and profit.

8.1 Data Gathering

The STCR Research Crop-wise Recommendations determined the necessary NPK values for a wide variety of crops by conducting research that was specific to each type of soil [10]. Figure 2 provides a visual representation of the dataset used for crop prediction. It includes information about the type of soil, the variety of crop, and the season, in addition to information about the amount of NPK that is currently available and that is required. On the Open Government Data (OGD) Platform in India, a dataset containing crop yield information was available. [11] Each of the following factors—crop year, season, area, and production—is broken out into its own column. Data on minimum and maximum temperatures as well as rainfall were supplied by Skymet's weather services [12]. The data on market prices incorporate information on crop and market prices [13].

8.2 Data Pre-processing

Using OpenCage Geocoder, Data Science Toolkit, and a climatic station that is located in the area, the crop yield data from an Indian district was combined

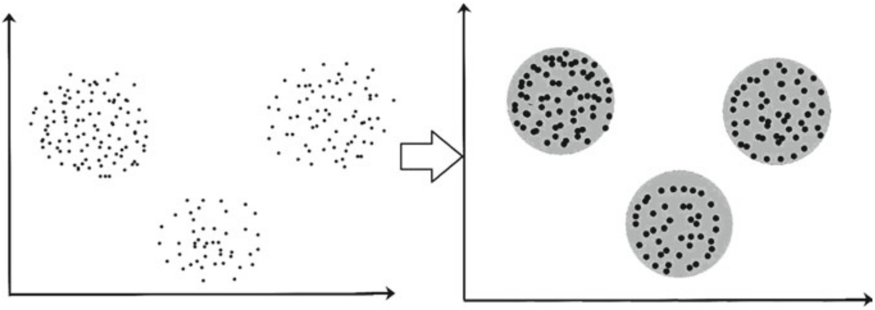


Fig. 3 Clustering

with climate parameters such as minimum and maximum temperatures and rainfall. Normalization of the datasets was necessary before they could be used as input for the training models [14–20]. The following are the prerequisite actions to take:

8.2.1 Clustering

Clusters are formed by crops that are grown in conditions that are comparable in terms of the soil and the climate. Figure 3 shows the methods of KNN, SVM, and LDA that are utilised in the process of accomplishing this objective. For instance, when establishing clusters, factors such as rainfall, temperature, and the type of soil are all taken into consideration. The model will generate a crop cluster that is appropriate for a given new observation by taking into account the current conditions of the soil and the weather.

8.2.2 Yield Prediction

The yield of the crop is predicted with the help of a decision tree regression model in this section. It is possible to prevent over pruning by limiting tree cuttings to a depth of no more than one quarter of an inch. As input features, we make use of the location (with a single Hot encoded), temperature, and precipitation.

8.2.3 Price Prediction

It is possible to determine the market price of the appropriate crops by employing a model that uses time series analysis. The Autoregressive Integrated Moving Average Model is utilised in order to accomplish this objective (ARIMA model). We determined whether or not the data were stationary by using the ADF test. One separate model will be trained for each crop and market in order to make a prediction about the possible selling price at that particular market location. The farmers can then use

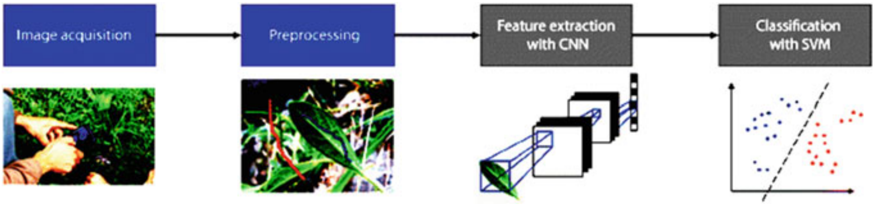


Fig. 4 Crop prediction using ML

this formula as a guide to assist them in selecting the most profitable crops to grow on their land in order to maximise their income.

9 Crop Detection

9.1 Image Acquisition

Using color cameras, various images of the soil samples that need to be categorized are taken and delivered as input to the system. Each type of soil's characteristics are gathered and kept in its own database. The final part of this process involves the detection of soil and crops using this database. The image's edges and contrast need to be improved for greater accuracy and clarity. The image's color maps are utilized to improve the contrast and edges. The crop prediction utilizing machine learning approaches is shown in Fig. 4.

9.2 Image Pre-processing

There are some inaccuracies in the image we first obtained. The results of the analysis are significantly influenced by the image's quality because it affects both the detection of the objects under investigation and the precision of the measurements that follow. To get an image free of errors, pre-processing steps are used. Since it involves improving the image by boosting contrast and correcting faults to produce a higher-quality image for use in later stages, this process is often referred to as "image augmentation." Before continuing with the processing, the image must be cleaned up of any flaws like noise or artifacts like scratches, lap tracks, comet tails, etc. In order to remove the noise and artifacts from the image, a filter known as the Smoothing filter is used. Low pass filters and high pass filters are the two different categories of filters. A low pass filter is a smoothing filter. It is applied to digital images to reduce high spatial frequency noise. A moving window operator is used in smoothing filters to change the value of each individual pixel in the image by some function of a nearby region of pixels. To affect every pixel in the image, the operator glides over it. The smoothing filter gradually improves the image by erasing the defects as a result of numerous rounds.

9.3 Feature Extraction

Following the k-means stage, in which the image is broken down into its individual components, the next stage is the feature extraction stage. This is the first step in the process that needs to be completed. During this phase, both the crop detection and the soil type classification processes are finished. To determine the type of soil present, scientists look at a wide variety of characteristics, such as the texture, colour, intensity, saturation, and hue of the sample. Feature extraction is accomplished with the help of the Gabor Filter. The Gabor Filter is a type of linear filter that analyses images in order to locate and highlight edges. Because of the similarities between the human visual system and the Gabor filter, it has been discovered that the frequency and orientation representations of the Gabor filter are particularly suitable for representing and differentiating textures. Utilizing a collection of frequency and orientation representations can be helpful in extracting the most information possible from an image. In addition, the Gabor filter can be used to obtain information such as the mean, the standard deviation, and the entropy. It is necessary to extract the colour of the soil, as it is the most important component. This distinction can be made with the help of colour moments, which is a metric that compares the colour characteristics of various images to one another. By comparing the values of similarity between photos to the values of images stored in an image database, it is possible to carry out activities such as image retrieval using the values of similarity between photos.

9.4 SVM Classification

The method known as Support Vector Machine (SVM) is utilised in the process of soil classification. It has found applications in a wide range of fields, including bioinformatics, recognition of text and images, as well as other areas. Because of how easy it is to use, SVM is a popular choice for an algorithm. This algorithm is being considered a potential replacement for the neural network algorithm. This is a description of how the process works, as it follows: An SVM training algorithm generates a model that sorts newly discovered examples into either of two categories. This classifier is a non-probabilistic binary linear classifier. Given a collection of training examples that have been labelled as belonging to one of two categories, it performs this action. Within an SVM model, there ought to be as much breathing room as is humanly possible between the examples of the various categories. After that, those newly created instances are mapped according to which side of the divide they are located on. SVM is helpful for finding the margin with the largest difference and performing separate plane analysis. This helps in determining whether or not the data points have sufficient support. As a consequence of this, the mapping looks like this: In this particular scenario, X is an instance of a class, and Y is a label for the category to which it belongs.

This algorithmic approach is based on distinguishing characteristics analysis and looks at predicted error minimization. To improve the training process, this technique considered empirical risk. In this case, structural analysis is used to calculate risk in order to reduce generalization error. Based on a review of the error margin under the class deviation heading, the closest training patterns are identified. This model is similarly built on the polynomial kernel representation in order to efficiently learn the elements and achieve greater accuracy.

9.5 Segmentation

The image is enhanced prior to the application of the segmentation technique by using image pre-processing techniques. K-means Image segmentation into its component parts is accomplished through the process of clustering. Partition clustering is a technique that is used to maximise a set of clustering criteria by dividing a given data set into disjoint subgroups. This technique is used as part of a partition clustering method. The clustering error criterion is the one that is utilised the vast majority of the time. This criterion calculates the squared distance that each point has from the centre of the cluster, and then adds up these distances for each point in the data set. The assignment of each pixel in the image to a cluster begins at the cluster's centre and continues throughout the image. The centre of the cluster is located by taking the average of all of the pixels. The next step in this cycle entails connecting each point in a particular data set to the centroid that is located closest to it.

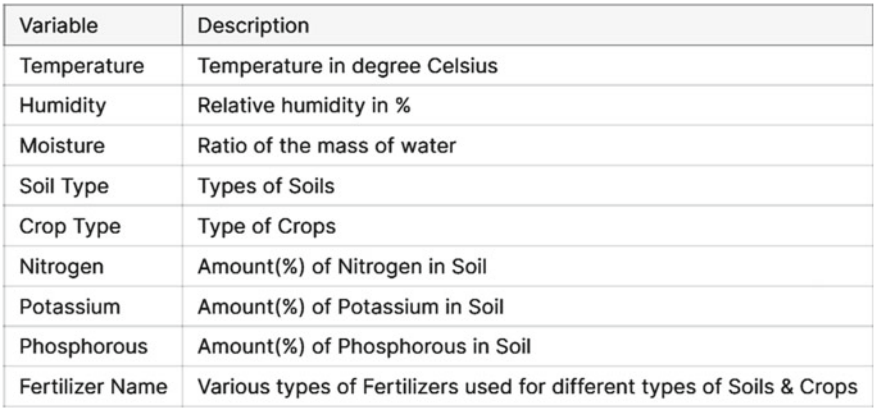
K-means The algorithm is as follows:

(1) K, a group of points ranging from × 1 to × n
(2) Disperse the centroids C1.........Ck throughout the universe.
(3) Repeat till convergence.
J = 1 for each cluster, K; The new centroid Cj from the previous step equals the average of all the points xi attributed to cluster J.
(4) Stop if none of the cluster assignments change.

10 Implementation and Result Analysis

10.1 Home Screen Along with Application Menu

With the help of this screen, the farmer will easily comprehend the app. Temperature, humidity, wetness, crop kind, soil type, nitrogen, potassium, and phosphorus content,



Variable	Description
Temperature	Temperature in degree Celsius
Humidity	Relative humidity in %
Moisture	Ratio of the mass of water
Soil Type	Types of Soils
Crop Type	Type of Crops
Nitrogen	Amount(%) of Nitrogen in Soil
Potassium	Amount(%) of Potassium in Soil
Phosphorous	Amount(%) of Phosphorous in Soil
Fertilizer Name	Various types of Fertilizers used for different types of Soils & Crops

Fig. 5 Home screen

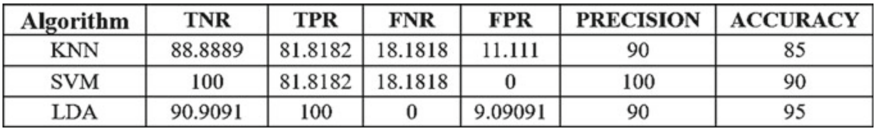
as well as the name of the fertilizer, are all displayed by the programme as illustrated in Fig. 5.

10.2 Confusion Matrix

A confusion matrix is used to assess the classification model's effectiveness. Measuring the effectiveness of our models by accuracy when we have uneven data can be deceptive. Classification models with binary or multiclass output are evaluated using performance metrics. Four of these pairings are listed in a table. There are two things to notice in the image up there.

(a) Predicted values: Values that the model predicts
(b) Actual Value: Datasets that include actual values.

We will utilize binary classification in this instance to comprehend the model. Positive points belong in the positive class, whereas negative points belong in the negative class. Rate is a measuring factor in a confusion matrix. The other 4 categories are TPR, FPR, TNR, and FNR. Figure 6 displays the precision and accuracy of a machine learning system based on a confusion matrix.



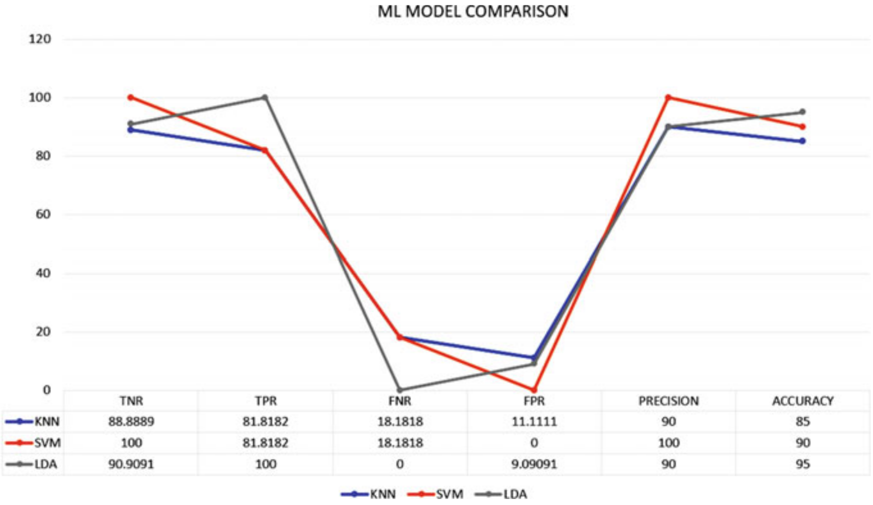
Algorithm	TNR	TPR	FNR	FPR	PRECISION	ACCURACY
KNN	88.8889	81.8182	18.1818	11.111	90	85
SVM	100	81.8182	18.1818	0	100	90
LDA	90.9091	100	0	9.09091	90	95

Fig. 6 Precision and accuracy analysis based on KNN, SVM, LDA

i. Values that are both actual and expected to be positive are known as True Positive (TP) values.
ii. Values that are anticipated to be positive but are actually negative are known as false positives (FP).
iii. False negatives are values that are positive when they should be negative (FN).
iv. Values that are both actual and expected to be negative are referred to as True Negative (TN) values.

10.3 ML Model Comparison

Agriculture is viewed as a vital business around the world where there are several challenges in calculating crops based on environmental conditions. This has become an issue for developing nations. Utilizing the most recent technologies, many organizations are reducing manual labor by using mechanical technology and IOT-based services. When removing physical labor, these strategies are especially useful, but not when generating predictions. In this study, crop yield is forecasted based on soil and temperature data utilizing the most modern ML technology such as KNN, SVM, and LDA. A dataset is built with various soil conditions as features and labels in order to forecast which type of each label will be connected to a certain crop. The sort of crop most suited for the existing conditions will be determined by the user's input of soil characteristics into the prediction procedure. Application assists in choosing the highest-yielding crops for hector. The examination of precision and accuracy based on the classification algorithms KNN, SVM, and LDA is shown in Fig. 7.



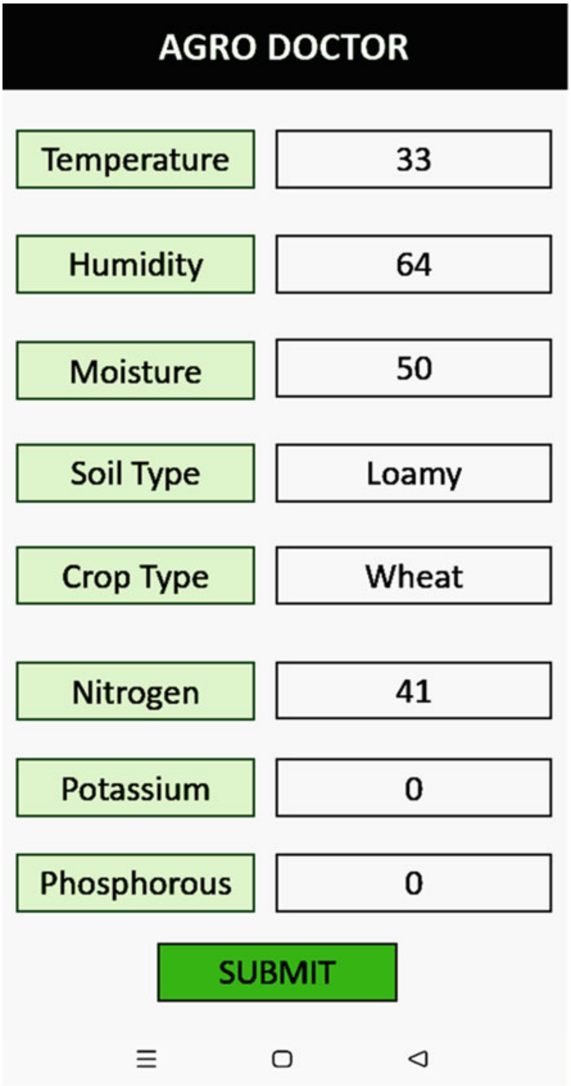
	TNR	TPR	FNR	FPR	PRECISION	ACCURACY
KNN	88.8889	81.8182	18.1818	11.1111	90	85
SVM	100	81.8182	18.1818	0	100	90
LDA	90.9091	100	0	9.09091	90	95

Fig. 7 Precision and accuracy analysis based on KNN, SVM, LDA

10.4 Notification and Information

This module will help farmers comprehend the soil conditions on their holdings. A notification will be sent based on the soil type and crop type of a soil. The app will also display the soil's temperature, moisture level, and humidity along with Nitrogen, Potassium and Phosphorous which is additional incorporated in this proposed model. The information on this page will assist other modules in carrying out their respective tasks. Figure 8 displays the suggested crop depending on the farmer's input.

Fig. 8 Precision and accuracy analysis



Temperature	33
Humidity	64
Moisture	50
Soil Type	Loamy
Crop Type	Wheat
Nitrogen	41
Potassium	0
Phosphorous	0

11 Conclusion

The proposed system offers the farmer a variety of modules in an attempt to address many of their problems. The technique offers soil prediction and crop prediction and helps to decrease losses in the agri field. The module lets the farmer know which crops are likely to thrive in his region based on soil and climate. With this module, farmers can sell their goods at higher prices. The smartphone application will allow us to learn everything about the farmers and their fields. In the suggested model, SVM provides 100% precision while LDA provides 95% accuracy for predicting soil. Using the machine learning technology, the system takes into account both the current soil conditions and climate in order to make better recommendations.

References

1. Jain N, Kumar A, Garud S, Pradhan V, Kulkarni P (2017) Crop selection method based on various environmental factors using machine learning. Int Res J Eng Technol (IRJET) 4(02)
2. Shyamalaprasanna A, Velnath R, Dhivya KT, Aishwarya S, Saravana G, Srimathi R (2021) Monitoring and controlling of industrial sewage outlet using IoT. In: 2021 International conference on advancements in electrical, electronics, communication, computing and automation (ICAECA), 2021, pp 1–5. https://doi.org/10.1109/ICAECA52838.2021.9675638
3. Karthikeyan R, Gowthami M, Abhishhek A, Karthikeyan P (2018) Implementation of effective crop selection by using the random forest algorithm. Int J Eng Technol 7(3):34. https://doi.org/10.14419/ijet.v7i3.34.19209
4. Ravanan V, Subasri R, Vimal Kumar MG, Dhivya KT, Kumar PS, Roobini K (2021) Next generation smart garbage level indication and monitoring system using IoT. In: 2021 Smart technologies, communication and robotics (STCR), 2021, pp 1–4. https://doi.org/10.1109/STCR51658.2021.9588961
5. Kaur M, Gulati H, Kundra H (2014) Data mining in agriculture on crop price prediction: techniques and applications. Int J Comput Appl 99(12):0975–8887
6. Priya P, Muthaiah U, Balamurugan M (2018) Predicting Yield of the crop using machine learning algorithm. Int J Eng Sci Res Technol. ISSN: 2277-9655; Krishnan M, Jabert G (2013) Pest control in agricultural plantations using image processing. IOSR J Electron Commun Eng (IOSR-JECE) 6(4):68–74. e-ISSN: 2278-2834,p-ISSN: 2278-8735
7. Rajan P, Radhakrishnan B, Padma Suresh L, Detection and classification of pests from crop images using support vector machine
8. Bashar A (2019) Agricultural machine automation using IoT through android. J Electr Eng Autom 1(2):83–92
9. Suma V (2021) Internet-of-Things (IoT) based smart agriculture in India-an overview. J ISMAC 3(01):1–15
10. STCR Research Crop Wise Recommendations. http://www.iiss.nic.in/downloads/stcr%20Crop%20wise%20Recommendations.pdf
11. Open Government Data OGD Platform India. https://data.gov.in/
12. Skymet Weather Services. https://www.skymetweather.com/
13. Market price data. https://agmarknet.gov.in/
14. ImageNet dataset. http://www.image-net.org/
15. Source of the dataset for non-pest plants. https://www.kaggle.com/emmarex/plantdisease
16. Tensorflow’s inception-v3 model. https://cloud.google.com/tpu/docs/inception-v3-advanced
17. Weather API. https://api.weatherbit.io/

18. Partitioning around medoids (PAM). https://web.archive.org/web/20111002220803/ http://www.unesco.org:80/webworld/idams/advguide/Chapt7_1_1.html
19. Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A (2019) The global burden of pathogens and pests on major food crops. 3
20. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojnam S, Rethinking the inception architecture for computer vision

Attack Detection in IoT Using Machine Learning—A Survey

Saeed Ali Haifa Ali and J. Vakula Rani

Abstract In the last decade, the İnternet Of Things(IoT) increased dramatically until it became an integral part of our daily lives. These devices are interconnected to the internet without the need for human intervention. Due to the weak configuration and unique characteristics of the internet of things has become a robust target for cyber-attack that worry the general user of these devices. Furthermore, IoT security challenges are increasing day by day and are subject to a variety of attacks. The traditional security measures, such as authentication, access control, network security, and encryption, for IoT devices and their vulnerabilities, are insufficient, ineffective, and cannot process these issues. Existing security methods must be improved to protect the IoT environment. ML/DL provided many solutions that assisted solve the challenges of the IoT and provided safety for it. The goal of this paper is to provide a study on the attacks in IoT architectures such as the sensing layer, network layer, and application layer, then present ML and DL that contributed to the solution in attack detection. In addition, we discuss the challenges of IoT architectures.

Keywords Internet of Things(IoT) · Machine-Learning · Deep-Learning · Attack detection · IoT security

1 Introduction

The internet of things become widely utilized nowadays, which connects devices to the internet to exchange data, IoT shares the sensor information that they gather by linking to an IoT gateway or another edge device where data is either sent to the cloud to be analyzed or analyzed locally; IoT has unique features such as resource constraints, heterogeneity, scalability, and the extensive behavior of the networks

S. A. Haifa Ali (✉) · J. Vakula Rani
CMR. Institute of Technology, Bengaluru, Karnataka, India
e-mail: haal21rs@cmrit.ac.in

J. Vakula Rani
e-mail: Vakula.r@cmrit.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_16

make the security and privacy in IoT more challenging and vulnerable to cyber-attacks, which aims to disrupt its environment. With the advancement of technology, the interconnection of IoT ecosystems with the enterprise network and the entire Internet also occurs, making IoT and industrial IoT devices a prime target for cyber-criminals and cyber-attacks. Either cause damage to devices, and files or threaten countries or stalk people either online or offline. Moreover, use them in attacks against other targets. On the other hand, there are three classes of IoT attacks that are commonly found in today's network environment which are, access attacks that access another user account or network devices through improper means, reconnaissance attacks that are used to gather information about the target network or system, and DDoS attacks which we explain it in Table 1. However, the cyber-attacks target the IoT layers, due to the role that each layer plays, which make them subject to a diversity of attacks. The attacks can compromise the devices or other nearby devices; control the flow of the network traffic. Furthermore, attract people to trust them and then commit cyber offenses. The number of IoT devices is expected to grow by over 25.4 billion in 2030. This increase in the number of IoT devices will increase the number of attacks and the opportunities to compromise them through hacking, cracking, phreaking, denial of services attacks, online fraud, pornographic offenses, etc. Encryption and password security are one of the various vulnerabilities that are exploited by attackers.

Machine Learning, is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience and Deep Learning is a type of machine learning that allows the machine to learn by itself by simulating the neurons in the human body. These algorithms can be used for detecting and mitigating the risk of attacks that come to provide solutions based on learning and prediction. ML and DL methods such as Super Victor Machine (SVM), Decision Tree (DT), RNN, DNN, Auto-encoder (Ae), etc. provide better solutions to detect and mitigate attacks. Although there are some existing surveys that have discussed attacks in IoT layers, there is a shortcoming and a lack of discussion of these types of attacks [1, 2] in depth. İn this study, we covered all types of promising attacks in IoT layers, and also presented some studies that addressed the promising attacks which disturb IoT security. The contribution of this paper is to concentrate on the attacks in IoT architectures, and the contribution of machine learning and deep learning to detect attacks. In addition, the challenges of IoT architecture.

2 IoT Architectures Attacks and Attacks Detection

IoT architecture consists of three important layers, perception layer, network layer, and application layer, and each layer has its mission. In Fig. 1, we display the IoT architecture, which illustrates the position and the component of each layer. The perception layer is the first layer and consists of devices and controllers to sense and gather information. Then, the network layer, which is made up of connections and protocols, and the main goal is to provide a connection between networks, and finally,

Table 1 Type of each attack and its principle

Attack	Principle
Reverse engineering attacks	A person-to-person attack in which the attacker establishes direct contact with the victim in order to persuade them to provide important information
Tampering attack	Type of physical attack in which the attacker can modify memory, obtains extra knowledge by reacting with a faulty appliance, and then try to break security
Spoofing attacks	The hackers impersonate an authorized device or user to steal data, spread malware, or bypass access control systems
Physical damage attack	Performed when the attacker is at a close distance from the device. A malicious user can take control of the computing or communication and cause damage to property and put lives at risk
RF interface attacks	Target devices that use communications methods such as Bluetooth, Bluetooth Low Energy (BLE), radio, and Wi-Fi
RFID cloning	Refer to the copying of an RFID electronic tag or smart card's information to a cloned tag, which will then have the same features as the original tag and potentially replace it
Code and malicious	Malicious software or malware that may corrupt client computers, databases, networks, and entire clusters of servers rapidly or over time
Injection attacks	A malicious code injected into the network which retrieves all the information from the database to the attacker
Theft of data	Steal the data and then put it in computers that have been broken into, servers, or other devices with the desire to violate the privacy or get private information
Sniffing attacks	Monitor and capture all data packets passing through a given network
Browser attack	Employ a Trojan horse covertly installed on a computer system that alters the user's web transactions in real-time, steel the messages in a public key exchange, and replaces the targeted security keys with fake ones
Denial of services (DoS)	Attacks whose goal is to flood the network or sites with unnecessary data sent by infected devices with the malicious program to control remotely, DoS causes slow services and congestion in the flow of data
Distributed senial of services (DDoS)	The attackers flood a server with internet traffic to deny users from accessing to services and websites
Social engineering	Set of tricks and techniques to deceive people and make them expose confidential and personal security information

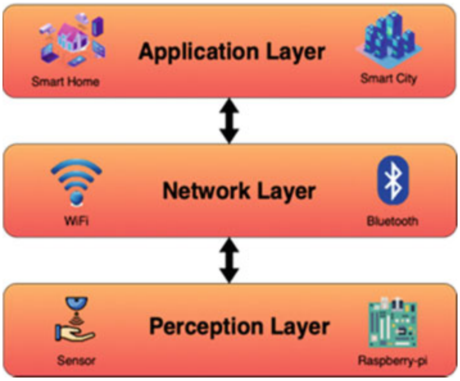
(continued)

Table 1 (continued)

Attack	Principle
Sinkhole attacks	A sinkhole attack is one in which a hacked node tries to attract network traffic by announcing a bogus routing update
Sybil attack	A group of nodes impersonating different peer identities in order to compromise an IoT ecosystem and it is used to broadcast bogus data from a random network
Jamming attack	Interfere with the radio frequency of the wireless sensor node, which will jam the signal and sent the communication back to the node
Man in the middle (MiTM)	It occurs when a man or hacker intervenes in the communication between the system and the user, or between two users, in order to steal the user's information
Phishing attack	Spoof website that pretends to be the real thing, The victims become imprisoned on the bogus website, and hackers steal all of their login information as well as credit card information
Social engineering	The attacker gains people's trust to do things against the security or policies and fools them into providing information to get access to information

the application layer along with the user can react to the software application. The goal of this layer is to give the user permission to access and recover files. However, all three layers suffer from promising attacks and these types of attacks discuss in Table 1. The main goal of this section is to understand the risk of attacks in each layer and provide the current solutions, which have been presented by researchers with the advantage of each research.

Fig. 1 IoT architecture



Physical Attacks: it has another two names the sensing layer and the perception layer; it is made up of sensors, and actuators. This layer in the internet of things is accountable to capture information, actuators, Zigbee, and RFID [3]. It suffers from various types of attacks that try to damage and destroyed devices such as Reverse engineering, Tampering Attacks, Jamming, Spoofing, and RF interference. Figure 2 illustrate the various type of physical attacks. The attacks enter the devices and manipulate them. Entry is through social engineering, where attacks get into devices and carry out real attacks on a large scale, such as destroying devices, eavesdropping, or other attacks. In order for physical attacks to be launched, the attacker must be close to the devices such as tampering with energy sources, endangering the communication mechanism, and reducing the life of the devices. The physical attack is nothing but a launch for the rest of the attacks like a smart lock, disabling the smart locks will potentially lead to steel money and kidnapping a member of the house. Detecting and securing this layer needs robust techniques like machine learning and deep learning methods. On the other hand, there are plenty of researchers discussing the problem of the attack on the physical layer. They have proposed many paper-based solutions to address the attacks on physical layers and one of these papers [4] has addressed the issues in the physical layer caused by jamming attacks using Random Forest (RF) machine-learning forest. The method was compared with two machine learning methods support vector machine (SVM) and K-nearest Neighbour (K-NN) in the multi-track status and one-route status. The Random Forest technique along with AdaBoost achieved significant results better than another method. Consequently, it was compared with a one RSSI status for a diverse number of nodes. Moreover, it was found that Random Forest has better accuracy in terms of testing and minimal false alarms compared to other methods.

In [5] analyzed the network traffic using an Artificial Neural Network (ANN) to expose malicious intrusion in IoT devices such as smart bulbs. The result of ANN compared the performances with previous studies results that used several supervised algorithms SVM, DT, and RF. It demonstrated reasonable accuracy and error rate of ANN respectively, 77.50 and 24.48% compared to earlier results used (SVM, RF, NB, and DT).

Liu et al. [6] have proposed a network invulnerable schema, called P4NIS that is supplied with three lines of defenses to detect and prevent eavesdropping attacks. Three lines of defense analyzed the packets. First encrypted the package in the upper layer with utilize cryptographic methods. Then took a program design for modifying encryption techniques, and after that, they utilized programmable forwarding policies that divided the encrypted packages into various network routes disorganized. The results showed that P4NIS rise awkwardness in eavesdropping attacks compared with modern methods. Furthermore, this schema can decrease the encryption cost by 69.85–81.24%.

Shaniqua et al. [7] proposed a machine learning model using SVM to taxonomy spoofing attacks and signals received by Unmanned Aerial vehicles (UAVs), as well as, they have performed K-fold examines to improve another learning pattern via selecting various rates of K-folds then was named them K-learning methods. Moreover, the proposed work by utilizing the characteristics of GPS to be like features.

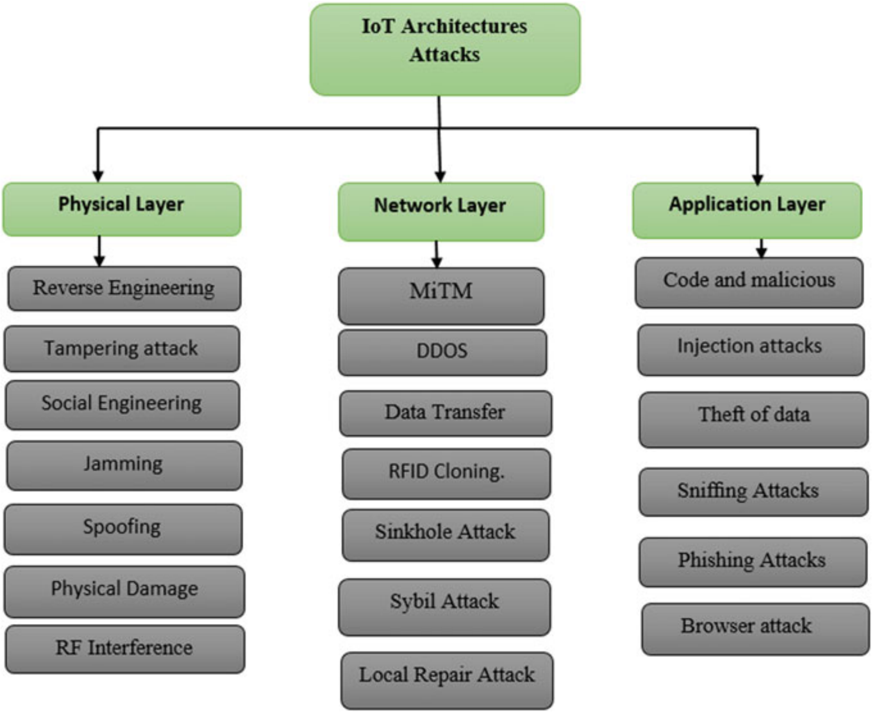


Fig. 2 Attacks in IoT architecture

Then they compared the model with five previous studies that have been carried out which scored significant accuracy, precision, recall, and F-score respectively (99, 98, 99, 98%).

Pathak et al. [8] analyzed IoT security sensors to discover tampering attacks in an office environment using the decision tree (DT) method. They compiled data from the genuine world and then, use machine learning to reveal tampering attacks by using two techniques. They used a genuine-time of the traffic type to train unsupervised machine learning methods for anomaly exposure. After that according to traffic, labels were created, and the decision tree supervised techniques has been used. Moreover, the result based on the Decision tree scored high accuracy of 91.61% and a low false-positive rate (FPR).

Network Attacks: This layer is used to convey data from the lower layer to the upper layer for processing [9, 10]. The major threats here are eavesdropping attacks, a man in the middle (MITM), Sybil attacks, Sinkhole attacks, DDoS attacks [11], and RFID Cloning. When the attackers hack this layer, they turn IoT devices into botnets. The aim of Network Layer attacks is to interrupt the route between the source and destination. This is one of the routing protocols that has been chosen. In network attacks, the attackers bombard the network with more traffic through compromised nodes or fake nodes, and from here, they allow an attack called Sybil attacks. Tampering with devices might allow attacks to make changes in the routing table and security key that would affect upper layers. On the other hand, the network layer is the most important layer due to its position in the middle between the physical layer and the application layer, therefore, securing this layer is essential and many proposed works addressed the issues in this layer. We present some proposed solutions based on machine learning and deep learning, which have been provided to solve the challenges in the network layer.

Kiran et al. [12] have built a model to identify Man-in- the- Middle attacks in the IoT network layer using machine-learning methods such as Naïve Bayes (NB), Support Vector Machine (SVM), and Adaboost. They built a testbed to simulate an IoT environment using a node, sensor, and wireless router and they built an adversarial system using a laptop. On the other hand, they collected data from sensors and transmitted it to ThinkSpeak platform using the wireless gateway. After that, the sensor value has been captured by a node and then transmitted to ThinkSpeak server, which was stored. At the last, the performance measurements of the classifier scored high of 100% in terms of accuracy, detection rate, and false alarm rate (FAR).

Newaz et al. [13] have proposed an Intrusion Detection System scheme, to demonstrate different vulnerabilities of personal medical device communication and then implemented five diverse cyber-threats such as man in the middle, Replay, false data injection, and DoS on commercially available healthcare devices using machine learning SVM, KNN, DT, RF. Then, they presented a method called HEKA, to watch and reveal cyber-threats in the network traffic. They have implemented the method in a testbed that contained eight medical appliances and estimated its accomplishment with four various attacks. Furthermore, the proposed showed that HEKA is successful in revealing diverse attacks with a score of e 98% of accuracy and a F1 score of 98%.

Farzaneh et al. [14] have presented a deep learning method called the Fuzzy-based method that used three metrics to detect Repair Attacks on routing protocol. Moreover, the result used the Cooja simulator in the operating system Contiki and showed that the method was able to detect the Repair attack and scored a high True Positive Rate (TPR) and minimal False Positive Rate (FPR).

Alaiz-Moreton et al. [15] have proposed a model to reveal and detect attacks like DOS, MiTM, and Intrusions in IoT environments by using an intrusion detection system (IDS). The work concentrated on making classification models that can feed an IDS using a dataset containing frames under attacks of an IoT system that utilized the MQTT protocol. They addressed two kinds of methods for taxonomizing the attacks, ensemble methods, and deep learning models. SVM, RF, Boosting Gradient, XGBoost, and LSTM have been used in this model and achieved a high score accuracy of over 90%. However, SVM achieved a result better than the other methods.

Anthi et al. [16] have proposed a three-layer Intrusion Detection System (IDS), which used supervised machine learning methods to detect network cyber-attacks (Denial of Service (DoS), Man-In-The-Middle (MITM), Spoofing, and Replay attack) on IoT networks. The system was assessed by training eight devices on a testbed. The system included three main functions, which are, to taxonomy the type as well as, profiled the natural action of IoT devices that are linked to the network, then detected wireless attacks against connected IoT devices, and after that, classified the attacks that have been deployed. The proposed achieved a high score in the achievement of the three tasks resulting in an F-measure of: 96.2, 90.0, and 98.0%.

Application Attacks: The application layer manages connecting end-users to endpoints using a user interface and deals with a large number of data transactions [17, 18]. The most challenging layer to be protected is the application layer. Plenty of the vulnerabilities found here are based on sophisticated user inputs that are complex to express with intrusion detection. Moreover, is vulnerable to software attacks such as malware, viruses, worms, Sniffing attacks, Phishing attacks, Injection attacks, and Browser attacks. This layer is also accessible and exposed to the world. SQL injections, a type of application attack, were responsible for breaches of data in the year 2014. It is the third most prevalent attack, coming after malware and DDoS. Other common vulnerabilities in this layer are included, such as security misconfiguration. Attackers have the ability to change application information and capture private data without being exposed by network protection mechanisms. The researchers presented many types of researches based on application attacks and we will present a few of them. Azmoodeh et al. [19] have proposed a neural network (NN) deep learning technique to detect Internet of Things malware in the IoT device's Operational Code using the deep learning method. They transferred OpCodes in a vector space and applied deep Eigenspace learning methods to classify malicious and benign applications. The proposed achieved an accuracy of 98.37% and a precision rate of 98.59%.

Mao et al. [20] have presented a learning-based aggregation analysis mechanism to detect phishing pages and determined the similarity of page layouts using machine-learning methods SVM, DT, and RF. In addition, the result of the classifier achieved above 95% accuracy. Then they compared the proposed work with other approaches and found that their work achieved better performance.

Moti et al. [21] have presented a structure to detect malicious patterns called MalGan by using the deep learning method Generative Adversarial Network (GAN), the performance analysis was carried out using VX-Heaven dataset, which contained malicious and normal software. The proposed MalGan scored a higher detection rate compared to four previous works of malware detection algorithms in terms of accuracy, F1, testing time, and training time.

Shafiq et al. [22] have proposed a feature selection approach named CorrAUC to detect Malicious Bot-net attacks, then relied on CorrAUC, they developed a feature selection method called Corrauc using machine learning RF, NB, SVM, and C4.5 DT. They applied two methods called TOPSIS and Shannon Entropy based on a bijective soft set to advocate selected features for malicious traffic in the IoT network. They used the Bot-IoT dataset to carry out their performance. The proposed showed a significant result of over 95%. Many feature selection approaches have been presented in [23], which increases accuracy, noise reduced.

Sarkar et al. [24] presented algorithms to determine the key parameters necessary to follow and sniff a Bluetooth Low Energy (BLE) connection in the connected state by concentrating on the adversary setting with a low-cost single radio and developing a suite of real-time algorithms. The algorithm was implemented in the open-source platform. At last, the performance of the implemented method achieved better accuracy and over 80% as well as, stability to BLE operational dynamics.

The types of attacks in each layer are presented in Fig. 2, in which we demonstrate twenty attacks that break into the devices either in the perception layer or, hacking and, phreaking the network layer or grabbing data and destroying files in the upper layer either closely or remotely.

In Table 1, we display the diverse types of attacks in IoT architecture, and these attacks threaten the general security of each layer and are considered one of the most important security concerns that harm and impede the work of each layer, The table shows the principle and the goal of each attack in each layer to make it effortless concept for the other researchers. We observe from the table that all the types of attacks aforementioned concentrated on hacking either by breaking into devices or communication to damage and sabotage. Moreover, these attacks focus on fooling people like the goal of social engineering either online or offline. on the other hand, we notice that the most common hackers target E-mails is sniffing and phishing attacks by sending lots of tricking emails such as advertised emails or impersonating to obtain private details about the address, name, and bank account detail or ID.

In Table 2 we summarize the three layers (perception, network, and application) in terms of the types of attacks in layers that disturb the security, we enhance that with the main purpose and primary task of each layer, furthermore, we present the weaknesses and the challenges of each layer that make them vulnerable to attacks. Through our research on weaknesses in IoT layers security, we observe that not securing devices in the lower layer, the reason behind that is the natural characteristics in IoT make devices vulnerable to hackers and the amount of data flowing upper layer, which make the network overcrowding.

The Physical layer-based security solutions are listed in Table 3, which consists of a few studies references regarding the attacks in the perception layer, we display the contribution of each study and which type of attack the researcher concentrated on and analyse, and we present the techniques that the researchers use it to solve the problem, we pay attention that the majority of researchers are interested to solve spoofing, jamming and malicious attacks in perception layer. In fact, these types of attacks are the most dangerous in the lower layer that disturbs the devices. Finally, we extracted the advantages of each study that most of them got high accuracy and detection rate.

The network layer-based security solutions are listed in Table 4. The table contains some studies regarding the attacks in the network layer, we discuss the contribution of each study and the attacks that the researchers concentrated on, and we display the techniques that the researchers use to solve the problems, we observe that the majority of researchers are focusing on DoS, DDoS, MiTM attacks and intrusion. Indeed, these types of attacks are critical that affect the task of this layer; we extract the advantages of each study that most of them got high accuracy and precision, and less false positive rate.

Table 2 Illustrates the attacks, the main purpose of each layer, and the challenges

Layer	Attacks	Main purpose	Challenges
Physical layer	• Reverse engineering • Tampering • Social engineering • Jamming • Spoofing • DOS • Fake node • Physical damage • RF interference	Gather data	• IoT is unmonitored and vulnerable to hackers • Destroying perception devices and manipulating gathered information • Resource constrains devices • Data privacy
Network layer	• Man in the middle • DDOS • Data transfer attack • RFID cloning • Spoofing • Sinkhole attack • Sybil attack • Routing information attacks	Transmit collected data	• TCP and IP protocol • Concentrate on the influence of network resource availability • Wireless networks due to devices connected to IoT networks via wireless communication • Due to a high volume of data, there is network congestion
Application layer	• Code and malicious • Injection attacks • Theft of data • Sniffing attacks • Phishing attacks • Browser attack	User-requested assistance	• Because it deals with a high volume of data transfers, it presents a huge issue for IoT security • Attacks against software (data privacy, control of access, and leakage of data)

Table 5 shows Application layer-based security solutions. The table is composed of several studies related to the attacks in the upper layer, we overview the contribution of each study and the attacks that the researchers work on solving, and we show the techniques that contribute to solving the problems, we observe that the researchers are focusing on malware and phishing attacks. These types of attacks are significant due to the task of the layer such as dealing with data and files; we summarize the advantages of each study that most of them got high accuracy and precision, high recall, and f1.

3 Challenges

Although we discussed the solutions in IoT regarding machine learning and deep learning, the potential challenges [37], have not been solved by these methods. One of these problems is securing the use of data, leaving the devices without shutting

Table 3 Physical layer attacks solutions

Author and References	Year	Contribution	Attack	Countermeasure method	Advantage
Upadhyaya et al. [4]	2019	Discussed machine-learning methods to detect jamming detection using data from emulation and genuine networks	Jamming	RF, real AdaBoost	High accuracy
Khatun et al. [5]	2019	Presented Artificial Neural Network (ANN), which can be carried out to process the traffic in the network to be able to detect malware in IoT	Malicious nodes	ANN	Reasonable accuracy and error rate
Goel et al. [25]	2019	Presented the cooperation between DNN methods and block-chain security that can create a model to eliminate tampers	Adversarial tampering	DNN	High accuracy
Aref et al. [26]	2017	Proposed a method called multi-agent reinforcement learning algorithm, based on Q-learning, to bypass jamming threats	Jamming	Q-learning	Timing
Han et al. [27]	2017	Proposed a PHY-layer authentication to minimize the connection between security factors and landmarks and then supply security based on the scheme	Spoofing	ML	FAR, miss DR

(continued)

Table 3 (continued)

Author and References	Year	Contribution	Attack	Countermeasure method	Advantage
Xiao et al. [28]	2017	Created a jamming detection connection game that takes advantage of user mobility to mend the signals against jammers	Jamming	Deep reinforcement learning	Fast convergence rate
Shi et al. [29]	2017	Proposed a device with free user authentication by essencing physiological and features established in everyday activities	Human daily activities	DL	High accuracy

Table 4 Network layer attacks solutions

Author and References	Year	Contribution	Attack	Countermeasure method	Advantage
Kiran et al. [12]	2020	Have built a model to identify Man-in- the- Middle attacks in IoT network layer	Man-in-the middle	SVM, NB, DT, Adaboost	Misclassification, High accuracy
Newaz et al. [13]	2020	Proposed an Intrusion Detection System (IDS), which is called HEKA, to watch medical appliances traffic and detect attacks	MITM, replay attack, false data injection, DoS	SVM, KNN, DT, RF	High accuracy, F1
Zaman et al. [1]	2020	Proposed a Fuzzy method to detect and reveal of Repair Attack on the routing protocol	Local repair attack	DL, Fuzzy Logic	High TPR, less FPR
Singh et al. [30]	2017	Concentrated on creating classification models that can feed an IDS using a dataset containing frames under attacks of an IoT system that uses the MQTT protocol	Dos, MITM, Intrusions	SVM, RF, Boosting Gradient, XGBoost, LSTM	High accuracy
Roopak et al. [31]	2019	Suggested four deep learning models with the utilized of CICIDS2017 datasets and compared them with machine learning algorithms	DDOS	DL	High accuracy, high precision
Newaz et al. [13]	2019	Discussed the three-layer with the use of supervised methods to detect cyber-attacks on IoT network	DOS, MiTM, spoofing, reconnaissance, and replay	ML	High F1 measures

(continued)

Table 4 (continued)

Author and References	Year	Contribution	Attack	Countermeasure method	Advantage
Mohamad et al. [32]	2017	Suggested IDS named compression header processor that examines 6LoWPAN to minimize the routing attacks	Routing attacks namely, hello flood, sinkhole, and wormhole attacks	ML	High accuracy
Li et al. [33]	2017	Proposed a Sybil attacks detection in WSN by taking advantage of the connection between path responses from various sensors	Sybil attack	DL, fuzzy Logic	High accuracy

Table 5 Application layer attacks solutions

Author and References	Year	Contribution	Attack	Countermeasure method	Advantage
Azmoodeh et al. [19]	2018	Presented an IoT botnet malware detection method based on the selection of Opcodes as a feature for classification mission	Malware	DL	High accuracy, high precision
Mao et al. [20]	2019	Aimed to enable automated page-layout-based phishing detection techniques using machine learning techniques	Phishing	SVM, DT, AB, RF	High accuracy, reasonable F1
Dovom et al. [34]	2019	Conveyed the Opcodes into a vector space and apply fuzzy then rapid fuzzy tree methods to detect malware and its type	Malware	SVM, DT, KNN, RF	High accuracy, high precision, high recall
Parra et al. [35]	2020	Presented an IoT micro-security add-on hosted in the device using a CNN model to detect URL-based attacks directed to a client's IoT devices. and LSTM model hosted at the back-end servers for detecting botnet attacks in IoT devices	Phishing attacks, DDoS attacks	CNN, LSTM	High accuracy
Ahmad et al. [36]	2021	Proposed an anomaly detection technique utilizing the deep neural network for the IoT network layers to classify traffic as benign and unbenign	DOS, DDOS, Reconnaissance, and information theft attacks	DNN	High accuracy

them out could be subject to hackers that can exploit the power and get private information, for this, it should secure network connection to prevent any intrusion to gain access to devices. On the other hand, with data leaks in the IoT ecosystem, hackers can access the data analyzed by the system. This may have private data like location, bank account data, and health registry. Nevertheless, taking advantage of insecure connections is not the only thing for hackers to get beneficial data. The entire data is sent and stored in the cloud, and then cloud-hosted services are in danger of exterior attacks. Therefore, breaches of data are possible from the devices themselves and the cloud environments to which they are connected. For machine learning and deep learning, which are considered as a solution to IoT security, they suffer from selecting the proper ML and DL algorithm for the mission [38]. Due to IoT being considered resource constraints, this problem limits the use of DL and ML algorithms to protect IoT networks. Finally, DL needs a huge number of training data to do the task.

4 Conclusion

IoT becomes crucial in our life nowadays; however, it is subject to attacks that disturb security. Securing IoT is difficult and traditional security cannot process IoT but can handle other networks. For this, machine learning and deep learning can detect and mitigate the risk of attacks. In this survey, various attacks type on the tree layer, physical, network and application layer have been discussed. We covered security attacks on the three layers using machine learning and deep learning methods. This survey targets to provide a study of the current solution of machine learning and deep learning that has been enhanced with previous researches.

References

1. Zaman S et al (2021) Security threats and artificial intelligence based countermeasures for internet of things networks: a comprehensive survey. IEEE Access
2. Lin J et al (2017) A survey on internet of things: architecture, enabling technologies, security and privacy, and applications. IEEE İnternet of Things J 4(5):1125–1142
3. Liang X, Kim Y (2021) A survey on security attacks and solutions in the IoT network. In: 2021 IEEE 11th annual computing and communication workshop and conference (CCWC). IEEE
4. Upadhyaya B, Sun S, Sikdar B (2019) Machine learning-based jamming detection in wireless iot networks. In: 2019 IEEE VTS Asia Pacific wireless communications symposium (APWCS). IEEE
5. Khatun MA, Chowdhury N, Uddin MN (2019) Malicious nodes detection based on artificial neural network in IoT environments. In: 2019 22nd International conference on computer and ınformation technology (ICCIT). IEEE
6. Liu G et al (2021) Softwarized iot network immunity against eavesdropping with programmable data planes. IEEE Internet of Things J 8(8):6578–6590
7. Shaniqua A, Mehmood A, Elhadef M (2021) Detecting signal spoofing attack in uavs using machine learning models. IEEE Access 9:93803–93815

8. Pathak AK et al (2021) Anomaly detection using machine learning to discover sensor tampering in IoT systems. In: ICC 2021-IEEE ınternational conference on communications. IEEE
9. Mrabet H et al (2020) A survey of IoT security based on a layered architecture of sensing and data analysis. Sensors 20(13):3625.
10. Hassija V et al (2019) A survey on IoT security: application areas, security threats, and solution architectures. IEEE Access 7:82721–82743
11. Amrish R, Bavapriyan K, Gopinaath V, Jawahar A, Vinoth Kumar C (2022) DDoS detection using machine learning techniques. J IoT Soc Mobile Analytics Cloud 4(1):24–32
12. Kiran KVVNLS et al (2020) Building a intrusion detection system for iot environment using machine learning techniques. Procedia Comput Sci 171:2372–2379
13. Newaz AKMI et al (2020) Heka: a novel intrusion detection system for attacks to personal medical devices. In: 2020 IEEE conference on communications and network security (CNS). IEEE
14. Farzaneh B et al (2020) A new method for intrusion detection on RPL routing protocol using fuzzy logic. In: 2020 6th International conference on web research (ICWR). IEEE
15. Alaiz-Moreton H et al (2019) Multiclass classification procedure for detecting attacks on MQTT-IoT protocol. Complexity 2019
16. Anthi E et al (2019) A supervised intrusion detection system for smart home IoT devices. IEEE Internet of Things J 6(5):9042–9053
17. Tawalbeh L et al (2020) IoT privacy and security: challenges and solutions. Appl Sci 10(12):4102
18. Asharf J et al (2020) A review of intrusion detection systems using machine and deep learning in internet of things: challenges, solutions and future directions. Electronics 9(7):1177
19. Azmoodeh A, Dehghantanha A, Choo KKR (2018) Robust malware detection for internet of (battlefield) things devices using deep eigenspace learning. IEEE Trans Sustain Comput 4(1):88–95
20. Mao J et al (2019) Phishing page detection via learning classifiers from page layout feature. EURASIP J Wirel Commun Netw 2019(1):1–14.
21. Moti Z et al (2021) Generative adversarial network to detect unseen internet of things malware. Ad Hoc Netw 122(2021):102591
22. Shafiq M et al (2020) CorrAUC: a malicious bot-IoT traffic detection method in IoT network using machine-learning techniques. IEEE Internet of Things J 8(5):3242–3254
23. Abdulwahab HM, Ajitha S, Saif MAN (2022) Feature selection techniques in the context of big data: taxonomy and analysis. Appl Intell 1–46
24. Sarkar S, Liu J, Jovanov E (2019) A robust algorithm for sniffing ble long-lived connections in real-time. In: 2019 IEEE global communications conference (GLOBECOM). IEEE
25. Goel A et al (2019) DeepRing: protecting deep neural network with blockchain. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops
26. Aref MA, Jayaweera SK, Machuzak S (2017) Multi-agent reinforcement learning based cognitive anti-jamming. In: 2017 IEEE wireless communications and networking conference (WCNC). IEEE
27. Han G, Xiao L, Vincent Poor H (2017) Two-dimensional anti-jamming communication based on deep reinforcement learning. In: 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE
28. Xiao L, Wan X, Han Z (2017) PHY-layer authentication with multiple landmarks with reduced overhead. IEEE Trans Wireless Commun 17(3):1676–1687
29. Shi C et al (2017) Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT. In: Proceedings of the 18th ACM ınternational symposium on mobile Ad Hoc networking and computing
30. Singh R, Singh J, Singh R (2017) Fuzzy based advanced hybrid intrusion detection system to detect malicious nodes in wireless sensor networks. Wirel Commun Mobile Comput 2017
31. Roopak M, Tian GY, Chambers J (2019) Deep learning models for cyber security in IoT networks. In: 2019 IEEE 9th annual computing and communication workshop and conference (CCWC). IEEE

32. Napiah MN et al (2018) Compression header analyzer intrusion detection system (CHA-IDS) for 6LoWPAN communication protocol. IEEE Access 6:16623–16638
33. Li Q et al (2017) Channel-based sybil detection in industrial wireless sensor networks: a multi-kernel approach. In; GLOBECOM 2017–2017 IEEE global communications conference. IEEE
34. Dovom EM et al (2019) Fuzzy pattern tree for edge malware detection and categorization in IoT. J Syst Architect 97:1–7
35. Parra GDLT et al (2020) Detecting Internet of Things attacks using distributed deep learning. J Netw Comput Appl 163:102662
36. Ahmad Z et al (2021) Anomaly detection using deep neural network for IoT architecture. Appl Sci 11(15):7050
37. Kamel DK (2021) Wireless IoT with blockchain-enabled technology amidst attacks. IRO J Sustain Wirel Syst 2(3):133–137
38. Amrish R, Bavapriyan K, Gopinaath V, Jawahar A, Vinoth Kumar C (2022) DDoS detection using machine learning techniques. J IoT Soc Mob Analytics Cloud 4(1):24–32

An Extensive Study on Logic Emerging IoT Adiabatic Techniques for Low-Power Circuit

T. Vijayalakshmi and J. Selvakumar

Abstract A low-power, energy-efficient circuit is essential for IoT edge devices, which increasingly perform data-intensive applications. Nanometer technology nodes push standard CMOS to its limits, which include increased leakage and increased power consumption. Appropriate algorithms for low-power circuits include adiabatic logic and approximation computing. It is possible to construct circuits that are more energy efficient by using adiabatic logic. The adiabatic logic's dual-rail construction and power clock approach, on the other hand, increase the overall footprint. More power is conserved by lowering the circuit's complexity and size while utilising approximation computing. For the Internet of Things (IoT), energy efficiency, and security, adiabatic circuits have the potential to work together. IoT-RF-powered devices can benefit greatly from adiabatic circuits even though they have been around for more than six decades, as demonstrated by some of the recent advancements. These enhancements are described in detail, with an emphasis on the main design challenges and opportunities associated with adiabatic circuits.

Keywords Adiabatic circuits · Nanometer · Internet of Things · Power clock · Energy consumption

1 Introduction

IoT refers to a network of interconnected devices that communicate data in order to facilitate the development of intelligent software. Smart homes, smart cities, and health care all use IoT strategies to monitor, anticipate, and regulate the world around us, which is a huge benefit. Mobile phones and wearables such as thermometers and moisture sensors (to name a few) are just a few examples of the wide variety of devices

T. Vijayalakshmi (✉) · J. Selvakumar
Department of ECE, SRMIST, Kattankulathur, Chennai, India
e-mail: vt1586@srmist.edu.in

J. Selvakumar
e-mail: selvakuj@srmist.edu.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*, Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_17

that fall into this category [1–3]. Data is collected by IoT devices and transferred to the cloud for processing. As a result, the network infrastructure is under greater strain as a result of the increasing volume of IoT edge devices and the data they generate. By 2022, there will be more than 28 billion IoT devices, and by 2025, there will be 75.44 billion [4, 5]. As a result, a novel paradigm for dealing with the massive amounts of data produced by IoT edge strategies is required.

An emerging computing example known as edge computing is one in which computation takes place at the network's edge. In the case of edge, nodes are located close to the end-user. As a result, they are less influenced have a low latency. As a result, the cloud server receives less sensitive data [5]. Consequently, researchers can use edge computing to create [6]. In order to create deep learning at the edge, IoT devices will need to have more processing power, which will increase power dissipation. In addition, the computing and energy resources of most IoT devices are constrained. As a result, IoT devices. Figure 1 shows the advantages of using adiabatic logic and approximation computing to create low-power and energy-efficient Internet of Things devices. Since battery life is a concern for mobile devices, the adiabatic approach is an absolute necessity, where adiabatic logic can be used to solve the need for long battery life in portable electronics [7]. It has been discovered that the most promising approach to increasing the battery life of integrated circuit devices is adiabatic logic. A long-lasting battery life is needed because of the rise in demand for handheld devices. More switching activity can be expected due to the fact that today's ICs operate at extremely fast speeds. The significant power dissipation means that big heat sinks are required. As a result, the gadget must be larger. Because of this, we must devise methods that utilise the least amount of energy possible. When it comes to portability, the biggest obstacle is power consumption. In a conventional CMOS circuit, power dissipation can take the following forms: (1) Dynamic Power Dissipation, which refers to the power expended by a device while it is switching. Static Power Dissipation (SPD) is the unpowered [8]. In the VLSI design process, there are a variety of ways to reduce the amount of power consumed. The scale of the supply voltage has also been altered to reduce power consumption. As a result, the circuit's speed is reduced when the supply voltage is reduced.

2 Background on Adiabatic Circuits

Irreversibility and heat generation were ideas that physicist Landauer considered in the 1960s when introducing adiabatic circuits for computers [9]. Adiabatic operations are defined as those that do not raise the environment's entropy during their execution [10]. Adiabatic operation can be difficult and impossible to achieve because of the sluggish current flow and inherent losses in the process in standard CMOS devices [11–13]. However, new logic families have been proposed to take advantage of adiabatic operation to reduce power consumption.

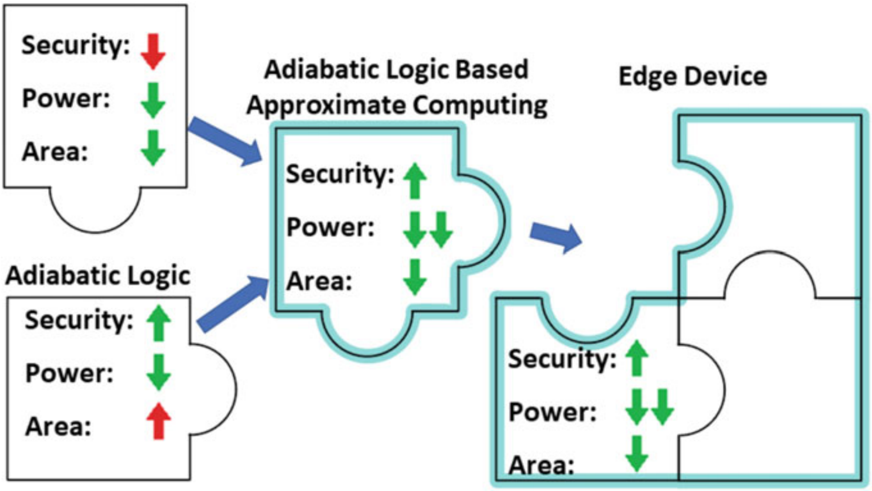


Fig. 1 Uniting the estimate computing and logic

A trapezoidal or sinusoidal power supply signal is the basic characteristic of these logic families, even if they range greatly in their ability to achieve full adiabatic operation. Since it synchronises the flow of data and is sometimes mentioned to as the power-clock signal, this signal also serves as the adiabatic circuit's clock [14]. Figure 2 depicts the adiabatic equivalent circuit for a given operation. The transistor's on-resistance and interconnect resistance are represented by R and C, respectively, while the output wire's capacitance is represented by C. A time of t_r is the power supply signal. Power loss across R can be reduced if time r is long enough to keep up with the time constant of the resistors in the circuit. When charging and discharging are taken into account, the total switching energy dissipated each cycle is calculated.

$$E_{ad}^{swi} = 2\frac{RC}{t_r} C V_{dd}^2 \tag{1}$$

Unlike normal static CMOS based process, in adiabatic process, a longer transition period lessens the overall switching energy, as indicated by (1). at which point in time the switching energy required by static CMOS process has reached a critical value $\left(E_{st} = \frac{1}{2}\alpha C V_{dd}^2\right)$ is Energy consumed by adiabatic operation is equal to E_{st} and can be found by comparing it to (Eq. 1).

$$t_r^{crit} = 4\frac{RC}{\alpha} \tag{2}$$

where is a measure of the amount of switching activity. As a result, adiabatic circuits use less switching energy than traditional circuits when t_r is greater than t_r^{crit}. As a result, adiabatic operation is well-suited to applications with low operating moderate

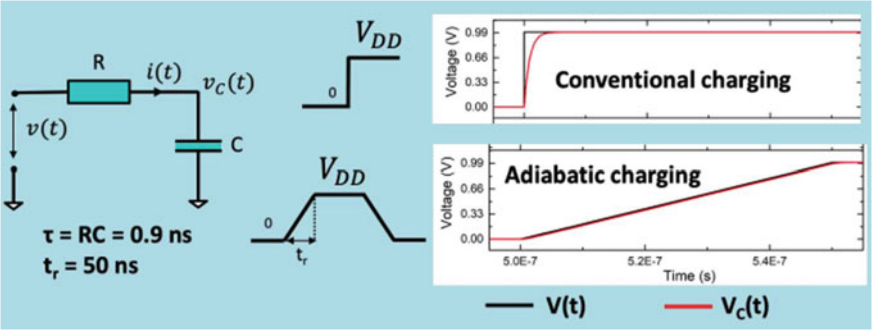


Fig. 2 Supply signal versus traditional charging with a continuous DC voltage

to high activity levels. At low frequencies, leakage energy can take upsurge overall energy consumption in adiabatic operation, thus it's crucial to keep this in mind. Because of this, a particular frequency has a significant impact on the amount of energy used overall [15, 16]. It can range from a few thousandths of a kilohertz to tens of thousands of kilohertz, depending on the technology. Because the output charges and recuperates each time the power supply signal is applied, it is practicable to power gate adiabatic logic circuits during idle mode to reduce the amount of energy that is dissipated. The power-clock generator's design characteristics should be considered when developing the power-gating circuitry.

3 Recent Developments

3.1 Adiabatic Circuits for RF-Powered IoT Applications

RF power harvesting IoT devices could benefit greatly from adiabatic circuits. RFID-based systems, which have traditionally had limited computer capabilities, are two examples of these applications. It is possible to adiabatically drive a digital logic in these devices that is already powered by RF because the wirelessly gathered signal is already in sinusoidal form For one thing, the difficulties of generating a power clock signal are lessened, significant power loss associated with the rectification process is eliminated, due to adiabatic operation when using this approach to improve RF-powered logic's energy efficiency, which is a key concern in this design.

Wireless link based on near-field inductive link for adiabatically powering an 8-bit ALU built in several families of adiabatic logic utilising 65 nm technology [17]. At 13.56 MHz, the wireless link transmits power at a dBm level of 24. At a distance of 6.5 cm, the power efficiency drops to − 37.4 dB, the simulated findings show a reduction in power usage of up to 30. Pass transistor adiabatic logic (PAL) required two out-of-phase power-clock signals to achieve this substantial decrease in power

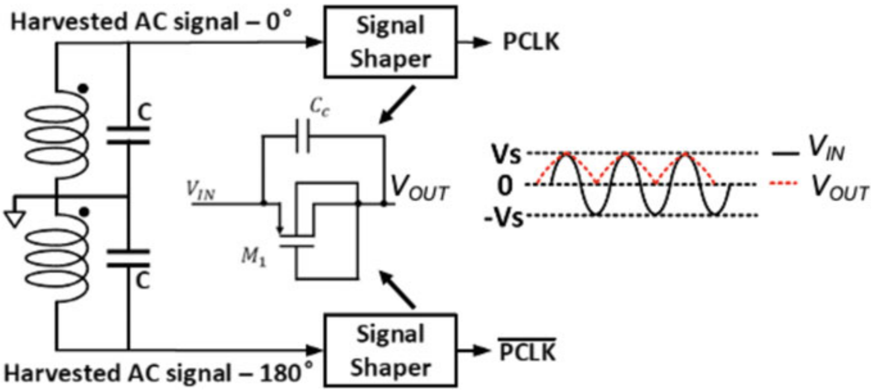


Fig. 3 In the case of a wirelessly powered application, directly harvesting from the wireless link the requisite power-clock signals for adiabatic logic

[17]. According to Fig. 3, two receiving coils with a 180-degree phase difference capture these two signals directly. In order to assure proper operation, the negative components of the wirelessly acquired signal must be eliminated. Figure 3 shows an example of a capacitor signal shaper that can accomplish this goal [18].

The output nodes of PAL stay fluctuating for a brief amount of time during process, which reduces the robustness of the architecture. As ECRL is more robust and allows for lower voltages, we've looked into using it for RF-powered applications as well [19]. ECRL, on the other hand, necessitates a 90° phase difference between each of the four power-clock signals. With passive LC components, a phase shifter is needed [20]. In order to reduce resistive loss, the size of these passive devices is increased, especially at low frequencies. For wireless power harvesting, existing adiabatic logic families have intriguing tradeoffs.

3.2 Adiabatic Logic Enhanced Hardware Security

Adiabatic circuits are more resistant to side- attacks in terms of hardware security. The lower SNR of these devices reduces the quantity of side-channel leakage [21, 22]. During the adiabatic operation recovery phase, there is a way for leaking. pMOS transistor's turn off when the output voltage approaches threshold. Consequently, during the subsequent charge cycle, less current from the power source flows, allowing information on the prior input signals to leak out.

The adiabatic operation is dependent on the use of ECRL with four-phase power clock signals developed in 65 nm technology. To put it another way: When using, the encryption efficiency (measured in Kb/sec/W) is boosted by roughly five. By reducing throughput by 18% and increasing physical area slightly, this large improvement in efficiency was made possible (2%). In addition, we used a correlation power analysis

(CPA) approach to figure out an adiabatic SIMON core's power-based side channel resistance [23–25]. As a result, we calculated correlation coefficients. Static CMOS-based unprotected SIMON core has an MTD of 1354 while unprotected adiabatic version has an MTD of 5718. It is therefore more than four times more resistant to power-based side-channel examination assaults if the adiabatic implementation is used in place.

4 Avoid Common Mistakes to Adiabatic Logic Design

4.1 Don't Use Diodes

True adiabatic design is all about avoiding diodes at all costs. It's essential to remember that if a diode is used as an essential functional element in a circuit, it will have to be substituted in the future as energy efficiency requirements grow more stringent. This is especially true when compared to, for instance, junction diodes, because they are inherently thermodynamically irreversible and generate an irreducible amount of entropy whenever they are used in that capacity. There is a built-in voltage drop across diodes, and this "diode drop" leads in an irreversible loss of QV for a given charge Q carried across the diode. For example, In other words, a dissipation-less diode is equivalent to an "Energy Maxwell's demon Hamiltonian dynamics laws embedded in all modern physics through quantum mechanics [26].

Diodes were commonly utilised in the charge return path of early adiabatic circuit designs, starting with Watkins. Until we can achieve even greater energy efficiency, this strategy may be beneficial if the diode drop is smaller than logic voltage swings.

4.2 Rules of Don't Disobey Transistor

Even if diodes are non-adiabatic, transistors can still be used for adiabatic operation as long as two basic principles are adhered to, even though they're not ideal switches:

- Never switch on a transistor if its source and drain terminals have a considerable voltage differential.
- When a transistor's channel is carrying a considerable amount of current, it should never be turned off.

As for the first rule, it should be self-evident. For example, we know that a dissipation of 12CV2 occurs when a dynamic node of capacitance C is connected directly to a static reference signal of voltage that differs by V. This dissipation is still 14CV2 when they converge to their average level even when both nodes are isolated and have capacitance C. Turning on the transistor results in a continuous power dissipation if the two nodes are coupled to voltage sources that are different from each other. It is generally impossible to verify that the voltages before the transistor is turned on are exactly the same in most adiabatic logic types because of noise issues, leakage, etc. However, we should strive to come as near to a match as feasible.

In fact, when the gate-to-source voltage passes a certain threshold, transistors aren't ideal switches, and they don't immediately transition from "perfect-on" to "perfect-off" (however slowly). A lossless diode can be built using perfect switches that are thermodynamically impossible [26]. In SCRL, however, the circuit layout may be easily modified so that this behaviour is avoided [27] by connecting the gate of a third complementary-type transistor to the gate of the outer FET in parallel with the inner FET. In order to minimise unwanted non-adiabatic dissipation while charging the source node, this second FET's strong on-state diverts the source node's current from the FET that is turning off, resulting in a slight voltage drop across it. Dual rail logic, which enables the designer to provide full-swing charging paths everywhere, can be used to avoid the rule being violated in inverted Boolean gate pull-up/pull-down networks of any desired complexity.

However, this scenario demonstrates that even if one is aware of the rule (2), one might readily break it by unintentionally. If gate-activated or source-activated transistor switch offs are ever used, it is imperative that designers be made aware of and emphasise the adiabatic circuits rule. Voltages at both the source and drain nodes must remain constant or fluctuate along identical paths when the transistor is turned off.

4.3 Use Reversible Logic

A large number of so-called adiabatic logic styles are not expressly nor implicitly irreversible. A direct proof of the link can be found in modern physics' most fundamental and well-known rules.

The Schrödinger wave equation is a fundamental part of the model's foundation as a quantum theory. To demonstrate the reversibility (bi-objectivity) of the time-evolution, any Hamiltonian dynamical system's state variables (i.e., the amplitudes of wave functions in quantum theory) are subject to a first-order differential equation. Even seemingly irreversible occurrences such as are explained away in pure quantum theory as predictable emergent phenomena predicted by a totally reversible underlying theory [28]. No deviations from this micro-reversibility have ever been seen. Reversible evolution of a microstate in interaction with an unknown environment cannot be tracked precisely by the modeller, and this results in macro-scale irreversibility. The link between adiabatic and reversible logic is obvious when considering the reversibility of mainstream quantum physics.

So, for example, let's say that an algorithm claims it can operate on a bit that can take one of two possible values and convert that bit to one with a single, unambiguous value. The number of possible physical states will not be reduced by this action. Because the logical content of the bit no longer distinguishes the states, other physical factors, such as the thermal vibrational state of atoms in the immediate environment, must now be used to separate them. Physical information in the system as a whole has not been altered at all. After erasing a given bit's state, the information about the given bit's state that was previously known is now unknown, which is exactly what entropy is: unknown information. So the total entropy has risen, as a result of this. State space size has risen by 2 times, hence this is a logarithm of the increase in state space. This value of entropy is equivalent to the Boltzmann constant k ln 2 when expressed in logarithm base e units.

Semi-adiabatic techniques, on the other hand, do not scale to ever-lower levels of dissipation when technical improvements or speed decreases, even if they are not conceptually reversible.

4.4 Don't Over-Constrain the Design

A number of logic systems have been proposed for creating large-scale circuits that approach true adiabaticity (and consequently total logical reversibility). The costs (in terms of space, time, and energy) of doing certain computations with existing schemes are asymptotically higher by unboundedly huge factors than with alternative schemes.

Adiabatic logic families [29], don't allow reversibility over numerous levels of sequential, pipelined logic. The SCRL project of Younis and Knight [30, 31] aimed to address this specific issue. A pipelined, reversible, fully adiabatic logic was provided by them. SCRL, on the other hand, turns out to be overly restrictive because it necessitates a transition for every logic node on every clock cycle. When implemented in SCRL, even memory that are simply storing data conduct active logic transitions for each stored bit [32]. Since SCRL does not enable stored bits to remain idle, it is asymptotically less cost-effective than an alternate adiabatic technique.

5 Conclusion

For long-term decades of cost-effective computing, which provides necessitates for a close approach to the real physical limits of computing. This is especially true when considering the requirement for nearly total physical (and thus logical) reversibility of the computing mechanism. These constraints rule logic circuit designs and transistors (or other current switches) must be monitored for their current carrying status when the device is turned off, as well as their voltage state when the device is turned on. Another requirement is that hardware algorithms of the highest asymptotically efficient performance can be expressed using this logic family. In the coming century, this forecast is as certain as the confluence of modern physics' and technology's juggernaut march. Meeting all of these conditions will be an absolute economic necessity. None of the adiabatic logic families created by other researchers have met all of these criteria. An order of magnitude improvement in energy efficiency can be achieved by using adiabatic circuits with frequencies ranging from tens of megahertz to hundreds of megahertz. Furthermore, adiabatic logic is better able to withstand attacks from power-based side channels, which makes it excellent for IoT devices with limited resources. Some recent advances in power-clock, performance constraints, and project automation are summarised in this study. The application of adiabatic logic in wireless powered systems has been shown.

References

1. Caro F, Sadr R (2019) The internet of things (IoT) in retail: bridging supply and demand. Bus Horiz 62(1):47–54
2. Chen J, Ran X (2019) Deep learning with edge computing: a review. Proc IEEE 107(8):1655–1674
3. Madushanki AAR, Halgamuge MN, Wirasagoda WAHS, Ali S (2019) Adoption of the internet of things (IoT) in agriculture and smart farming towards urban greening: a review
4. Networking CV (2016) Cisco global cloud index: forecast and methodology, 2015–2020. White paper. Cisco Public, San Jose, 2016
5. Yu W, Liang F, He X, Hatcher WG, Lu C, Lin J, Yang X (2019) A survey on the edge computing for the internet of things. IEEE Access 6:6900–6919
6. Azar J, Makhoul A, Barhamgi M, Couturier R (2019) An energy efficient IoT data compression approach for edge machine learning. Future Gener Comput Syst 96:168–175
7. Moon Y, Jeong D-K (1996) An efficient charge recovery logic circuit. IEEE J Solid-State Circuits 31(4)
8. Rajesh A, Raju BL, Reddy KCK (2014) Reduction of power dissipation & parameter variation in VlSI circuits for SOC. Int J Rev Electron Commun Eng (IJRECE) 2(3)
9. Landauer R (1961) Irreversibility and heat generation in the computing process. IBM J Res Dev 5(3):183–191
10. Frank MP (2005) Introduction to reversible computing: motivation, progress, and challenges. In: Conferences on computing frontiers, pp 385–390
11. Moon Y, Jeong D-K (1996) An efficient charge recovery logic circuit. IEEE J Solid-State Circuits 31(4):514–522
12. Gong C-SA, Shiue M-T, Hong C-T, Yao K-W (2008) Analysis and design of an efficient irreversible energy recovery logic in 0.18-m cmos. IEEE Trans Circuits Syst I Regul Pap 55(9):2595–2607
13. Maksimovic D, Oklobdzija V, Nikolic B, Current KW (1977) Clocked cmos adiabatic logic with integrated single-phase power-clock supply: experimental results. In: International symposium on low power electronics and design, pp 323–327
14. Celis-Cordova R et al (2019) Design of a 16-bit adiabatic microprocessor. In: IEEE international conference on rebooting computing, 2019, pp 1–4
15. Wan T, Karimi Y, Stanacević M, Salman E (2017) Perspective paper—can AC computing be an alternative for wirelessly powered iot devices? IEEE Embed Syst Lett 9(1):13–16
16. Wan T, Karimi Y, Stanacevic M, Salman E (2017) Energy efficient AC computing methodology for wirelessly powered IoT devices. In IEEE Int Symp Circuits Syst
17. Wan T, Karimi Y, Stanaćević M, Salman E (2019) Ac computing methodology for RF-powered IoT devices. IEEE Trans Very Large-Scale Integr Syst 27(5):1017–1028
18. Huang Y, Wan T, Salman E, Stanacevic M (2019) Signal shaping at interface of wireless power harvesting and ac computational logic. In: IEEE international symposium on circuits and systems, May 2019
19. Moon Y, Jeong D-K (1996) An efficient charge recovery logic circuit. Solid-State Circuits IEEE J 31(4):514–522

20. Wan T, Salman E, Stanacevic M (2016) A new circuit design framework for IoT devices: charge recycling with wireless power harvesting. In: IEEE international symposium on circuits and systems, May 2016
21. Avital M et al (2015) DPA-secured quasi-adiabatic logic (SQAL) for low-power passive RFID tags employing S-boxes. IEEE Trans Circuits Syst I Regul Pap 62(1):149–156
22. Lu S, Zhang Z, Papaefthymiou M (2015) 1.32 GHz high-throughput charge-recovery AES core with resistance to DPA attacks. In: IEEE symposium on VLSI circuits, June 2015, pp C246–C247
23. Kumar SD, Thapliyal H, Mohammad A (2018) Finsal: finfetbased secure adiabatic logic for energy-efficient and DPA resistant IoT devices. IEEE Trans Comput-Aided Des Integr Circ Syst 37(1):110–122
24. Wan T, Salman E (2018) Ultra low power simon core for lightweight encryption. In: IEEE International symposium on circuits and systems, May 2018
25. Leff HS, Rex AF (eds) (2003) Maxwell's Demon 2: entropy, classical and quantum information, computing. Institute of Physics Publishing
26. Schlaffer A, Nossek JA, Is there a connection between adiabatic switching and reversible computing? Institute for Network Theory and Circuit Design, Munich University of Technology, http://citeseer.nj.nec.com/schlaffer97is.html
27. Frank MP (1999) Reversibility for efficient computing, manuscript based on Ph.D. thesis, Dec. 1999, http://www.cise.ufl.edu/~mpf/-manuscript, §7.6.4, pp 197–199
28. Zurek WH (2002) Decoherence, einselection, and the quantum origins of the classical. Preprint http://arxiv.org/abs/quant-ph/-0105127
29. Hall JS (1992) An electroid switching model for reversible computer architectures. In: PhysComp '92 (ibid. [2]), pp 237–247
30. Younis SG, Knight TF Jr (1994) Asymptotically zero energy split-level charge recovery logic. In: International workshop on low power design, pp 177–182. http://www.cise.ufl.edu/~mpf/-scrl94.pdf
31. Younis SG (1994) Asymptotically zero energy computing using split-level charge recovery logic. Ph.D. thesis, MIT EECS Dept. http://www.cise.ufl.edu/~mpf/younis-phd.ps
32. Vieri C, Ammer MJ, Amory Wakefield L (1998) "Johnny" Svensson, William Athas, and tom knight "designing reversible memory," unconventional models of computation, Springer, pp 386–405

A Critical Review of Agri-Food Supply Management with Traceability and Transparency Using Blockchain Technology

Sanket Araballi and P. Devaki

Abstract Agriculture is the backbone of our society; hence, the Indian economy is heavily reliant on farmers. Farmers are in charge of crop cultivation and account for around 51% of all agricultural production. Despite their contributions, individuals do not realise the benefits or earn sufficient profit for a variety of reasons, including a lack of understanding and supply chain management inefficiency. Smart technologies that require supply chain management models are used to tackle these difficulties. This aids in the financial transaction's monitoring at each stage. Blockchain technology has recently emerged as a transparent supply chain management platform. The goal of this study is to show how various supply chain management systems can track their transparency. We show the various problems with the current system in this review for a traceable transaction that can help farmers in tracing the financial transaction. Furthermore, AI (Artificial Intelligence) is suggested for future research direction.

Keywords Transaction traceability · Agri-Food

1 Introduction

Agriculture is the world's backbone, having a direct impact on a variety of fields, ultimately influencing the main purpose of human existence. Agriculture is important for a country's economy, as well as providing security, and nutritional value, and promoting population health. Agriculture carries several risks, such as changing weather patterns from season to season. Agriculture products' costs fluctuate dramatically due to soil degradation, weeds, pests, and non-sustainable crops, which affect yield and Climate change. Figure 1 shows the Agri-Food Supply management and explained later.

S. Araballi (✉) · P. Devaki
Department of Information Science and Engineering, NIE, Mysuru, Karnataka, India
e-mail: sanket.araballi@gmail.com

P. Devaki
e-mail: devaki@nie.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_18

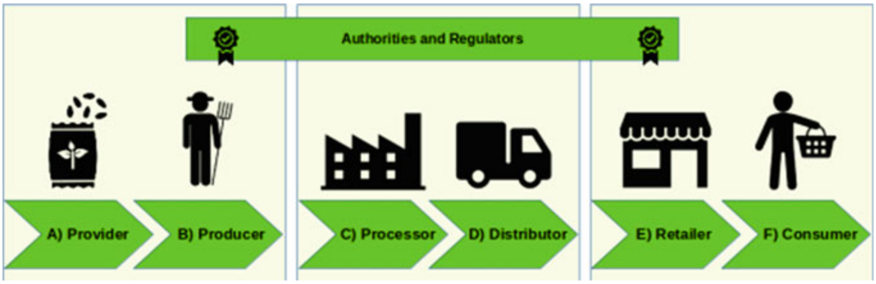


Fig. 1 Simplified version of the Agri-Food supply chain management process

The above figure i.e. Figure 1 shows the simple version of the supply chain management process, the roles of the actors involved are stated below [1]:

(A) **Provider**: is responsible for providing raw materials like seeds, nutrients, pesticides chemical fertilizers etc.
(B) **Producer**: the farmer is responsible for actions starting from the planting of seedlings to harvesting the crops.
(C) **Processor**: this actor is involved in the simple processing of complex tasks.
(D) **Distributor**: The Distributor is accountable to transfer the end product of the processor from the processors' stage to the site of retailers.
(E) **Retailer**: The retailer is accountable to sell the products, by representing in provisional stores to supermarkets.
(F) **Consumer**: This actor is the end output in the chain, who act laterally through the entire process, the authorities deliver standard measures, regulations, laws, rules and policies that are added and conform with.

Furthermore, there are various phenomena and process which is carried out for ideal farming [2–4]; these are explained as follows:

i. **Raw materials**: The producers and the providers stock blockchain technology with sales information and purchase history of the raw materials supplied, which include technical data of the products and their price. To automatize the process smart tags are used.
ii. **Planting**: The producers stock up the blockchain data around the planting process e.g.; the total calculation of seeds used; the sensors automate this data entry process. By connecting various weights for smart contracts that spontaneously explode, that create accounts when an anomaly is sensed (more seeds need to be registered unlike those that are purchased).
iii. **Growing**: The sensors are planted at various points that separately store the blockchain data to keep track of the plants grown and the environment considered. Smart contracts that track records when an anomaly is detected, sensor values are said to be out of threshold.
iv. **Farming**: Farmers are responsible for storing the blockchain data at every stage in the system model, like the amount of input applicable. Sensors are capable

of automating the entry process for data generated, which involves chemical sensors and multisensory systems that include smart contracts which automatically detect fire henceforth, create records when an anomaly is detected, and sensor values achieved outside of the threshold.

v. **Harvesting**: the farmers are responsible to store blockchain data about harvesting. Sensors can automate the data entry connected through scales. Smart contracts are responsible for autonomous fire, henceforth to certify the process right from seed harvesting that is compliant with certain regulations.

vi. **Delivery to the processor**: farmers are responsible for transferring the ownership to the local distributors, via the blockchain. Sensors and smart contracts responsible to automate the process create records when anomalies are sensed in the delivery phase.

vii. **Processing**: Consider the simple scenario of packaging a processor, the latter stores the blockchain technology information approximately by the amount received by the product or distributors. The packed amount of the product is misplaced during the processing phase. Smart contracts automate the data or create a new process when an anomaly is detected during the delivery phase.

viii. **Retailing**: retailers are responsible to store the blockchain data about the amount received as a product from distributors. At regular intervals sensors autonomously save the blockchain information of the status details of the retail environment. Note: smart contacts asynchronously fire, for certain records when an anomaly is detected.

ix. **Consuming**: retail shopkeepers store the blockchain details of the products sold, as these customers validate the history of a product before purchasing it. Smart tags are connected along with each package for consumers the regain the end system product.

Traceability

In Agriculture, Transaction traceability anticipates a significant volume of data created by the supply chain. Employees in the early track and traceability system take track of data in the field, which is later manually recorded to a handbook or into a computer model [5]. This approach poses risks that as erroneous data recording and ineffective resource usage. In the last decade, we saw a rapid increase in communication skills to automate processes and products, resulting in the Internet of Things (IoT) framework [6]. Figure 2 shows the agriculture traceability with respect to IoT. The quick expansion of (the Internet of things) and sensor technology favours data collection methods that are both fast and reliable. This system combines blockchain technology for identifying the product, analysing the ingredient, transport, and data recording. This strategy integrates blockchain technology in identifying the product, analysing the ingredient, shipping, and data gathering, along with overall system integration. Various supply chains take advantage of technologies such as barcodes, QR codes, and RFID (WSNs). RFID systems protect information and data for farmers, distributors, merchants, and consumer management solutions for agri-foods [7]. To trace and monitor the “farm to fork route,” RFID technology aids in the management

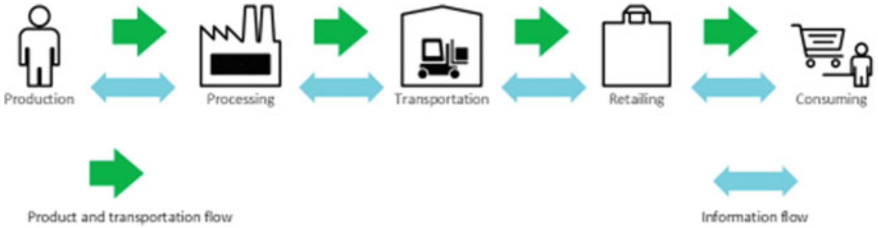


Fig. 2 Agriculture traceability

of the agri-food supply chain. When a problem with food safety emerges, the source of the problem and, as a result, the solution is discovered quickly. Blockchain technology incorporates storage of the transactional data and then saves and simplifies programs to implement this. To enable this application so that it is distributed [7]. In the article founder of Ethereum said that decentralization refers to three levels of (de) centralization.

1. Logical (de) centralization;
2. Political (de) centralization;
3. Architectural (de) centralization [8].

To this day many frameworks utilize the certification process that ensures not deliver complete distinguishability to the adjacent nodes of the network. To recognize the essential problem in supply chain management applications, such as product supply chains, in a setting where “smart contracts” can be used in a promiscuous manner. Nick Szabo coined the phrase “smart contract” in 1997. His main idea was smart contracts similar to contracts in the real world with a slight difference from originality. These are digital, small programs applied and kept in blockchains. To feature this break-free logical code into it. Smart contracts are responsible to get a few attributes, they are absolute and spread across blockchain technology. This absolute nature ensures that we cannot meddle with the code of the contract. For being distributed this secure evaluation of smart contracts are similar to blockchain applications like that in agri-food supply chains and the health sector [9]. When an operation is scanned and found thoroughly in a complete digitalized way. The exact operation is confined to that block and added to the supply chain. Once the agreement is executed at the chain. A statement issued by a variety of information relevant to a blockchain is added [10]. Each consumer in the network holds a duplicate of the smart contract resulting in the following:

1. History of all smart contracts,
2. History of all transactions;
3. The current state of all smart contracts [11].

Ethereum is implemented for smart contract sustenance. As a result, numerous programming languages enable software coding processes in conjunction with blockchain technology.

A widely used tool to support Ethereum's solidity programming language, a Turing-complete language that establishes the sequence of particular rules that govern how a programme operates and executes [11].

1.1 Motivation and Contribution of Research Work

In general, supply chain management is focused on a traceability model that makes use of blockchain's features, in which all stakeholders access each operation and data connected to a particular product. The main objective is to cover the entire data collection and management process. For each transaction involving participants in the agriculture supply chain. This entire system is known as a "farm-to-fork" approach since it allows to monitor, track, and trace the value of agricultural products. In comparison to centralized methods, Tian's solution has both benefits and drawbacks. Thus, motivated by the problem faced by farmers for transparency, this research review conducts a critical review of supply chain management with an aspect of transaction traceability and transparency.

2 Related Work

The Agri-food supply chain system includes several levels of transactions, each level has a dissimilar set of terms and conditions [12]. Various systems have different features and the function has been working together with supply-chain, which includes food processing, transportation, storage and distributors. Following that, an examination of the electronic traceability systems for agricultural products that describe the condition described above was conducted. As blockchain technology matures and enters with a variety of applications, research is carried out extensively that concentrate using of DTL for agriculture and traceability systems. Table 1 shows the comparison of various mechanisms based on different parameters like traceability, Cost, Inventory Management and so on.

2.1 Traceability

As a supply chain, blockchain allows users to track agri-food items among stakeholders. Feng et al. [1] have analyzed the traceability issue, blockchain addresses on coordinating, verifying, linking and recording transactions. Blockchain lets trace the products from farms to consumers, [2, 12]. Chan et al. [5] proposed an Information asymmetry which is addressed by an Agri-food supply chain management software built on blockchain for traceability and transparency, allowing information to be shared with supply chain stakeholders. Stakeholders can audit transactions on

Table 1 Mechanism comparison

Review basis	Significance	Papers
Traceability	As a supply chain, blockchain allows users to track agri-food items among stakeholders that have analyzed the traceability issue, blockchain addresses coordinating, verifying, linking and recording transactions. Blockchain lets trace the products from farms to consumers	[1, 2, 5, 7–9, 12]
Efficiency	Blockchain increases the performance of the business through minimal effort	[5, 10, 11, 13, 14]
Privacy	To entail business matters confidential blockchain ensures security property and anonymity to maintain relationships undisclosed	[15–18]
Cost	Blockchain technology helps reduce transaction fees and ease fair prices throughout the chain	[5, 8, 15, 19–22]
Inventory management	The procedure of observing the transfer of goods initiated from production houses to warehouses and transferring these provisions to retail shops is a primary aspect of supply chain management	[23–25]

the blockchain, in addition to providing transparency [6]. Because the blockchain is absolute, the data in it cannot be modified with [7–9]. The transaction history can be viewed and copied by any valid user [8]. They have discussed using the blockchain in food authenticity, which ensures that no changes were made. When it was in the hands of a specific party, a variety of things could happen. As a result, blockchains could be used to combat food theft and improve traceability. However, data collection is required for traceability. Food traceability is complicated by a lack of records [8]. Without telling blockchain consumers, products can be intentionally harmed [2]. Furthermore, ensuring the validity, confidentiality, and integrity of raw input data is problematic [1]. The accuracy of data collected by sensors or by people is not guaranteed. Besides, these analytical methods are unable to screen various food product parameters, particularly environmental parameters. Third parties, like the government, tracking the blockchain network to solve the issue of data manipulation.

2.2 *Efficiency*

Blockchain increases the performance of the business through minimal effort. Leng et al. [11] and Hasan et al. [12] suggested that blockchain technology increases the complete efficiency, throughput, and reliability of relevant frameworks, this eases the growth of the business. Blockchain is responsible for digitizing the product and certifications that provide real-time information about food products while reducing the time it takes to trace from a week to a few seconds. This functionality enables tracing the contaminated products based on plant or animal disease outbreaks [5]. Human intervention is reduced by trust and self-organizing nature while integrating IoT devices to improve supply chain efficiency [13, 14]. For instance, uses blockchain to boost security independence which exhibits agricultural robotic swarm operations. They leverage blockchain technology to enable the information recorded for autonomic payment transfer. Blockchain technology spontaneously along the smart contracts could implement and assess transactions to stop. Further, added mutilation with cautions Applicants assess declarations carried out and inform when relevant parties give in to the condition such as quality, timing, and quantity. Smart contracts deal with mountable and easy business at a cheap price, to enhance the complete performance including the manufacturing services.

2.3 Privacy

To entail business matters confidential blockchain ensures security property and anonymity to maintain relationships undisclosed. Zhao et al. [15] stated blockchain technology enhances the confidentiality of the enterprise. Blockchain ensures features like anonymity, traceability, reliability, security and transaction visibility to ensure the correctness of agri-food products. Blockchain is associated with more privacy issues, in comparison with benefits that were briefly mentioned. Privacy can be comprised of permanent information to ensure data visibility transparency. Zhao et al. [15] have discussed many issues on privacy leakage and various efforts to ensure privacy. Consider, the example of stakeholders who may want to keep their information confidential to ensure business competitivity. Blockchain technology which with inhibits private or permissioned blockchain is preferred over a public blockchain. A challenge encountered in blockchain technology implementation is the accurate selection of the above technology. The privacy management and transactional privacy cannot be ensured because all participants are not competitors because all applicants in the network access the information [16–18]. There exists no single technique that hides all parties' involvement, as well as the transaction amount at the same pace. One of the drawbacks of blockchain technology is ensuring Privacy. These issues need to be considered for the management of data, essentially data ownership and data retention [16].

2.4 Cost

Blockchain technology helps reduce transaction fees and ease fair prices throughout the chain as suggested by Chan et al. [5]. Another challenge associated with blockchain technology is the cost of computing and expurgating the equipment for the smooth running of the system to adopt a blockchain. Adopting blockchain technology requires several transformations and variations in the organization that are expensive [15] and time-consuming. The cost of blockchain is a hindrance to adopting blockchain technology [19]. Many countries face complications since the high degree of computation essential [15]. Hence, [19] stated that blockchain technology is “SME” friendly. In the early stages, once the technology is advancing the high level of uncertainty and market volatility may result in challenges of adoption, because of the old mindset of many stakeholders as well as they may not be mindful of technology and its benefits (e.g., [5–7]) The ideology of blockchain exists that skills employees and organizations. Further satisfying business solutions are limited. Blockchain technology may impose risk on the public to ensure the volatility of cryptocurrencies [5]. The stakeholders adopt a partial education and necessary skills to adapt to blockchain technology [15]. The training framework still gives inadequate results [5]. Cost is sustainable in blockchain technology [19]. The adoption requirements are altered based on different phases of the chain [7, 12]. Hasan et al. [11] suggest minimal changes in the blockchain as a substitute for adopting a new one. Other technologies are most suitable in various cases [7, 12, 14]. Enhancing the knowledge of users’ understanding of blockchain improvises adoption nature [19]. The government plays a key role in minimizing the hindrance caused for adoption by setting different examples [5, 12]. To maintain the capability of businesses for sustaining the blockchain improvises the social, and economic environmental characteristics [5]. Blockchain aids in supervision that assists to abolish the adulteration of food, improvise food safety concerns enhancing the value and solving tough problems, and hence enhancing sustenance (e.g., [3, 8]). Blockchain technology helps allocation of resources [3], to increase the demand–supply and predict quality [4, 7]. Good supervision minimizes the loss and product waste [4], In addition, this provides support for emission reduction [5]. Water management uses blockchain technology to record water quality data to reach sustainability [5–7]. Economically blockchain can improvise the living condition of poor people in developing countries [18] and necessitate food security, (by giving food coupons [10]). Consumers are impacted more by environmental and legal work conditions [8, 20, 21]. Loop [22] stated that blockchain technology acts as a tool for keeping track of social and environmental issues blockchain technology can be used to influence environmental, social, and economic legislation, as well as reduce corruption. Misconduct is held accountable outside the organisation.

2.5 Inventory Management

The procedure of observing the transfer of goods initiated from production houses to warehouses and transferring these provisions to retail shops is a primary aspect of supply chain management. This technique isn’t worth the demand forecasting along with inventory management and both are interlinked. If the provider is not able to satisfy the need of customers, the company’s efforts are not utilized and the customers are not satisfied. Consecutively if the supply surpasses the demand, efficient raw materials are depleted for not giving extra profit because there is no possibility to trade additional inventory. High inventory levels accordingly, for various reasons that confine organizations for profit [23]. Further, the money lent rates increase the expenses of the inventory storage increases well. However, to calculate the supply volume before designing working plans, ML [24] techniques Such dangers should be avoided. Machine learning and data analytics approaches can help with labour management, inventory system automation, vehicle deployment plan optimization, and supply chain decisions that require agility. The ability to accurately estimate demand is crucial for supply chain planning and management [25]. While demand patterns for food remain stable, they demand ambiguity for leisure products like clothes and technology. Blockchain Implementation Challenges significantly.

3 Blockchain Implementation Challenges

From the above blockchain technology has gained wide importance and great extension through supply chain management. Blockchain technology and DLT solve most of the prevailing issues, however, these aren't a solution, to consider the datum of the agri-supply chain which is subtle in terms of criticism and, in many scenarios, they function under difficult pressure. Parallelly to this blockchain technology must also consider various social-technical and financial effects in the agri-supply chain sector or its undone rules and unbreakable relations that aren't considered. In a further elusive phrasing, the agriculture sector chain is a multifaceted task relevant to blockchain adoption for the following issues.

1. Stakeholders have low-level technical knowledge.
2. The products in the blockchain transform in their entirety.
3. The procedure of role and business allows a huge number of stakeholders that are involved is diverse;
4. The food supply chain is dispersed geographically spread across various continents throughout the globe, it is a fact of the matter for interoperation and deployment across various obstacles.

Issues relevant to information management specifically for information ownership and data retention in blockchain technology are essential and needed to be cautiously measured [3]. Blockchains are immutable collections, the data accuracy introduced by the sensors or manually or by people is not guaranteed. Hence, if a sensor breakdown, then the data through the blockchain is inaccurate [5]. These are challenges of monitoring, integrating, and evaluating a certain category of information along the supply chain, in general. Environmental data, for example, is difficult to navigate and analyze using objective methods [8]. Another problem arising here is through the collaboration and continuous integration of blockchain along with legacy traceability problems, to preserve blockchains parallelly through a distributed nature like DLT. The scalability of this transaction system sustains an open challenge which serves as the context for massive executions and deployment [3]. Thus, the proper selection of (public, private, permissioned) that implement the blockchain technology is considered an issue. Nevertheless, this benchmark of choice for which the performance is not discussed always. This is because agri-food supply chains involve multiple parties in the transaction's completion, which necessitates the use of a public blockchain for a significant period and compute capacity. In this regard, the sort of blockchain being explored should be carefully evaluated. This might affect the traceability of the performance of the system. From a financial aspect blockchain increases Because of the transparency of the agricultural food supply chain and consumer trust, utilization is based on a large quantity of energy and financial cost [1]. Companies are required to invest a significant amount of money and time. The amount of time it takes to train staff and get essential material. The time and money invested in traceability costs exceed the entire cost of the product in the short and long term, necessitating a cost–benefit analysis. Given the obstacles ahead, this is an intriguing element to explore. Blockchain applications in the agricultural supply chain were investigated. This research aids in the discovery of 18 food traceability boundary conditions. Five of them have worked with blockchain directly. A vast number of concerns, according to the authors, are related to regulatory requirements for the internal supply chain and industrial processes that necessitate organizational adjustments to offer sustenance while maximizing the aid of traceability. Access to blockchain technology, governance and sustainability, policy and regulation, and the key role of government as a leading example in digitalizing public administration are all important factors. These argue with investment in technology and education that is analysed for further production and validate proof by taking care of the benefits to further yield and demonstrate the proof given the advantages of new blockchain technology.

The previous sub-section gives an outline of the challenges of blockchain technology for the probable integration of the agriculture sector. Considering these challenges future research is to tackle and overcome these issues. Safety must be ensured on the blockchain technology, parallelly making it more flexible. Blockchain technology may also be considered an established method concerning financial technology and cryptocurrencies, as long as the supply chain is still deemed to be in its initial stage. As a result, in addition to technical difficulties and tasks, the additional debate will focus on the necessary legislation and policies for incorporating blockchain technology which is a brittle and sensitive sector as that like agriculture. Henceforth, the technologies which are developing faster involve AI by combining with DLT which leads to a smart agriculture platform, where services are rendered by various services, components and stakeholders which are interrelated. Smart agriculture has various benefits despite enhancing traceability systems, and more effective yield by using big data and machine learning algorithms. To carry out this study, needed to be driven toward a more realistic method for making platforms for pilot applications lately. For widening the borders of DLT in combination with big data, AI and machine learning approaches, by making a smart and safe agriculture sector.

4 Conclusion

The primary aim of this survey is to carry out a thorough literature review, regarding traceability and blockchain technology which is applied to the agriculture sector. Traceability is a technique that extensively studied these years. Many regulations, directives and laws throughout the world by considering the traceability issues of agri-food products that are set up. Also, blockchain technology is a matter of widespread research. Recently, some research study is carried out on the implanting the blockchain on agriculture traceability systems through their arrival. Simultaneously an enhanced trend of start-ups and pilot applications. The study, carried out through the entire context of this paper represented that blockchain technology seems appealing, but a few limits needless to be taken care of and addressed, like regulations, relationships amongst stakeholders, data ownership, scalability etc. For a good understanding of blockchain technology, the possibility to devise executions, researchers, and developers have many advantages from a widespread valuation system. Towards the conclusion of the implementation, the novel model is considered to be effective, this model leads to authorising these concerns. Minimize the cost, the risk is reduced, minimize the time taken, and enhance trust and transparency. Stakeholders who entail in adopting an innovative technique of functioning, once these are considered that our proposed method is welcoming, enhances efficiency and adds worth. Considering the overhead scenario, we can consolidate the new blockchain technologies through the primitive phase of agriculture as a widespread task that is analysed step-by-step, we can effectively engage the stakeholders who are directly affected by the supply chain.

References

1. Feng H, Wang X, Duan Y, Zhang J, Zhang X (2020) Applying blockchain technology to improve Agri-food traceability: a review of development methods, benefits and challenges. J Cleaner Prod 260. Art. no. 121031
2. Kshetri N (2018) 1 blockchain's roles in meeting key supply chain management objectives. Int J Inf Manage 39:8089
3. Thakur M, Donnelly KA-M (2010) Modeling traceability information in soybean value chains. J Food Eng 99(1):98105
4. Badia-Melis R, Mishra P, Ruiz-García L (2015) Food traceability: New trends and recent advances. A review. Food Control 57:393401
5. Chan KY, Abdullah J, Shahid A (2019) A framework for traceable and transparent supply chain management for Agri-food sector in Malaysia using blockchain technology. Int J Adv Comput Sci Appl 10(11):149–156
6. Caro MP, Ali MS, Vecchio M, Giaffreda R (2018) Blockchain-based traceability in Agri-food supply chain management: a practical implementation. In: Proceedings of IoT vertical topical summit agriculture Tuscany (IOT Tuscany), May 2018, pp 1–4
7. Watanabe H, Fujimura S, Nakadaira A, Miyazaki Y, Akutsu A, Kishigami JJ (2015) Blockchain contract: a complete consensus using blockchain. In: Proceedings of IEEE 4th global conference consumer electronics (GCCE), Oct 2015, pp 577–578

8. Queiroz MM, Telles R, Bonilla SH (2019) Blockchain and supply chain management integration: a systematic review of the literature. Supply Chain Manage Int J 25(2):241–254
9. Liang G, Weller SR, Luo F, Zhao J, Dong ZY (2019) Distributed blockchain-based data protection framework for modern power systems against cyber attacks. IEEE Trans Smart Grid 10(3):3162–3173
10. Leng K, Bi Y, Jing L, Fu H-C, Van Nieuwenhuyse I (2018) Research on agricultural supply chain system with double chain architecture based on blockchain technology. Future Gener Comput Syst 86:641–649
11. Hasan H, AlHadhrami E, AlDhaheri A, Salah K, Jayaraman R (2019) Smart contract-based approach for efficient shipment management. Comput Ind Eng 136:149–159
12. Behnke K, Janssen MFWHA (2020) Boundary conditions for traceability in food supply chains using blockchain technology. Int J Inf Manage 52. Art. no. 101969
13. Li Z, Wang WM, Liu G, Liu L, He J, Huang GQ (2018) Toward open manufacturing: a cross-enterprises knowledge and services exchange framework based on blockchain and edge computing. Ind Manage Data Syst 118(1):303–320
14. Ferrer EC (2018) The blockchain: a new framework for robotic swarm systems. In: Proceedings of future technologies conference Cham, Switzerland: Springer, 2018, pp 1037–1058
15. Zhao G, Liu S, Lopez C, Lu H, Elgueta S, Chen H, Boshkoska BM (2019) Blockchain technology in Agri-food value chain management: A synthesis of applications, challenges and future research directions. Comput Ind 109:83–99
16. Demestichas K, Peppes N, Alexakis T, Adamopoulou E (2020) Blockchain in agriculture traceability systems: a review. Appl Sci 10(12):1–22
17. Kosba A, Miller A, Shi E, Wen Z, Papamanthou C (2016) Hawk: the blockchain model of cryptography and privacy-preserving smart contracts. In: Proceeding of IEEE symposium security privacy (SP), May 2016, pp 839–858
18. Lu Q, Xu X (2017) Adaptable blockchain-based systems: a case study for product traceability. IEEE Softw 34(6):21–27
19. Duan J, Zhang C, Gong Y, Brown S, Li Z (2020) A content-analysis based literature review in blockchain adoption within food supply chain. Int J Environ Res Public Health 17(5):1784
20. Lee HL, Mendelson H, Rammohan S, Srivastava A (2017) Technology in agribusiness: opportunities to drive value. Stanford Graduate School Bus., CA, USA, White Paper, Aug. 2017. [Online]. Available: https://www.gsb.stanford.edu/faculty-research/publications/technologyagribusiness-opportunities-drive-value
21. Loop P (2016) Blockchain: the next evolution of supply chains. Mater Handling Logistics 71(10):22–24
22. New S (2010) The transparent supply chain. Harvard Bus Rev 88(10):76–82
23. Goldratt EM, Cox J (2016) The goal: a process of ongoing improvement. Routledge, Abingdon, UK
24. Abolghasemi M, Beh E, Tarr G, Gerlach R (2020) Demand forecasting in supply chain: the impact of demand volatility in the presence of promotion. Comput Ind Eng 142:106380. [CrossRef]
25. Zhou H, Benton W Jr (2007) Supply chain practice and information sharing. J Oper Manage 25:1348–1365. [CrossRef]

Face-Anti-spoofing Based on Liveness Detection

Shivani Mangal and Khushboo Agarwal

Abstract Many applications, like crossing points, banking, and mobile banking, are now using Face Recognition (FR) systems. The widespread usage of FR systems has heightened concerns about the security of face biometrics against spoofing assaults, in which a picture or video of a valid user's face is employed to attain unauthorized access to resources or activities. Even though numerous FAS or liveness detection techniques (which identify if a face is live or spoofed at the moment of acquisition) have been developed, the problem remains unsolved because of the complexity of identifying discriminatory and operationally affordable spoof characteristics and approaches. Furthermore, particular facial sections are frequently repetitive or correspond to image clutter, resulting in poor overall performance. This paper proposed a neural network model for face-anti-spoofing which outperforms the other models and shows an accuracy of 0.91%.

Keywords Face Anti Spoofing · Face recognition · Convolutional neural network

S. Mangal (✉) · K. Agarwal
Computer Science and Engineering, Madhav Institute of Technology and Science, Gwalior, India
e-mail: 122shivanimangal@gmail.com

K. Agarwal
e-mail: ka.agarwals@mitsgwalior.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_19

1 Introduction

Biometrics technology has gained in popularity as a result of the quick expansion of Internet technologies, and it is now extensively used in intelligence protection, criminal proceedings, financial and social stability, clinical training, and other disciplines. The face identification system is more simply accepted by the public than extant biometric identification systems owing to its excellent security, genuineness, and non-contact, and has formed an important research path for academics and industries [1]. The face recognition (FR) technology, on the other hand, is open to malware activity by unauthorized users, posing a serious threat to the system's integrity. As a result, creating a facial anti-spoofing system with higher identification performance, quick response time, and high robustness is critical [2].

The method of determining whether the recently collected facial picture is from a living human or a deceiving face is known as face anti-spoofing (FAS) detection. FAS research has been particularly engaged in recent times both domestically and overseas, owing to its significant academic significance. Printing, video replay and 3D mask attacks are the most popular spoofing assaults. Real and misleading faces have some variations, which are mostly expressed in image texture data, movement details, and perspective details [3]. We can create several FAS systems to identify the actual and counterfeit faces by taking benefit of these distinctions. FAS identification research has progressed fast in recent years, yielding numerous useful research outcomes. This study will examine the methodology based on deep learning (DL), as well as the technique's merits and weaknesses, as well as the FAS development trend.

With DL's continued advancement and remarkable performances in the field of FR, an increasing number of investigators have used FAS to investigate more comprehensive techniques for combating face deception. DL, as opposed to the old manual feature extraction (FE) technique, may autonomously learn photos, retrieve more critical and plentiful facial features, and assist in effectively distinguishing real from fake faces.

They first suggest a (CNN) [4] to extract features in FAS, which paved the way for a new branch of DL in the field of FAS [5]. The recognition impact was significantly lower than that of conventional approaches because the technologies were not yet established. Furthermore, the superiority of DL in feature extraction prompted a significant amount of research to pursue DL-based FAS. FAS based on DL has progressively advanced through network updates, TL [6], a combination of various characteristics, and domain generality, and has now exceeded the previous technique due to the unwavering dedication and repetitive tries of several researchers [7].

2 Related Work

Despite significant developments in facial recognition systems, face spoofing remains a significant risk. Most academic and corporate FR systems can be fooled by the following: an image, a video, a 3D face model of a genuine user; a reverse-engineered face image from the template of a genuine user; a sketch of a genuine user, etc. We present a quick summary of published facial impersonation recognition techniques. CNN has proven superior to alternative learning frameworks in a variety of computer vision tasks. For facial pictures, a distinctive feature representation approach known as HGC-CNN is employed to identify face spoof attacks with color photos. It's a multi-feature learning system that combines capsule NN and hypergraph regularisation concepts. Capsule NN can incorporate a variety of characteristics, including intensity values, LBP, and picture quality. Hypergraph regularisation can also be employed to learn relationships between samples. The expressive ability of extracted features is improved even more when locality information is included. SVM was utilized in the studies since the new representation is consistent with existing classifiers. The suggested approach outperformed the prior approach on FSA detection with color photos, according to experimental data on the NUAA database and the Multispectral spoofing database [4]. An another approach that combines two CNN streams presented by Yousef Atoum et al. They utilize both the whole-facial image and regions taken from a similar face to differentiate the spoof from live faces, as with most previous methods in face anti-spoofing that only use the entire face to identify presenting attacks. The first CNN streaming is based on the characteristics of patches collected from different face areas. This stream proves to be resistant to all types of presentation attacks, particularly on lower-resolution face photos. The second CNN stream uses the whole facial image to estimate face depth. The outcomes of this CNN's trials suggest that our depth estimation, especially on higher-resolution images, can produce impressive outcomes [8]. Gene LBPnet, a novel technique for CNN based on LBP for face spoofing detection, is presented by Karuna Grover & Rajesh Mehra. On the NUAA dataset, this methodology outperformed previous state-of-the-art algorithms. Using various assessment parameters, it has been demonstrated that the suggested approach provides excellent accuracy (98%) and a low Equal Error Rate, leading to improved recognition of spoofing attacks and thereby improving system security spoofing attempts [9]. To mutually assess the complexity of face pictures and the rPPG signal of face footage, the suggested system integrates CNN and RNN structures. To discriminate between real and fake faces, the approximated depth and rPPG are combined. They also provide a new FAS database for faces that includes a wide range of lighting, subject, and pose variants. The SiW dataset, which covers more subjects and modifications than previous datasets, is introduced. Lastly, they illustrate the technique's advantage in the experiment [10]. For face liveness identification, Zahid Akhtar et al. propose seven unique strategies for obtaining exclusionary patches in a facial image. A particular classifier is given the properties of specified discriminative picture patches. For the ultimate categorization of authentic and spoof faces, the categorization outcomes of these regions are pooled using a majority-voting-based scheme. In comparison to prior efforts, experiment outcomes on two publically accessible datasets reveal comparable outcomes [11]. To improve the security level of a FAS system, they introduced a novel model for identifying liveness attack images in this article. The variation between the attributes of actual and false faces is taken into account in the approach. As a result, integrating types of image information improves attack effectiveness greatly when compared to using a single approach [12].

3 Proposed Methodology

- Step 1: Collect the CASIA v2 image dataset which is freely available.
- Step 2: Cleaning the data and removing the noisy data.
- Step 3: Identifying and removing noisy images and perform data shuffling.
- Step 4: Reshaping the data features, and samples and splitting them into training and testing.
- Step 5: Passing the data into the training model.
- Step 6: Train and test samples (3331, 833) for fake and real images and split into 70% for training and 30% for testing.
- Step 7: After completion of training measure performance parameters accuracy, recall and precision (Fig. 1).

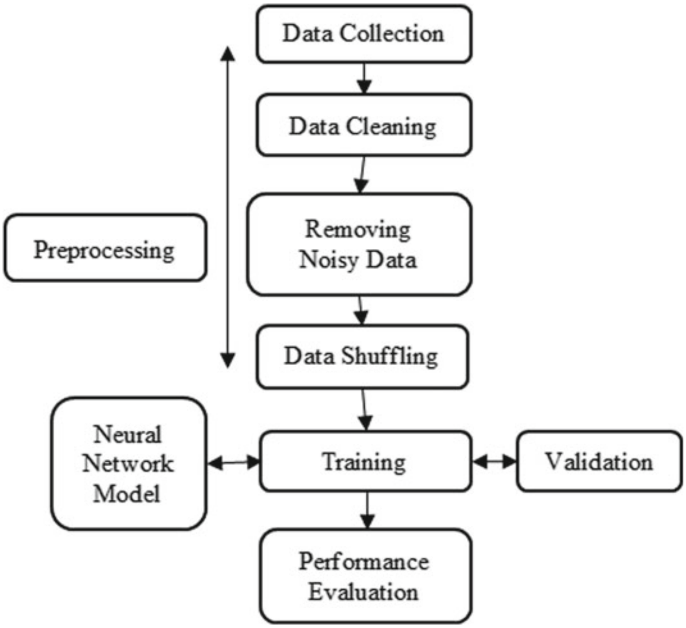


Fig. 1 Flow chart of proposed methodology

3.1 Dataset Gathering

The suggested method is evaluated using the CASIA v2 picture dataset, which is frequently used to identify image forgery and is freely available. There are 4795 photos in all, with 1701 legitimate and 3274 fake.

3.2 Data Pre-processing

The goal of pre-processing is to optimize graphic data by overwhelming unwanted deformities or improving particular graphic properties that are important for subsequent processing and evaluation.

(a) Data cleaning is the act of determining and restoring (or eliminating) corrupted or erroneous information from a record set, table, or database. It includes recognizing insufficient, improper, faulty, or redundant data and then updating, changing, or deleting the dirty or imprecise data.

(b) Checking Noisy Images: Image noise is a sort of ambient sound that produces erratic changes in image intensity or color details. The image detector and circuits of a scanner or digital camera can make it. Movie coarse and the inevitable impulse noise of an optimal photoelectron can likewise cause image noise. We can check the original and noisy images in the dataset and convert all the images to error analysis for better performance.

(c) Data Shuffling: The shuffling strategies try to jumble up data while retaining logical linkages among columns if desired. It rearranges data from data inside a feature (for example, a column in pure flat format) or a collection of attributes randomly (e.g. a set of columns). Figure 2 shows the original and ELA image.

3.3 Model Parameter

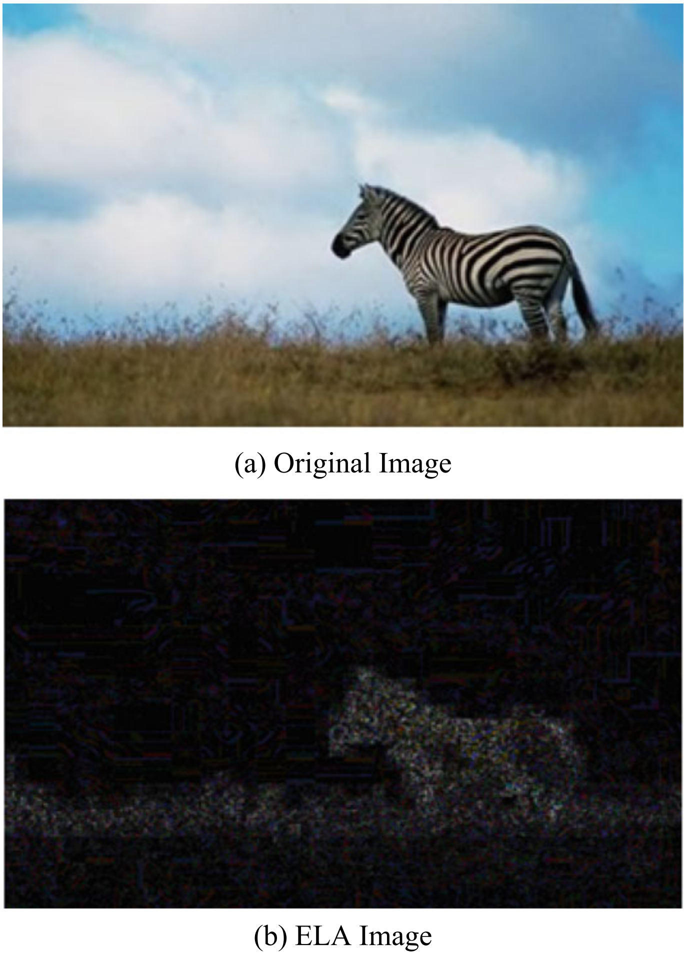
Figure 3 shows the model parameter and explain is below:

(1) *Conv2D:* Conv2D is a 2-D convolution layer that produces a sequence of results by twisting a convolution kernel with the layers' data [13].

(2) *Max-Pooling:* Pooling that chooses the largest component from the section of the feature map encompassed by the filters is known as max pooling. As a consequence, the result of the max-pooling layer would be an FM with the most important characteristics of the previous FM [14].

(3) *Dropout Layer:* Dropout is a strategy for avoiding overfitting in a model. At every iteration of the training stage, Dropout consists of setting the outbound edges of hidden nodes (Hidden components are made up of neurons) to 0 [15].

(4) *Flatten Layer:* The process of converting data into a 1D array for usage in the following layer is known as flattening. The CL result is flattened to produce a



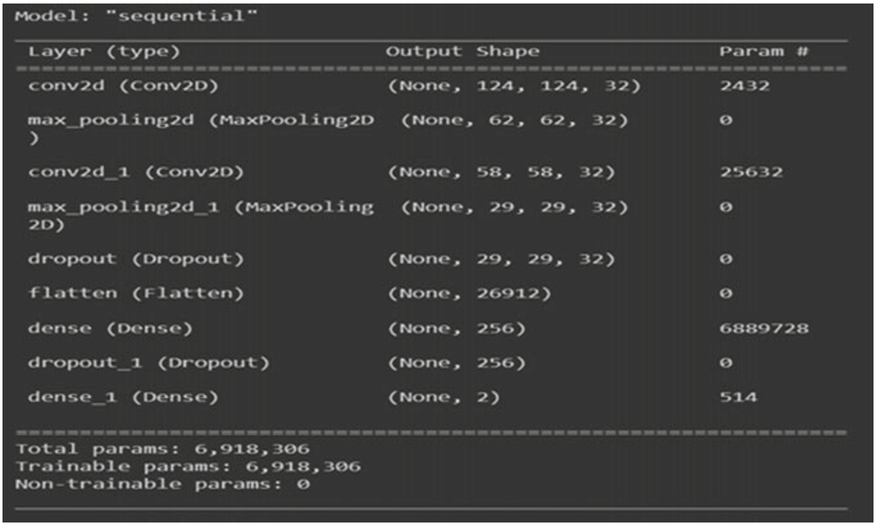
(a) Original Image

(b) ELA Image

Fig. 2 Figure showing the original and ELA image

single long feature representation. It's also related to a fully-connected layer, which is the definitive classification technique [16].

(5) *Dense Layer:* A DL in any NN is tightly linked to the layer before it, indicating that each of the layer's neurons is linked to each of the layer's neurons. It is the most commonly used layer in ANN. The outcome of the DL is an ‘m’ dimensional array. As a consequence, the layer is typically used to change the dimensionality of the vector. The vector is also subjected to processes such as rotation, scale, and translation by these layers [17].



```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 124, 124, 32)      2432

 max_pooling2d (MaxPooling2D  (None, 62, 62, 32)       0
 )

 conv2d_1 (Conv2D)           (None, 58, 58, 32)        25632

 max_pooling2d_1 (MaxPooling  (None, 29, 29, 32)       0
 2D)

 dropout (Dropout)           (None, 29, 29, 32)        0

 flatten (Flatten)           (None, 26912)             0

 dense (Dense)               (None, 256)               6889728

 dropout_1 (Dropout)         (None, 256)               0

 dense_1 (Dense)             (None, 2)                 514

=================================================================
Total params: 6,918,306
Trainable params: 6,918,306
Non-trainable params: 0
_________________________________________________________________
```

Fig. 3 Model parameter

The neural network model is sequentially trained. The employed NN model with layers is shown in Fig. 3. The dataset is separated into the training of 70% and testing of 30% images for fake (3331) and Real (833) images. The NN is trained and used the RELU and Sigmoid as the activation function. The first layer of the network is the conv2D layer with 2432 parameters, after that the max-pooling2D layer is employed proceeding again to conv2D and max-pooling layer. The dropout, flatten and dense layers were then employed in a cascade manner. Table 1 shows the hyper parameters of training where the ADAM optimizer is used with 20 epochs for a batch size of 32.

Table 1 Hyper parameters of training

Optimizer	ADAM
Loss function	Binary cross-entropy
Metrics	Accuracy
Epochs	20
Batch size	32
Validation split	0.2
Shuffle	True

4 Simulation Result

4.1 Performance Matrix

Precision and accuracy: The degree to which a measured value is near its true value is known as accuracy. Precision refers to how closely all of the measured values are related. To put it another way, accurateness is the proportion of right categories to total classifications.

Recall/Sensitivity: Sensitivity is defined as the proportion of true positives to the whole number of actual positives. Similarly, specificity, also known as the true negative rate, is the proportion of genuine negatives to total negatives [18].

F1-Score: When a model's accuracy is greater than 90%, it is considered to be accurate, we also include the F1 score as a statistic that provides a better indication of cases that have been wrongly classified. The harmonic mean of precision and recall is employed to compute this. When TP and TN are more significant, accuracy is utilized. When the class distribution is unequal and FP and FN are more important, the F1 score is a better statistic [19]. All of the metrics formulas are as shown below.

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

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

$$recall = \frac{TP}{TP + FN} \tag{3}$$

$$F\text{-}score = \frac{2}{1/precision + 1/recall} \tag{4}$$

$$specificity = \frac{TN}{TN + FP} \tag{5}$$

$$sensitivity = \frac{TP}{TP + FN} = recall \tag{6}$$

4.2 Confusion Matrix

In a classification issue, a Confusion Matrix is a tabular representation of prediction outcomes with count values split down by class. It demonstrates how a classification model performs while making predictions, as the name implies. It reveals the types of errors made by the classifier as well as the errors themselves [20]. Better and worse classification results are represented by the points above and below the line, accordingly. The matrix is shown in Fig. 4.

Figure 5 shows the accuracy and loss graphs for the evaluated results. Table 2 shows the comparison of the base and proposed results with the proposed system accuracy of 0.91.

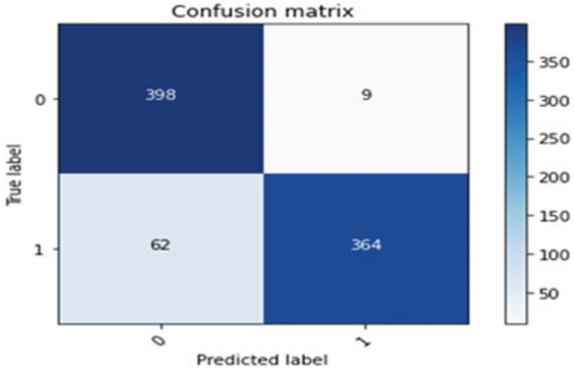


Fig. 4 Confusion matrix showing the true and the predicted label

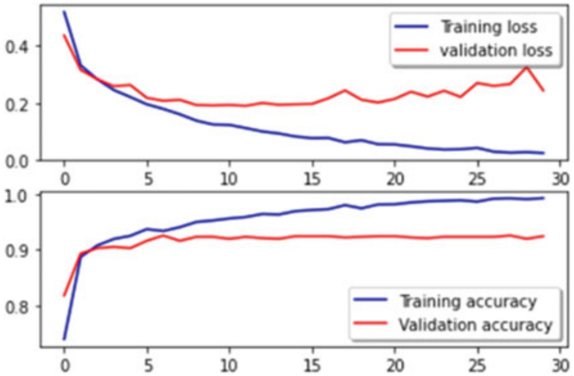
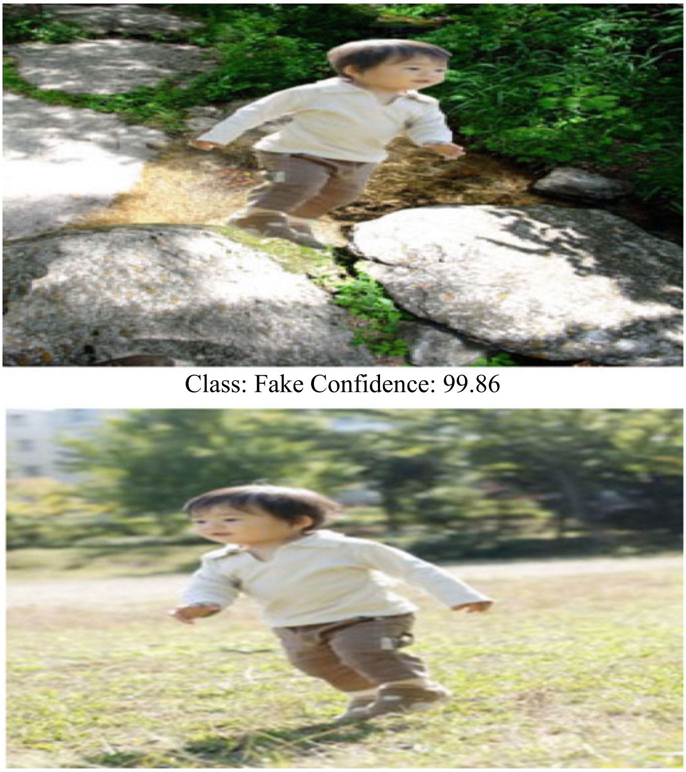


Fig. 5 Figure showing the loss and accuracy graph

Table 2 Comparison of the base and the proposed results

Results	Base Paper	Proposed
Recall	0.81	0.85
Precision	0.86	0.97
F1 Score	0.81	0.91
Accuracy	0.85	0.91



Class: Fake Confidence: 99.86

Class: Real Confidence: 99.56

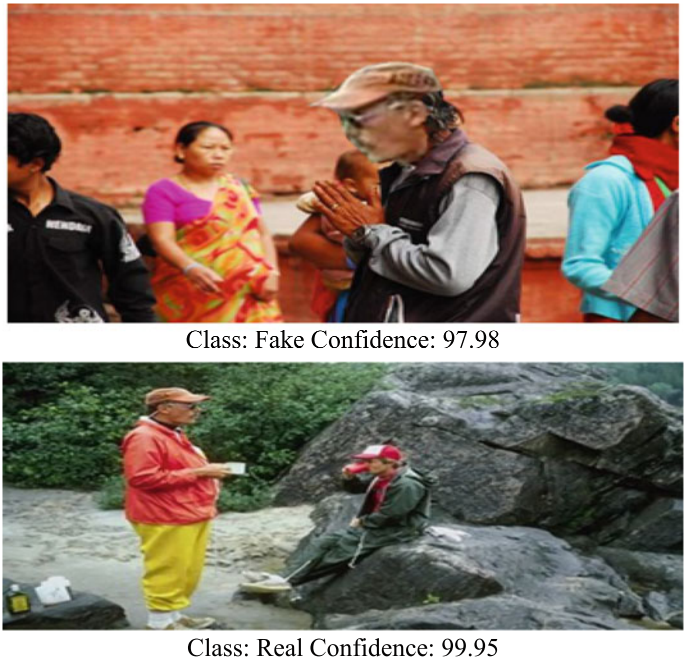
Fig. 6 An example of a resultant image compared with the original image

The precision, recall, and f1-score of the proposed model are 0.97, 0.85, and 0.91. The fake and real confidence of the resultant images is shown below.

Figures 6, 7 and 8 shows fake and real confidence images for spoofing techniques with duplicate photographs of people whose original images areas maintained in a database. It means that if an intruder wanted access to the authorized system, he or she may have used these several techniques.

5 Conclusion

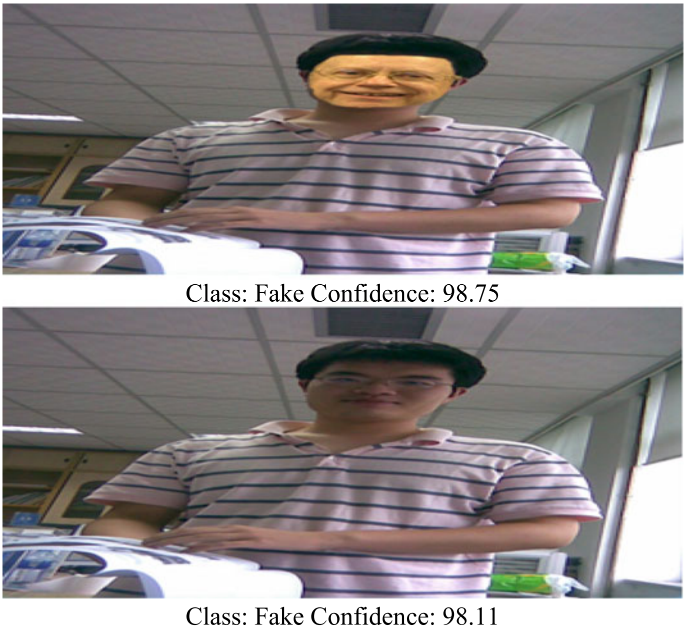
Face Recognition has become an essential technique for achieving protection as AI has become more widely used in real life. FAS has become a pressing issue in the fight against harmful attacks. The research of face spoofing identification has been continuously monitored and revised, from the starting of manual FE methods based on image texture, image quality, and depth information, to using DL to instantly extract



Class: Fake Confidence: 97.98

Class: Real Confidence: 99.95

Fig. 7 An example of a resultant image compared with the original image



Class: Fake Confidence: 98.75

Class: Fake Confidence: 98.11

Fig. 8 An example of a resultant image compared with the original image

features, merged with network up-gradation, feature assimilation, and domain generalization, and the efficiency and effectiveness of identification have now attained a significant state.

References

1. Hashemifard S, Akbari M (2021) A compact deep learning model for face spoofing detection. [Online]. Available: http://arxiv.org/abs/2101.04756
2. Al-Huda Taha N, Hassan TM, Younis MA (2021) Face spoofing detection using deep CNN. Turkish J Comput Math Educ 12(13):4363–4373
3. Ming Z, Visani M, Luqman MM, Burie JC (2020) A survey on anti-spoofing methods for facial recognition with RGB cameras of generic consumer devices. J Imaging (6)12. https://doi.org/10.3390/jimaging6120139
4. Liang Y, Hong C, Zhuang W (2021) Face spoof attack detection with hypergraph capsule convolutional neural networks. Int J Comput Intell Syst 14(1):1396–1402. https://doi.org/10.2991/IJCIS.D.210419.003
5. De Souza GB, Papa JP, Marana AN (2021) Efficient deep learning architectures for face presentation attack detection. 112–118. https://doi.org/10.5753/sibgrapi.est.2020.12992
6. Khalid IA (2020) Transfer learning for ımage classification using tensorflow. Towards Data Sci
7. Zhang M, Zeng K, Wang J (2020) A survey on face anti-spoofing algorithms. J Inf Hiding Priv Prot 2(1):21–34. https://doi.org/10.32604/jihpp.2020.010467
8. Liu Y, Jourabloo A, Liu X (2018) Learning deep models for face anti-spoofing: binary or auxiliary supervision. In: Proceedings of the IEEE computer society conference computter vision pattern recognition, pp 389–398. https://doi.org/10.1109/CVPR.2018.00048
9. Das PK, Hu B, Liu C, Cui K, Ranjan P, Xiong G (2019) A new approach for face anti-spoofing using handcrafted and deep network features. In: Proceeding—IEEE ınternational conferences servey operations and logistics and ınformatics 2019, SOLI 2019, no November, pp 33–38. https://doi.org/10.1109/SOLI48380.2019.8955089
10. Akhtar Z, Foresti GL (2016) Face spoof attack recognition using discriminative ımage patches. J Electr Comput Eng. https://doi.org/10.1155/2016/4721849
11. Atoum Y, Liu Y, Jourabloo A, Liu X (2018) Face anti-spoofing using patch and depth-based CNNs. In: IEEE ınternational joint conference on biometrics (IJCB) 2017, vol 2018-Janua, pp 319–328. https://doi.org/10.1109/BTAS.2017.8272713
12. Grover K, Mehra DR (2019) Face spoofing detection using enhanced local binary pattern. Int J Eng Adv Technol 9(2):3365–3371. https://doi.org/10.35940/ijeat.b3834.129219
13. Pouyanfar S et al (2019) A survey on deep learning. ACM Comput Surv 51(5):1–36. https://doi.org/10.1145/3234150
14. Masita KL, Hasan AN, Shongwe T (2020) Deep learning in object detection: a review. In: 2020 International conference artificial ıntelligence big data, computing data communication system (icABCD 2020)—proceeding no. August, 2020. https://doi.org/10.1109/icABCD49160.2020.9183866
15. Manalu BU, Tulus, Efendi S (2020) Deep learning performance in sentiment analysis. In: 2020 4th International conference on electrical telecommunication and computer engineering (ELTICOM 2020)—proceeding, pp 97–102. https://doi.org/10.1109/ELTICOM50775.2020.9230488
16. Shirahatti AP, A survey of deep learning for sentiment analysis. V(I):1–7
17. Weng W, Zhu X (2021) INet: convolutional networks for biomedical image segmentation. IEEE Access 9:16591–16603. https://doi.org/10.1109/ACCESS.2021.3053408
18. ML (2020) Classification: precision and recall. Machine learning crash course. https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall

19. R (2020) ROC Curves. Machine learning crash course. https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc
20. Narkhede S (2018) Understanding confusion matrix. Towardsdatascience. https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

PDR Analysis and Network Optimization of Routing Protocols for Edge Networks

Archana Ratnaparkhi, Radhika Purandare, Gauri Ghule, Shraddha Habbu, Arti Bang, and Pallavi Deshpande

Abstract AdHoc On-Demand Distance Vector (AODV) is a notable and broadly utilized protocol for MANETs. The Mobile AdHoc Network or MANET, without any infrastructure, is an assortment of remote nodes conveying and communicating over a wireless network. All wireless devices working in AdHoc mode inside range are allowed to have communication with each other in absence of base station. The routers are capable to roam and communicate arbitrarily and organize themselves as per the requirements when the nodes structure themselves into a random topology. Radio signals possess range limitations due to which multihop communication in MANETs is inevitable. The performance of traffic situations utilized in a mobile AdHoc network is responsible for the transmission and gathering of data between source and destination in a MANET.This paper provides a comprehensive comparative analysis of the routing protocols with respect to variation in node configuration.

Keywords Adhoc networks · Routing protocols · Packet delivery ratio · Throughput · Dynamic environment

1 Introduction

Design of effective routing strategy has been a matter of research for decades. Evolution and advancements in routing strategy using optimally minimal resources is the crux. The designing of a reliable and efficient routing strategy is a very challenging problem due to the limited resources in Mobile Ad-hoc Networks (MANETs). The variables in network conditions such as network partitioning, network size and traffic density dictate the routing strategy to be used in order to make efficient use of the limited resources. The routing protocol is also expected to cater to different levels of QoS required by a diverse range of users and applications. With the advent

A. Ratnaparkhi (✉) · R. Purandare · G. Ghule · S. Habbu · A. Bang · P. Deshpande
Vishwakarma Institute of Information Technology, Pune, India
e-mail: archana.ratnaparakhi@viit.ac.in
URL: http://www.viit.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_20

of MANETs, the significant algorithms such as link state algorithm and distance vector have undergone numerous enhancements. In huge Mobile Adhoc Network, traditional link-state and distance vector algorithms do not scale, as periodic updates in routes consume a significant amount of the available bandwidth, consume more power on the side of each node and increase channel contention. To improve on these, numerous routing protocols have been proposed. These protocols can be classified as proactive, reactive and hybrid protocols [1]. Basic functional difference between proactive and On-demand routing protocols lies in the way the routing information is maintained. Routing information is maintained by every node to every other node in the network in case of proactive protocol. On-demand routing protocols maintain information only for the active nodes, thus reducing the overheads in proactive routing protocols. Improving on these two kinds of routing protocols, hybrid routing protocols were proposed. These protocols allow the nodes that lie in close proximity to work in unison and form a backbone to cut down on route discovery overheads, thus increasing stability. These protocols work by maintaining the routes proactively to the nodes nearby and determining routes to distant nodes through an appropriate strategy for route discovery.

1.1 Dynamic Source Routing (DSR)

As DSR protocol involve carrying the complete packet from source to destination, it makes the protocol ineffective in large networks, as an increase in diameter of the network will increase the packet overhead. Hence, the overhead may occupy most of the bandwidth in large scale networks that are dynamic in nature. This routing protocol is advantageous over protocols such as LMR, TORA and AODV, including better performance in moderately sized networks.

1.2 Ad-hoc On-demand Distance Vector (AODV)

This protocol is advancement over sequence numbering and periodic beaconing procedure of DSDV and employs a route discovery procedure like DSR. There are a couple of major differences between AODV and DSR. In DSR, each packet contains complete routing information, where the packets carry only the destination address in AODV. This implies that AODV has fewer routing overheads than DSR. Also, the address of each node along the route is carried by route replies in DSR. AODV trumps over DSR in its adaptability in highly dynamic networks. Although, large delays may possibly be experienced in route construction, and failure to establish a link may initiate an alternative route discovery, consuming more bandwidth and introducing excess delays with an increase in the size of the network [2, 3].

2 Related Work

In this work, analysis of six protocols such as ZigBee, UWB, Bluetooth, WiFi, WiMax and GSM is carried out. The best and most suitable protocol for an application of intelligent sensor is selected by evaluating it on the basis of the transmission time, the data coding efficiency, the bite error rate, and the power and the energy consumption with network size [4]. The cross layer protocols such as Low Energy Adaptive Clustering, Self Organized TDMA Protocol, Flexible TDMA Protocol, Energy Efficient Fast Forwarding Protocol and D-MAC are analyzed by the authors depending on their performance with the help of NS2 [5]. A survey is done on requirements, challenges in technicalities and already available work on MAC layer protocols related to the support of M2M communication. The issues related to efficiency, scalability and accessibility of fair channel for M2M communication are mentioned in detail. The existing MAC layer protocols and their applicability to M2M communication is also reviewed with the protocols that already developed only for M2M communication. The paper also talks about future research and open problems in it and discussed about in-process standardization efforts [6]. A MAC access algorithm for IEEE 802.15.4 LRWPAN is proposed and implemented by authors based on CSMA/CA protocol, backoff delay, the probability of collision, and retransmission in total networks are the things algorithm focused on reducing. The idea is depended on BE and CW parameter's dynamically modifying value according to traffic state. The unused slot time of superframe is also manipulated to increase the efficiency rate [7]. Examination of the transmitter of a vehicle hub is started which imparts in-high unique environment and undisturbed network change circumstances. The dissection of regular parcel achievement proportion with throughput of VANET in realistic traffic environment is done. Authors also shown through demonstration that transmitter is affordable for vehicle to vehicle communication. It also shows that in rush hour gridlock situation it shows better results [8]. In order to find the optimal network simulator different open source network simulators are compared depending on parameters like CPU utilization, memory usage, computational time, and scalability by simulating a MANET routing protocol [9–13].

3 Proposed Work

Experimentation has been done on NS2 platform. The flowchart as seen in Figs. 1 and 2 starts with initialization of environmental and network parameters, delay, queue and then timing information. In the following experiment we have taken both TCP and UDP AODV/DSR protocol and noted the change in Packet Delivery Ratio with respect to change in the distance between their destination nodes. Here we have taken 4 nodes i.e. source and destination nodes for TCP protocol and same as for UDP AODV/DSR protocol and changed the distance of either one or both destination

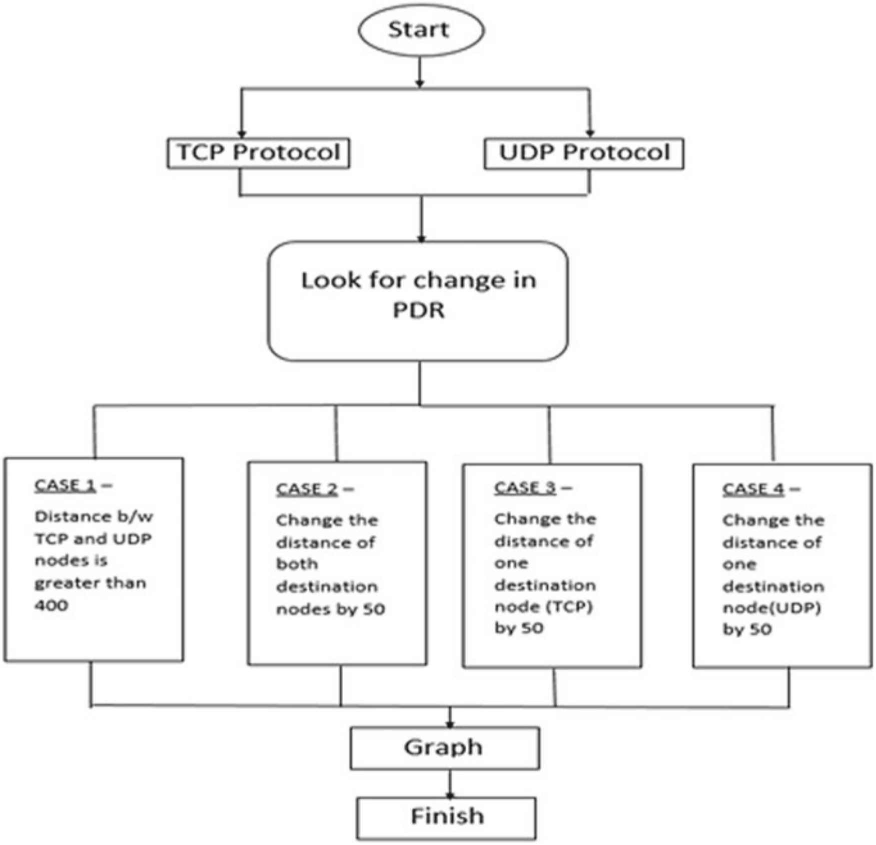

Start
TCP Protocol
UDP Protocol
Look for change in PDR
CASE 1 – Distance b/w TCP and UDP nodes is greater than 400
CASE 2 – Change the distance of both destination nodes by 50
CASE 3 – Change the distance of one destination node (TCP) by 50
CASE 4 – Change the distance of one destination node(UDP) by 50
Graph
Finish

Fig. 1 Flowchart

nodes. Simulation parameters as indicated in literature have been used and tabulated in Table 1.

Case 1: The below NAM window snapshot as seen in Fig. 3 shows that when the distance between UDP which is source one and TCP i.e. source two, nodes is greater than 400, we see that packet loss occurs. However when the distance between the destination node of UDP i.e. node 3 and source node of TCP i.e. node 0 is greater than 400, it results to no packet loss. This is discussed later in result analysis.

Case 2: The below snapshot as seen in Fig. 4 of NAM window shows the change in distance of both the destination nodes parallelly by the distance parameter of 50, keeping fixed source nodes. Here both UDP and TCP source nodes are fixed and the destination nodes i.e. node 1 and node 3 are moved parallel to each other. The different results at different points will be discussed later.

Case 3: The snapshot as seen in Fig. 5 of the NAM window shows the change in distance of only one destination node is changed by the distance parameter of

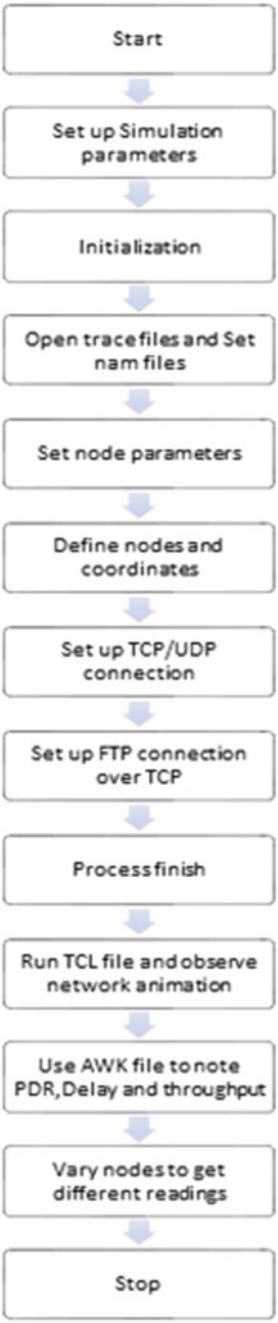

Start
Set up Simulation parameters
Initialization
Open trace files and Set nam files
Set node parameters
Define nodes and coordinates
Set up TCP/UDP connection
Set up FTP connection over TCP
Process finish
Run TCL file and observe network animation
Use AWK file to note PDR, Delay and throughput
Vary nodes to get different readings
Stop

Fig. 2 Programming steps

Table 1 Simulation parameters

Parameter	Value
Channel type	Wireless channel
MAC protocol	IEEE 802.15.4
Routing protocol	AODV, DSR
Traffic type	CBR/FTP
Simulation time (s)	100
Number of nodes	4
Number of sources	2
Range parameter (m)	50
Queue type	Drop tail
Radio propagation models	Two ray ground
Antenna model	Omni-directional antenna

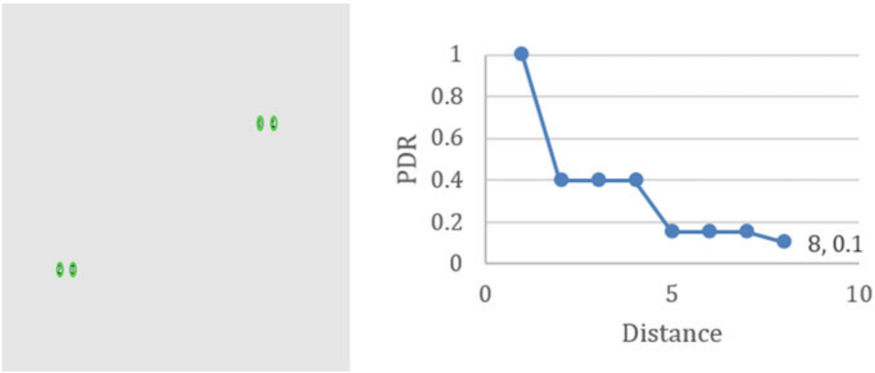


Fig. 3 Case 1: variation of PDR with distance (NAM window)

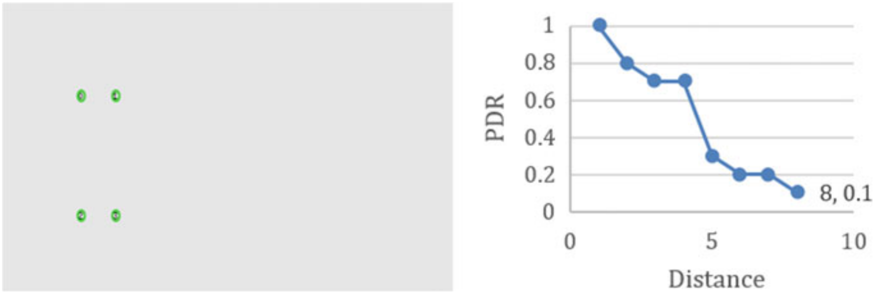


Fig. 4 Case 2: variation of PDR with distance (NAM window)

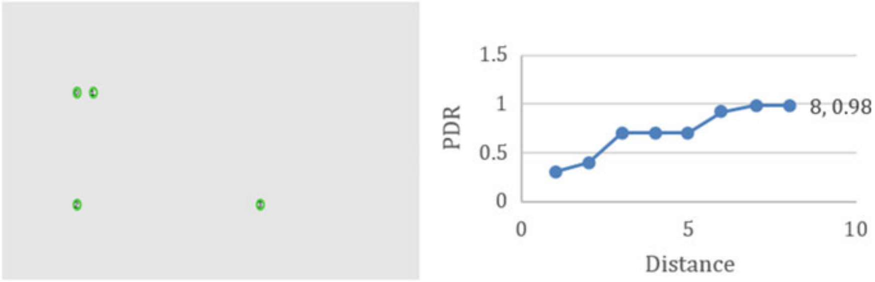


Fig. 5 Case 3: variation of PDR with distance (NAM window)

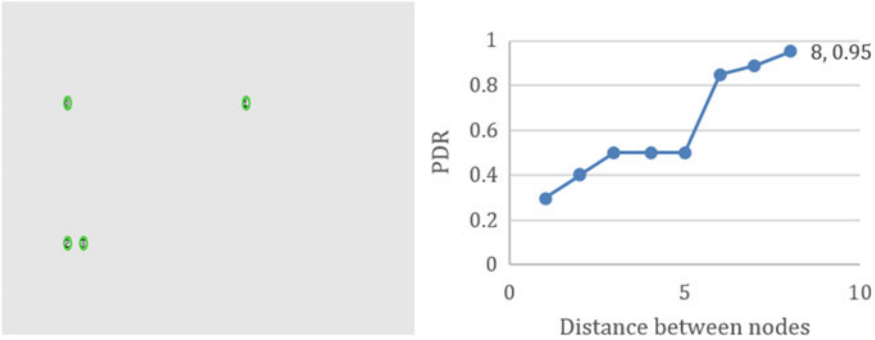


Fig. 6 Case 4: variation of PDR with distance (NAM window)

50, keeping all nodes at fixed position. Here we have change the position of TCP destination node keeping source of TCP and nodes of UDP fixed.

Case 4: The below snapshot of the NAM window shows the change in distance of only one destination node (i.e. UDP) by the distance parameter of 50, keeping all nodes at fixed position (Fig. 6).

4 Result Analysis

The following observations were noted during the experiment.

- When the distance between UDP and TCP nodes is greater than 400 the PDR is always high i.e. 1.
- In case 1, when the distance between source and destination nodes is greater than 300 the PDR decreases to 0 gradually.
- In case 2, when the distance between source and destination node of TCP is greater than 300, the UDP protocol will drops packets, as a result PDR decreases abruptly towards 0.

- In case 3, when the distance between source and destination node of UDP is increased the PDR increases continuously.
- In case 4 as the graph shows, DSR shows some fluctuations for some input values of TCP and UDP compared to AODV.

5 Conclusion

In this paper, the desired work is achieved using an NS2 simulator under varying parameters, traffic states and conditions to study and analyze the working and performance of AODV/DSR protocols by taking both TCP and UDP connections. We have observed, noted and calculated the change in Packet Delivery Ratio or PDR by varying the distance between the destination nodes. We have taken 4 nodes for our experiment, source and destination nodes for TCP protocol as well as UDP AODV/DSR protocol. By changing the distance of either one or both nodes, we came across multiple cases. We concluded that the PDR is always high or 1 when the distance is greater than 400. The PDR decreases to zero when the distance is greater than 300. The PDR decreases continuously when the distance between the nodes of UDP is increased. Compared to AODV, DSR shows fluctuations or irregularities for few input values of TCP and UDP. This work can be extended further to include more dynamic movement in nodes and the parameters can be studied. Furthermore mobile adhoc networks which are far more complex can be studied as part of further enhancements.

Acknowledgements We are extremely thankful to Department of Electronics and Telecommunication and Research and Development wing for their constant support and encouragement.

References

1. Singh B, Hans R (2015) TCP and UDP based performance analysis of AODV, DSR and DSDV routing protocols under different traffic conditions in mobile adhoc networks. Int J Future Gener Commun Netw 8(2):73–92
2. Singh R, Tripathi P (2021) Performance evaluation of TCP (Transmission Control Protocol) and UDP (User Datagram Protocol) over AODV routing protocols. Int J Modern Sci Eng Technol. Academia.edu. Retrieved November 2, 2022, from https://www.academia.edu/11945555/Performance_Evaluation_of_TCP_Transmission_Control_Protocol_and_UDP_User_Datagram_Protocol_over_AODV_Routing_Protocols
3. Abolhasan M, Wysocki T, Dutkiewicz E (2004) A review of routing protocols for mobile ad hoc networks. Ad Hoc Netw 2(1):1–22
4. Biradar RV, Patil VC, Sawant SR, Mudholkar RR (2009) Classification and comparison of routing protocols in wireless sensor networks. Spec Issue Ubiquitous Comput Secur Syst 4(2):704–711
5. Chakkor S, Cheikh EA, Baghouri M, Hajraoui A (2014) Comparative performance analysis of wireless communication protocols for intelligent sensors and their applications. arXiv:1409.6884

6. Gajjar S, Pradhan SN, Dasgupta K (2012) Performance analysis of cross layer protocols for wireless sensor networks. In: Proceedings of the international conference on advances in computing, communications and informatics, pp 348–354
7. Rajandekar A, Sikdar B (2015) A survey of MAC layer issues and protocols for machine-to-machine communications. IEEE Internet Things J 2(2):175–186
8. Bouazzi I, Bhar J, Atri M (2017) New CSMA/CA prioritisation based on fuzzy control mechanism. Int J Intell Eng Informat 5(3):253–266
9. Karunakar P, Matta J, Singh RP, Kumar OR (2020) Analysis of position based routing Vanet protocols using Ns2 simulator. Int J Innov Technol Exploring Eng (IJITEE) 9(5)
10. Khan AR, Bilal SM, Othman M (2012) A performance comparison of open source network simulators for wireless networks. In: 2012 IEEE international conference on control system, computing and engineering. IEEE, pp 34–38
11. Voitenko I (2022) https://github.com/iuriivoitenko/simpleMANET. GitHub. Retrieved June 13, 2022
12. Shakya S, Pulchowk LN (2020) A novel bi-velocity particle swarm optimization scheme for multicast routing problem. IRO J Sustain Wirel Syst 2:50–58
13. Sathesh A (2020) Artificial intelligence based edge computing framework for optimization of mobile communication. J IoT Soc Mob Analytics Cloud 2(3):160–165

Privacy Threat Reduction Using Modified Multi-line Code Generation Algorithm (MMLCGA) for Cancelable Biometric Technique (CBT)

Pramod D. Ganjewar, Sanjeev J. Wagh, and Aarti L. Gilbile

Abstract Nowadays individual's identity verification is required at many places for authenticity like Government Sector, Private sector, Public Sector etc. Many existing systems are based on either physical documents or on biometric parameters. Biometric system has become a more convenient way for authenticity check. In biometric system, some samples of biometric parameters are collected and stored at the server side for further use. These samples would be used for the verification of the identity of the person. The actual required biometric parameter will be compared with all the existing samples available in the system to match with the registered person. If it matches with one of the existing sample, then that will be authenticated by the system and allowed to perform further operations. But what if, the samples collected by the authority get misused? That needs security from the owner of the system. So, the identity of the samples must be hidden from the operators by some way, which has been focused in this work. A new approach called as Cancelable Biometric Technique (CBT) using Modified Multi-Line Code Generation Algorithm (MMLCGA) is used for storing biometric samples using template. The cancelable method converts the gathered samples and stores it into the system for hiding its original identity. In the verification phase, the system will convert it back to the original sample to be used for identity matching of the person or user. This technique provides more privacy, because of which privacy threats can be reduced. The time and accuracy of the proposed technique is better by 15% and 1.4% respectively, when compared to the existing technique Multi-Line Code Generation Algorithm (MLCGA).

Keywords Cancelable biometrics · Revocable · Biometric sensor · Multi-biometrics · Biometric security · Template protection

P. D. Ganjewar (✉) · A. L. Gilbile
MIT Academy of Engineering, Alandi (D.), Pune, Maharashtra, India
e-mail: pdganjewar@mitaoe.ac.in

S. J. Wagh
Government College of Engineering, Karad, Maharashtra, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_21

1 Introduction

The regular authentication system uses tokens like PIN, passwords, cards, IDs etc. These tokens can be misused by anybody, without the presence of real entity, which is the disadvantage of the regular authentication system. The authentication system that uses biometrics is better than the existing one. These systems uses biometric features like fingerprint, thumb impression, IRIS structure, voice, keystrokes, signatures etc., as every user has unique biometric features. There is no need to memorize these features for further use. There is no chance of misplacing or forgetting these features like passwords or other identification mechanisms. Also, it cannot be used without the presence of the real entity. But there is an issue with the storage of these biometric templates. They are generally stored as it is, at any place as per the company guidelines. If the database of these biometric templates gets stolen, then they can be misused, as these are stored in its original form. These templates need to be protected in such a manner that, even after getting the access to such templates, it cannot be misused. The templates need protection in such a way that even after recovery of these templates, the original templates cannot be obtained. In template protection, biometric features need to be arranged in some well- organized way, so that original biometric information is more secure. In case of misuse of the transformed biometric template, new template will be formed by changing the transformation parameters of the biometric template and the new template gets restored on the misused template. The primary goals of this work are as follows:

- **Non-reversibility**: After performing many operations, original templates should not be regenerated from the protected template.
- **Accuracy**: The operations performed for the template protection, should not hamper on the quality of the original template.
- **Diversity**: Same biometric cannot be used by different applications. It will increase the chances of revocability.

2 Related Work

Kaur and Khanna [1] presented the concerns regarding the security of the biometric systems. The pseudo-biometric data is procured and kept at safe place and used for verification of the identity of the users. The loss of the privacy can be reduced using pseudo-identity and it allows the revocability in cases of compromise. This method extracts pseudo biometric from its original biometrics and also reduces its size by half of its original size. They analyzed the recognition and protection performance of the proposed method based on the data taken from unimodal and multimodal data generated from biometric data of palm-vein, palm-print, face etc. A new transformed template will be generated based on the distance calculation between the feature points and some random point, using Euclidean distance and the performance also get improved.

Wu et al. [2] addressed template protection using cancelable biometric techniques, for electrocardiograms (ECG) biometrics to decrease the threats. The property of ECG, which makes them difficult to forge or steal is its inherent biological and intrinsic life sign. The feature of ECG to be considered is its R peak. The principle of "Signal Subspace collapse" was used to generate various templates associated with ECG to revoke compromised identity like passwords or PINs. To recognize the identity of the users from their ECGs, the popular multiple signal classification technique was used. In this, without understanding the transformation of the distortion, the identification will be carried out unlike other cancelable techniques. This is the reason that the recovery of original ECGs from their template is very difficult. They have analyzed it using the various experiments on real ECGs of 285 subjects.

Jin et al. [3], proposed two, template protection methods using ranking based hashing techniques, which were motivated by cancelable biometric and "Index-of-Max" (IOM) hashing. Two realizations named as Uniformly Random Permutation and Gaussian Random Projections are demonstrated from this hashing notion. This system ensured strong concealment to biometric information, more robust against biometric feature variations.

Gomez-Barrero et al. [4], suggested that unlink ability and irreversibility are two international standards for protecting biometric templates. The new method for determining the unlink ability of biometric template was proposed. The efficiency of the technique was verified using four techniques called as Homomorphic Encryption, block re-mapping, biometric salting and Bloom filters.

Jin et al. [5], illustrated that, specific strings of binary with fixed length are produced from the kernel learning mechanism for fingerprint information. The first listed procedure was—Minutiae description extraction, second was a method of kernel transformation that was based on kernelized hashing to produce a vector of fixed width, third was Binarization (Segmentation) and fourth was to match. The datasets used for experimentation were FVC2002 and VC2004 to describe the possibility of proposed framework in case of template randomness, accuracy of matching and proficiency.

Lee et al. [6], developed a new technique to produce a cancellable fingerprint template without alignment. Based on constant value of rotation and translation a minutia will be calculated. The value of the minutia is based on the orientation of its adjacent local areas. The geometric interconnections between the enrolled models will be retained and transformations will be performed. The transformed templates will be operated to verify the unregistered users of the system. The success and changeability are the criteria used for analysis in the experiments to verify the accuracy of the results.

Kelkboom et al. [7], proposed an improved method FRR, in which during enrollment and verification process various biometric instances were used. In this noise, bit errors and HD were decreased. The evaluation of the output of the biometric device was proposed based on Gaussian analytical methods, which specified the number of instances used during the measuring and checking process. The error-detection trade off curve was used for the assessment of the machine efficiency, which incorporates FAR and FRR. The testing was done based on the biometric Face Recognition database Grand Challenge V2 and FVC2000.

Wong et al. [8] suggested a revocable fingerprint template called multi-line code with low complexity. The proposed multi-line code improved performance. The better network protection offering diversity and revocability using the fingerprint template was created efficiently. However, the proposed method required large storage capacity as the template was stored as a real number string.

Jin et al. [9] proposed a protection technique to protect the fingerprint minutiae. Graph-based Hamming Embedding was used to protect against inversion by using a minutiae descriptor, as Minutiae Vicinity Decomposition (MVD). User specific Minutiae Vicinities Collection scheme was then used to improve Randomized MVD. Graph-based Hamming Embedding was used to embed the MVD. Binary template has a strong concealment of the minutiae vicinity that protects the location and orientation of minutiae effectively.

Jin et al. [10] presented a fingerprint template protection mechanism to secure the fingerprint minutiae. In order to create a binary template that could be strongly secured against inversion, Randomized Graph-based Hamming Embedding (RGHE) was implemented. This approach, adopted a minutiae descriptor, was called Minutiae Vicinity Decomposition (MVD). The user-specific minutiae vicinities collection scheme then enhanced discrimination of the randomized MVD. Using the graph-based Hamming Embedding technique, it was inserted into a Hamming space. This technique has many advantages like, effectively protects the location and the orientation of the minutiae points, the MVD's well-preserved discriminability, the templates are quick, and it is a revocable method. However, the MVD features are highly likely to reveal the minutiae vicinity.

Li et al. [11] addressed the development of privacy-preserving protocols based on the combination of garbled circuit and homomorphic encryption for securing fingerprint minutiae. The suggested work promised both template and transaction security. To encrypt the template stored on the server, the user's private key is used. By means of re-encryption, the template may be modified or revoked. Two hybrid protocols were introduced in this paper with the combination of homomorphic encryption and garbled circuit to satisfy the matching minutiae. The key downside of the scheme is that, often less efficiency was incurred.

Akdoğan et al. [12] suggested a new secure key agreement protocol using biometrics with unordered set of features. The suggested protocol used fingerprints without using any randomly generated key or any random data in the key itself. A threshold-based quantization process was used to group the minutiae in a predefined neighborhood. Based on the measured similarity score on the common collection of minutiae, the acceptance or rejection decision will be made.

Li et al. [13] implemented a security analysis framework by combining information-theoretic and computational security approach. Using decision level fusion, a fingerprint based Multi Biometric Cryptosystem (MBC) was built. This was based on two elements called new security analysis framework for bio-cryptosystem-oriented and a realistic MBCD construction based on fingerprints. To further secure

MBCD, hash functions were used in each single biometric trait. However, more storage space was required.

Subramaniam et al. [14] proposed a system to generate cryptographic keys using biometrics and conventional cryptographic algorithms. These keys were used to generate shared session key between E-Passport and ES. The best method for authentication was developed by combining conventional cryptography. The proposed approach fulfilled all the security goals and can be used for real time applications to transfer information safely.

Barman et al. [15] presented an approach to generate the cryptographic keys from cancellable fingerprint template of sender and receiver. The cancelable fingerprint templates were securely transmitted using a key-based steganography technique. The templates were combined with a concatenation-based feature level fusion technique and their elements are shuffled using the shuffle key to create a unique session key for communication using hash. A revocable key for symmetric cryptography was created using irrevocable fingerprints and the privacy of fingerprints is secured by the Cancellable transformation of fingerprint template. It lacks a session-based cryptographic key, however.

Vijayakumar [16] preferred Multimodal Biometric Recognition (MBR) technique over unimodal biometric systems to provide more security. The features like iris, face, finger vein, and palm print were used for getting the highest accuracy to identify the exact person. The use of multiple features from the person improved the accuracy of biometric system.

Smys and Wang [17] introduced a CTI system using blockchain to handle the issues of sustainability, scalability, privacy, and reliability. This is capable of measuring organizations contributions, reducing network load, creating a reliable dataset, and collecting CTI data with multiple feeds. This had been tested using various parameters to determine the efficiency of the proposed methodology. Experimental results showed that it can save upto 20% of storage space.

3 Proposed Method

The method of cancelable biometric based template safety is proposed in this paper to provide the security and privacy against the problems resulting from the phenomenal use of biometric systems. The cancelable biometric turns a user's original biometric identity into a pseudo-biometric identity that is used for purposes of storage and matching. In the event of breach, the use of pseudo-identity mitigates privacy threats and facilitates revocability. The proposed system is the modification of the Multi-line Code Generation Algorithm (MLCGA) invented by Wong's [8], which converts the fingerprint image to the line which contains the minutiae in number format. In the proposed system, to reduce the complexity of the algorithm, Line Code Generation step is modified, in such a way that, on the line passing through the minutiae, instead of drawing multiples circles over the line only one circle is drawn with large radius, as shown in Fig. 2.

3.1 System Modules

The cancelable fingerprint technique is implemented using the modified multi-line code algorithm. The entire fingerprint verification system comprises of three main modules.

- **Pre-processing**: Pre-processing includes processing of scanned images to improve the image quality. The feature extraction is also performed in this step. In this, minutiae points on the fingerprints are detected and marked. These minutiae points are unique for every individual.
- **Template generation**: Fingerprints are captured using scanners in the form of images. Here, the captured fingerprint image is converted to the template. The template generation is performed using modified multi-line code algorithm. Generated template is different than actual fingerprint of user. So this pseudo template is then stored into the database.
- **Template matching**: When an user marks his attendance through biometric system, the user will be searched into the system, and then the user's captured fingerprint is used for verification purpose. This captured fingerprint is again converted using multi-line code algorithm and it will be matched with previously stored template fingerprint in the database. If the fingerprint is matched then the user will be an authenticated user and he will have permission to enter the premise.

3.2 System Architecture

As mentioned, the existing system is based on Wong's [8] system, called as Cancelable Biometric Technique using Multi-line Code Generation Algorithm (CBT-MLCGA). The proposed system is the modification of this to simplify the complexity of multi-line code generation step in the proposed algorithm, and is called as Cancelable Biometric Technique using Modified Multi-line Code Generation Algorithm (CBT-MMLCGA). Also, in the proposed system, different algorithms other than Wong's system are applied in feature extraction. The system architecture of the proposed system i.e., CBT-MMLCGA is shown in Fig. 1.

The system architecture of the proposed system called as CBT-MMLCGA has two main parts:

- Enrollment
- Verification.

Enrolment. When the user comes for the first time into the system, he is new to system. Now he needs to register himself using personal details and enrol himself with his fingerprint. This process is known as Enrolment. Scanned fingerprint of user will be an input to the system. Along with fingerprint, user should enter his personal information like full name, username, date of birth, etc. While enrolment

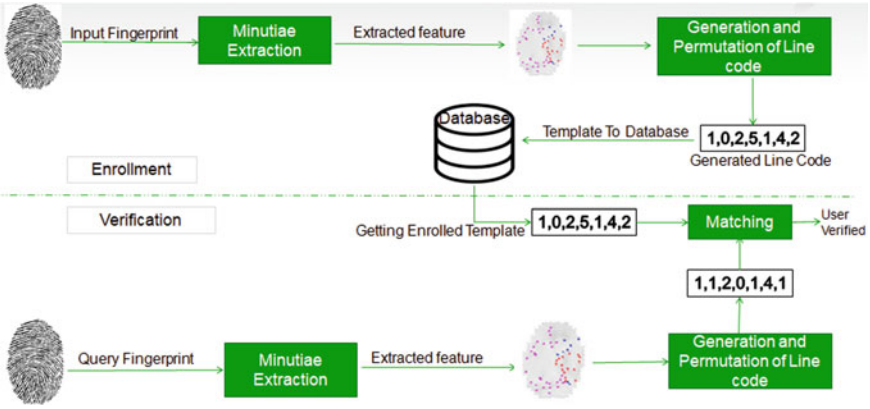


Fig. 1 System architecture

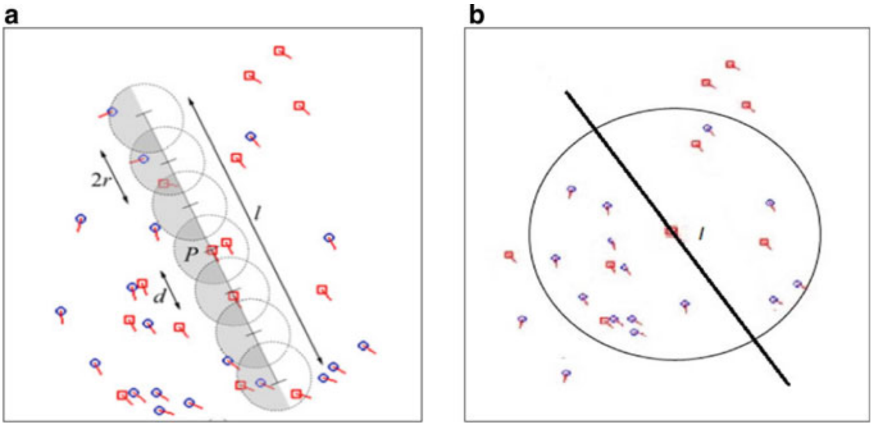


Fig. 2 **a** Plane view of Wong's system, **b** view of proposed system

using multi-line code algorithm, the fingerprint template is generated and stored into database.

Verification. When the user comes for the second time into the system, it means the user is already registered in the system. So his identity needs to be verified with the help of his fingerprint. Only verified users will be allowed to enter into the system. This process is known as Verification. Here, the scanned fingerprint will undergo multiple preprocessing steps like minutiae extraction, contextual filtering, binarization, thinning and minutiae detection. This step will give a number of minutiae points which will be unique for every individual. While verification, scanned fingerprint is again converted into template using multi-line code algorithm and this new template is matched with the previously stored template. The fingerprint verification system has the following four stages:

- Extraction of minutiae
- Generation of multi-line code
- Permutations in multi-line code
- Fingerprint matching.

Extraction of Minutiae. The minutiae extraction includes the following five stages:

> ***Segmentation***: The image contains unwanted part other than the region of interest. Likewise in the fingerprint image, which also contains some unwanted parts that need to be removed by procuring the region of interest using segmentation by excluding the background.
>
> ***Contextual filtering***: The contextual filtering is required to improve the image quality. So the quality of fingerprint image will also be improved using this. The noise from the fingerprint image will be removed by conserving the fingerprint structures. The Gaussian filter is one of the option for the contextual filtering.
>
> ***Binarization***: In image processing, it is required to convert grey scaled image into black and white. Therefore, using binarization the grey scaled actual image is transformed to a binary image in black and white.
>
> ***Thinning***: In image processing, it is required to reduce the ridges to pixels, and that can be done using thinning. Similarly, the ridges of the fingerprint are reduced to one pixel long using image to make the further processes easy. In this work, Zhang–Suen image thinning algorithm is used.
>
> ***Minutiae detection***: In the image processing, it is required to identify the bifurcation and end points of the ridges. These points are referred as minutiae and are also called as feature points.

Generation of multi-line code. Once the minutia are extracted from the given fingerprint image, Wong's technique is applied as shown in Fig. 2a, where one straight line of length (i.e. l) is drawn through the selected minutia. Then, all the sample points of that line are identified and a circle is drawn around it. Now, the minutia coming inside that circle is counted. The total minutia coming inside each circle are arranged in a sequence for generating a line code that describes the reference minutia. The proposed system is illustrated in Fig. 2b, where a single circle will be drawn instead of drawing multiple circles on a line. The total minutia coming under this circle will be counted and used for generating the line code.

Permutation in multi-line code. In case of the revocability and template diversity, some external factors must be added to bring variations into the created line code. In case of simple permutation on line code, user defined secret keys are used. The secret key changes the permutation order of the minutia code. The secret keys to be used are generated by using Pseudo-Random Number Generator (PRNG) algorithm. The secret key generated should be unique i.e., it should not be allocated to more than one user/application. This is achieved by using PRNG, which generates unique random numbers.

Fingerprint Matching. The Template Fingerprint (TF) and Query Fingerprint (QF) are used for matching. Every line code from QF is matched with each line code from TF. Based on the result of line code matching, authenticity is checked.

4 Implementation Details

The cancelable biometrics are complicated for implementation. The proposed MMLCGA reduces the complexity of multiline code algorithm used in CBT system. In the modified multi-line code, step Line Code Generation converts the fingerprint image into a line, which contains the minutiae in number format. To reduce the complexity of the algorithm, Line Code Generation step is modified. Here, on the line passing through the minutiae, instead of drawing multiples circles over the line only one circle is drawn with large radius. This reduces the computation time of algorithm and improved accuracy of the system. Thus, it helps in reducing the complexity of the existing CBT-MLCG algorithm.

The implementation is divided into different steps that are described in short as:

- Binarization
- Contextual Filtering
- Thinning
- Minutiae Extraction
- Line Code Algorithm.

4.1 Binarization

Image binarization is the process of taking a greyscale image and converting it to black-and-white, essentially reducing the information contained within the image from 256 shades of grey to black and white i.e. binary image. This is sometimes known as image thresholding, although thresholding may produce images with more than 2 levels of grey. It is a form of segmentation, whereby an image is divided into constituent objects. The pseudo code of Binarization is given below:

```
Load the image.
Calculate the height and width of image.
For loop:
    Scan the image through its height and width
    Get the RGB value of point
    Set lower value to white and higher value to black
End loop.
```

4.2 Contextual Filtering

The contextual filters are mainly used to suppress either the high frequencies in the image i.e., smoothing the image, or the low frequencies i.e., enhancing. A Gaussian filter is a linear filter. It's usually used to blur the image or to reduce noise. The Gaussian filter will blur edges and reduce contrast.

4.3 Thinning

It is commonly used to tidy up the output of edge detectors by reducing all lines to single pixel thickness. Thinning is normally applied to binary images and produces another binary image as output. Here, Zhang-Suen thinning algorithm is used.

4.4 Minutiae Extraction

Here, the ridge endings, ridge dot, and bifurcations in image are found marked as minutiae.

4.5 Line Code Algorithm

In this step, the image is converted to the line of the numbers like {15, 26, 35, 91, …}. The pseudo code of Line Code Algorithm is given below:

```
Store all the featured points in the list
Draw a line through one of the featured points
Draw circle on the line with radius r
For loop:
    Scan image for its height and width
    Find if featured point belongs to the drawn circle
    If yes, then add that point into Array
End loop
Permute the calculated line with some random number.
```

So, at the end, the line is generated with the featured points that are present in the drawn circle. The algorithm is implemented in Java, and Fingerprint Verification Competition (FVC) datasets are used for result analysis. Standard FVC datasets are downloaded from http://bias.csr.unibo.it/fvc2004/databases.asp. Here, FVC2000, FVC2002 and FVC2004 datasets are used. Every dataset has fingerprint samples for 10 users. These fingerprint samples are in JPG format.

5 Results and Discussion

The existing algorithm (CBT-MLCG) and proposed algorithm (CBT-MMLCG) are implemented on Java platform and tested based on the standard FVC datasets like FVC2000, FVC2002 and FVC2004. This is the extension of work [18].

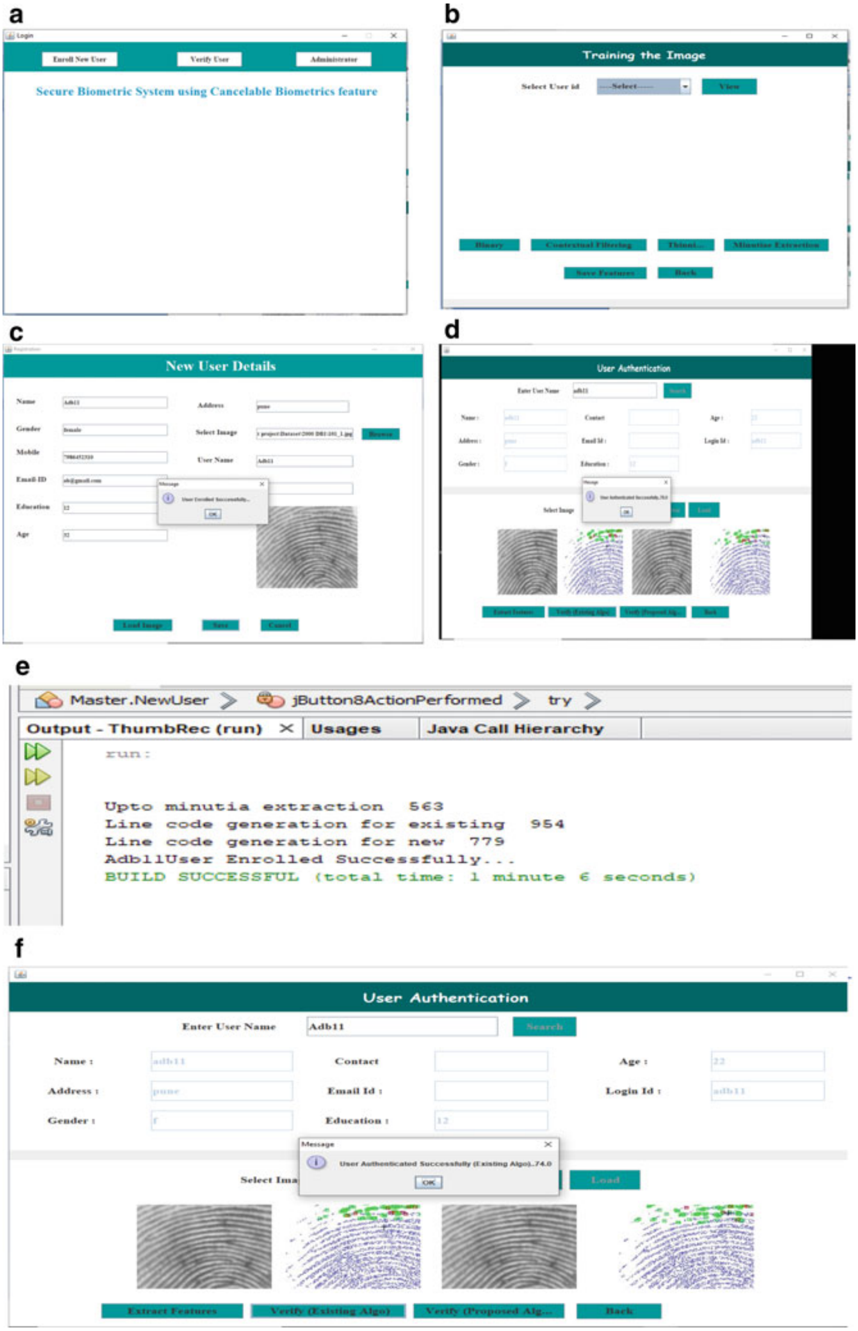
The user needs to enrol in the system first and will be verified with the existing fingerprint at the time of authentication. The enrollment phase is indirectly a training phase as in, when any new user gets enrolled in the system, training data gets updated and used in the future for authentication. Only the authorized/enrolled users will have the access to the system. The working of the system is shown in Fig. 3a–g.

The comparative analysis of existing (CBT-MLCGA) and proposed (CBT-MMLCGA) is performed based on the two parameters called as computation time and accuracy of the algorithm. The computation time using CBT-MLCGA and CBT-MMLCGA is calculated and shown in Fig. 3e, for one user from FVC2000 dataset. Similarly, it is calculated for all the users from FVC2000, FVC2002 and FVC2004 datasets and recorded in the Table 1 and their comparative analysis is shown in Fig. 4. By looking at the results shown in Table 1, the average time of the proposed technique (CBT-MMLCGA) is approximately 15% less than that of the existing technique (CBT-MLCGA).

The accuracy of the algorithm using the existing technique and proposed technique is calculated shown in Fig. 3f, for one user from FVC2000 dataset. Similarly, the accuracies for all the users from FVC2000, FVC2002 and FVC2004 datasets are computed and recorded in the Table 2 and their comparative analysis is shown in Fig. 5. By looking at the results shown in Table 2, the average accuracy of the proposed technique (CBT-MMLCG) is improved by approximately 1.4% over the existing technique (CBT-MLCG).

6 Conclusion

Technique like cancelable biometrics has the potential to boost the security and confidentiality of a traditional biometric system. It uses multi-line code algorithm for the verification of the fingerprints. This existing technique is modified, and a new technique called Modified Multiline Code Generation Algorithm, which reduces the complexity of Multiline code algorithm, has been proposed. In the modified multiline code algorithm, the step Line Code Generation convert the fingerprint image into a line, which contains the minutiae in number format. To reduce the complexity of the algorithm, Line Code Generation step is modified by drawing a single circle instead of multiple circles on the line passing through the minutiae. This reduces the computation time of algorithm and improves the accuracy of the system. The algorithm is implemented in Java and analysis have been performed using Fingerprint Verification Competition datasets. Here, FVC2000, FVC2002 and



Secure Biometric System using Cancelable Biometrics feature

Select User id: ----Select-----

Name: Adb11

Gender: female

Mobile: 7986452310

Email ID: ab@gmail.com

Education: 12

Age: 32

Address: pune

Select Image: r project\Dataset\2000 DB1\101_1.jpg

User Name: Adb11

Message: User Enrolled Successfully... OK

Enter User Name: adb11

Name: adb11 Contact: Age: 22

Address: pune Email Id: Login Id: adb11

Gender: f Education: 12

Message: User Authenticated Successfully..74.0 OK

```
run:

Upto minutia extraction  563
Line code generation for existing  954
Line code generation for new  779
Adb11User Enrolled Successfully...
BUILD SUCCESSFUL (total time: 1 minute 6 seconds)
```

Enter User Name: Adb11

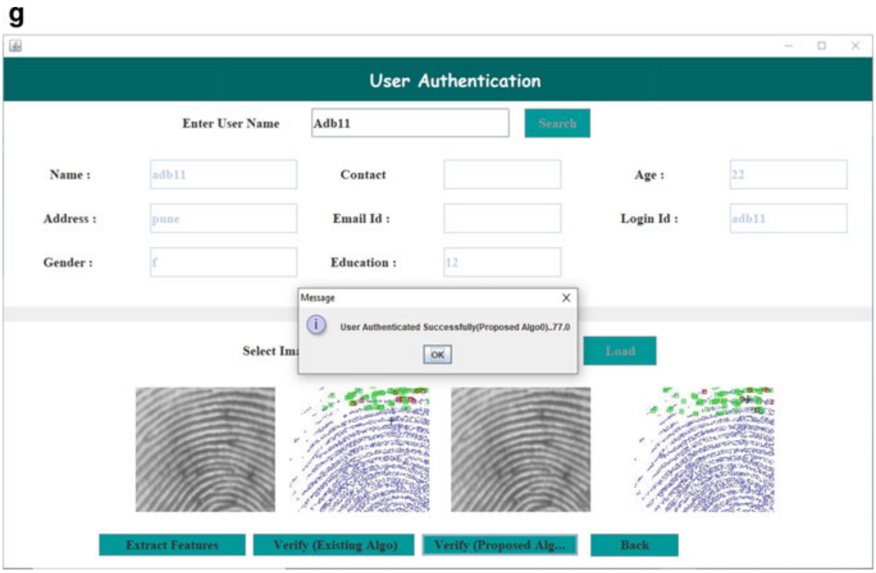
Name: adb11 Contact: Age: 22

Address: pune Email Id: Login Id: adb11

Gender: f Education: 12

Message: User Authenticated Successfully (Existing Algo)..74.0 OK

Fig. 3 **a** Initial screen, **b** training set, **c** user enrollment, **d** user authentication, **e** computation time, **f** accuracy of authentication using CBT-MLCG, **g** accuracy of authentication using CBT-MMLCG



Enter User Name Adb11

Name : adb11

Contact

Age : 22

Address : pune

Email Id :

Login Id : adb11

Gender : f

Education : 12

Message

User Authenticated Successfully(Proposed Algo0)..77.0

OK

Fig. 3 (continued)

Table 1 Time complexity

Dataset	FVC2000		FVC2002		FVC2004	
User's	Time complexity (ms)		Time complexity (ms)		Time complexity (ms)	
	CBT-MLCG	CBT-MMLCG	CBT-MLCG	CBT-MMLCG	CBT-MLCG	CBT-MMLCG
User-1	954	779	480	420	1221	1016
User-2	938	851	802	725	839	728
User-3	604	477	493	443	845	714
User-4	276	173	477	417	745	645
User-5	522	460	526	368	673	527
User-6	453	368	493	455	569	487
User-7	451	261	455	383	554	529
User-8	451	374	484	378	549	477
User-9	439	327	462	459	549	476
User-10	188	178	529	443	562	632
Avg. time complexity	527.6	424.8	520.1	449.1	710.6	623.1

FVC2004 datasets are used. The time and accuracy of the proposed technique (CBT-MMLCGA) is better by 15% and 1.4% respectively, when compared to the existing technique (CBT-MLCGA).

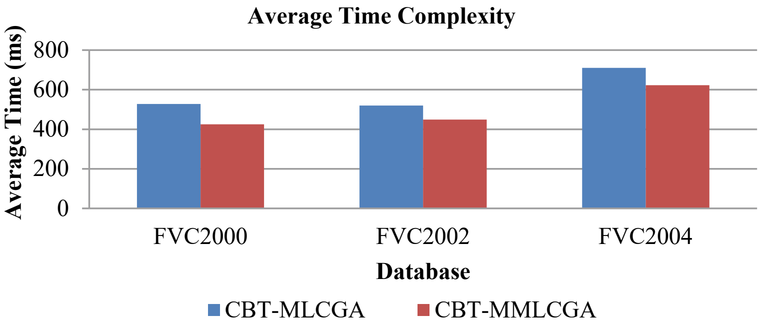


Fig. 4 Average time complexity

Table 2 Accuracy of the algorithms

Dataset	FVC2000		FVC2002		FVC2004	
User's	Percentage accuracy		Percentage accuracy		Percentage accuracy	
	CBT-MLCG	CBT-MMLCG	CBT-MLCG	CBT-MMLCG	CBT-MLCG	CBT-MMLCG
User-1	78	80	93	94	64	66
User-2	40	42	55	57	72	73
User-3	56	59	76	79	80	80
User-4	92	93	56	57	93	96
User-5	54	55	94	96	56	57
User-6	52	53	70	71	80	83
User-7	94	95	76	78	82	84
User-8	46	47	76	77	94	95
User-9	65	65	65	67	72	74
User-10	85	84	88	90	92	93
Avg. % accuracy	66.2	67.3	74.9	76.6	78.5	80.1

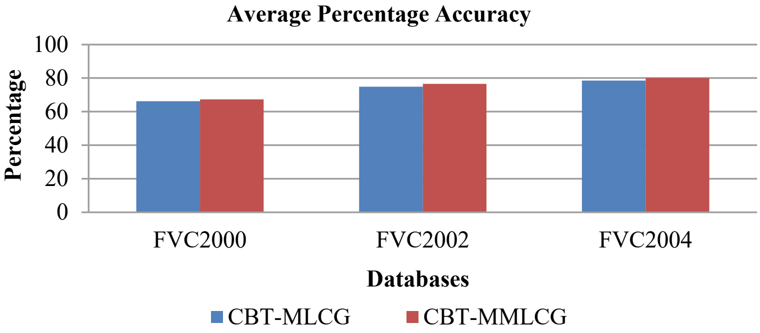


Fig. 5 Accuracy of the algorithms

Acknowledgements I would like to express my special thanks to my supervisor Dr. Sanjeev. J. Wagh for their constant support and encouragement in my research work. In this work my PG student Ms. Arti Gilbile, helped me for implementation of this algorithm, without her support this would not be possible.

References

1. Kaur H, Khanna P (2019) Random distance method for generating unimodal and multimodal cancelable biometric features. IEEE Trans Inf Forensics Secur 14
2. Wu S-C, Chen P-T, Lee Swindlehurst A, Hung P-L (2019) Cancelable biometric recognition with ECGs: subspace-based approaches. IEEE Trans Inf Forensics Secur 14
3. Jin Z, Lai Y-L, Hwang JY, Kim S, Teoh ABJ (2018) Ranking based locality sensitive hashing enabled cancelable biometrics: index-of-max hashing. IEEE Trans Inf Forensics Secur 13
4. Gomez-Barrero M, Galbally J, Rathgeb C, Busc C (2018) General framework to evaluate unlinkability in biometric template protection system. IEEE Trans Inf Forensics Secur 13
5. Jin Z et al. (2016) Generating fixed-length representation from minutiae using kernel methods for fingerprint authentication. IEEE Trans Syst Man Cybern Syst 46(10):1415–1428
6. Lee C et al. (2007) Alignment-free cancelable fingerprint templates based on local minutiae information. IEEE Trans Syst Man Cybern Part B 37(4):980–992
7. Kelkboom EJC, Molina GG, Breebaart J, Veldhuis RNJ, Kevenaar TAM, Jonker W. Binary biometrics: an analytic framework to estimate the performance curves under Gaussian assumption. IEEE Trans Syst Man Cybern Part A Syst Humans 40(3)
8. Wong WJ, Wong MLD, Kho YH (2013) Multiline code: a low complexity revocable fingerprint template for cancellable biometrics. J Central South Univ 20(5):1292–1297
9. Jin Z, Bok-Min G, Teoh ABJ, Tay YH (2014) Non-invertible analysis on graph-based hamming embedding transform for protecting fingerprint minutiae. In: IEEE 2014 international conference on electronics, information and communications (ICEIC), pp 1–2
10. Jin Z, Lim MH, Teoh ABJ, Goi BM (2014) A non-invertible randomized graph-based hamming embedding for generating cancellable fingerprint template. Pattern Recogn Lett 42:137–147
11. Li M, Feng Q, Zhao J, Yang M, Kang L, Wu L (2014) Minutiae matching with privacy protection based on the combination of garbled circuit and homomorphic encryption. Sci World J. https://doi.org/10.1155/2014/525387
12. Akdoğan D, Altop DK, Levi A (2015) Secure key agreement using pure biometrics. In: 2015 IEEE conference on communications and network security (CNS), pp 191–199
13. Li C et al (2015) A new biocryptosystem-oriented security analysis framework and implementation of multibiometric cryptosystems based on decision level fusion. IEEE Trans Inf Forensics Secur 10(6):1193–1206
14. Subramaniam U, Subbaraya K (2015) A biometric based secure session key agreement using modified elliptic curve cryptography. Int Arab J Inf Technol (IAJIT) 12(2)
15. Barman S, Samanta D, Chattopadhyay S (2015) Fingerprint-based crypto-biometric system for network security. EURASIP J Inform Sec 2015(1):1–17
16. Vijayakumar T (2021) Synthesis of palm print in feature fusion techniques for multimodal biometric recognition system online signature. J Innov Image Process (JIIP) 3(02):131–143
17. Smys S, Wang H (2021) Data elimination on repetition using a blockchain based cyber threat intelligence. IRO J Sustain Wirel Syst 2(4):149–154
18. Gilbile A, Ganjewar PD. Design of secure biometric using cancelable techniques. In: Springer international conference on data science, machine learning and applications (ICDSMLA 2020). EBook ISBN 978-981-16-3690-5, Hardcover ISBN 978-981-16-3689-9. http://doi.org/10.1007/978-981-16-3690-5

Systematic Literature Review—IoT-Based Supply Chain Management in Industry 4.0

Sreeparnesh Sharma Sivadevuni and Sathish Kumar Ravichandran

Abstract The twenty-first century has seen considerable implications towards social trends, as well as advances in technology and industrial achievements. The fourth industrial revolution is the consequence of a move towards automation and a decrease in the amount of human participation in most industries (Industry 4.0). Through an in-depth assessment of the existing available work, this article investigates the influence of technology on the Internet of Things (IoT) has and continues to have on Supply Chain Management (SCM) during the era of Industry 4.0. This analysis of the relevant literature not only provides a plethora of fresh information on Industry 4.0 but also indicates areas that require more attention and gives suggestions for the future. Investigating academic developments related to Industry 4.0 will assist in the process of closing this knowledge gap. A comprehensive review of academic articles on the impact of IoT on supply chain management in Industry 4.0 was conducted up to the end of April 2022.

Keywords Internet of Things (IoT) · Industrial revolution 4.0 (IR 4.0) · Supply chain management · Systematic literature review (SLR) · State-of-the-art · Qualitative research · Quantitative research · Research agenda

1 Introduction

The rapid pace at which economic conditions are changing in the contemporary world is the driving force behind the significance of this research. Automation, Robotics, Digital Technology, Information Technology, Industrial Automation, Augmented Reality, Virtual Reality (AR/VR), and other similar developments are all in response to growing consumer demand and advances in technology. These characteristics are referred to as the "Digital Economy" or "Industry 4.0," [1]. As a result of increased consumer expectations about the timeliness and level of quality of the services they

S. S. Sivadevuni (✉) · S. Kumar Ravichandran
Department of Computer Science and Engineering, School of Engineering and Technology, Christ University, Bangalore, Karnataka 560001, India
e-mail: Sreeparnesh.s@res.christuniversity.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_22

get, Supply Chain Management (SCM) must now contend with a whole new set of issues [2]. The use of cutting-edge technology is necessary for the sustainable development of the economy over the long term. The acceleration of industrialization throughout the globe is being driven by advances in science and technology around the world [3]. This is particularly true in the industrial sector. There are four main phases of technological evolution that are widely accepted. To date, no one can agree on exactly what constitutes an industrial revolution, but these stages are well-known, nevertheless [4]. It took two centuries for water and steam-powered mechanical manufacturing facilities to give way to electrical mass production, and it took another two centuries for electronic gadgets and information technology to completely automate (IT) [5]. New industrial facilities that were powered by water or steam were the driving force behind each of these revolutions. The Internet of Things (IoT) and Cyber-Physical Systems (CPS) have been noticed by governments and organizations all over the world, and these entities are now making use of the opportunities that these two technological advancements provide [6].

For any sector to be successful in a complex environment, it must have a very agile workforce. The workforce also needs to have a high degree of tolerance and abilities in risk reduction. In addition they must have structural flexibility that enables the persons involved to prepare for rapid reaction to the challenges that they may face with the capability of an organization's supply chain to adapt to significant changes in the operational environment [7]. However, additional resources, such as buffer inventory and surplus capacity, as well as a higher cost of coordination, are required for greater flexibility and resilience [8]. If companies are going to achieve an appropriate balance between the needed degree of resiliency and flexibility in addition to the associated costs, they need to be able to predict and track the whole supply chain and they must have the necessary velocity to react swiftly to changes. Firms must have a complete awareness of the supply chain to achieve this delicate equilibrium. So-called concepts that can lead supply chain managers are the "4R's," which stand for responsiveness, reliability, resilience, and relationships [9, 10].

2 IoT Based Supply Chain Management in Industry 4.0

Supply Chain Management (SCM) is undergoing a paradigm shift because of the Internet of Things (IoT), one of the most recent developments in information technology [11]. Communications within the supply chain are raised to a new level thanks to the Internet of Things, which introduces new possibilities such as human-to-thing communication and autonomous coordination among "things" while they are being stored in a facility or transported between supply chain entities [12]. These new abilities provide a tremendous potential to deal with SCM issues in a more effective way. As a result of the IoT, the supply chain can now be seen with a degree of transparency, agility, and adaptability that has never been possible before [13]. It is possible to get unparalleled insight into all parts of the supply chain when the data released by smart things are successfully gathered, processed, and transformed into meaningful information. It is possible that the organization may be made aware of potentially hazardous situations both inside and outdoors due to this increased visibility [14]. New levels of supply chain efficiency may be achieved when these signals are reacted to in a timely way. In the absence of information, it is not the availability of information that has been lacking; rather, it is the technology for collecting and analyzing massive volumes of data and the time lag between the collection of data and action that has been lacking [15]. As a result of the Internet of Things, supply chains will be able to react to changes in real-time and reach historically unachievable levels of agility and adaptability, allowing them to adapt to market demands in real-time [16]. Remote management of supply chain operations, better cooperation with partners, and the provision of more exact information to assist in more effective decision-making will all be made possible thanks to the Internet of Things [17].

After conducting a brief survey of literature, this article discusses the IoT and the influence it has on SCM in the age of Industry 4.0. This study covers major features of IoT in SCM including the concept of IoT, the primary IoT technological pieces required in its implementation in a supply-chain context, numerous SCM applications, aspects of contemporary manufacturing, and applications that are currently applied [18]. To provide an answer which is more suitable to the stated main research question, the purpose of this work is to review and analyze the academic progress in a methodical manner. This will allow us to provide an answer that is more relevant to the stated main research question. Because of this, we will be able to deliver an answer that is more relevant to the primary research topic that was posed. The first stage result from a larger research effort that will include reviews of other high-level projects and plans related to the fourth industrial revolution around the world [3]. These reviews will be included in the larger research effort. As a component of this more extensive research initiative, this review will be given.

3 Principles and Methods

3.1 Review Principles

Three essential review guidelines were created to guarantee that all articles could be appraised uniformly with fewer subjective opinions: Inclusion and exclusion criteria must be clearly stated [19]. Each manuscript should have a clear set of inclusion/exclusion criteria. The following segregation shows six key inclusion and exclusion criteria, as well as their subgroups, in this review.

Exclusions

1. The full text of the document is not accessible in English; just the paper's title, its abstract, and its important phrases are translated into the language.
2. An evaluation of a paper that is incomplete.

3. Articles that are not scholarly in nature do not belong in this category. Examples include editorials, conference assessments, the table of contents, and prefaces.
4. IoT, SCM and Industry 4.0 review, survey, debate, and issue solution are not the subject of a paper. IoT, SCM and Industry 4.0 is merely given as an example to illustrate the point. It is only in the context of a future research direction, viewpoint, or need that the terms "IoT, SCM and Industry 4.0" is employed [6]. A reference to IoT, SCM and Industry 4.0 can only be found in a few places. It is solely used as a phrase or reference in relation to IoT, SCM, and Industry 4.0.

Inclusions

5. Without IoT, SCM and Industry 4.0, a study of the fourth industrial revolution's role in SCM will be incomplete. If you're writing an academic article is written for utilization of IoT, SCM and Industry 4.0 as a way of describing some of the difficulties, concerns, or trends are necessary to cover its importance.
6. A clear and unambiguous focus on IoT, SCM and Industry 4.0 may be seen throughout the research activities presented in this paper.

3.2 Systematic Literature Review Method

It was necessary to conduct an extensive study of Industry 4.0 in IoT and SCM owing to this issue, as well as the results of previous studies [20]. The review for this article's goals was conducted using a structured review technique. Methods for systematic reviews typically include five stages: formulation of research questions, Locating study, selection and evaluation of study sites, analysis and synthesis of results, and reporting and implementation of findings [21]. These five stages guided the selection of this approach (Fig. 1).

Step 1: Question Formulation

As a first step, research on IoT and associated difficulties in supply chain logistics began with a study of the number of studies on the topic as well as a review of their context and methodology [22]. They also looked at previous studies to see what had worked and what had not to come up with a list of suggestions. For the purposes of data collection and analysis, two questions are presented in Fig. 2. Quantity of Evaluated and Chosen Studies in Each Database is shown in Table 1.

Step 2: Locating Studies

Finding, selecting, and evaluating relevant research to answer the review's specific questions was accomplished. Google Scholar was utilized as the initial search engine since it displays the most results from all databases. The five keyword phrases used were "Industry 4.0 and Supply Chain," "Industry 4.0 and IoT," "IoT in Supply Chain," "Smart Supply Chain," and "E-Supply Chain" [23]. The title, abstract, or keywords included in the publications were evaluated to locate related studies. As the last step, it was decided to do an extensive search of the main research databases, such as

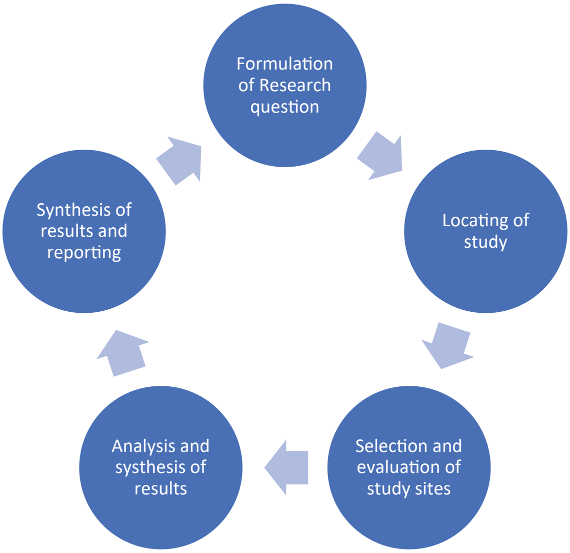

Formulation of Research question
Locating of study
Selection and evaluation of study sites
Analysis and systhesis of results
Synthesis of results and reporting

Fig. 1 The systematic review in current research is outlined below

Question 1: Industry 4.0's IoT supply chain trends: what can we expect?

Critiria Analysis: methods *How many studies have been done? Publication dates a repository for research*

Question 2: Will the Internet of Things make it easier to plan, control, and coordinate the operations involved in supply chain management?

Critiria Analysis: *These categories were chosen based on a combination of factors including consensus among experts, methodological approaches and opportunities for additional study.*

Fig. 2 Question formation and criteria analysis

Table 1 Quantity of evaluated and chosen studies in each database

Data base reviewed	Number of papers selected
Taylor and Francis	9
IEEE	16
Springer	14
Emerald	6
Wiley	8
MDPI	4
Total	57

Taylor and Francis, Emerald, Elsevier, IEEE, MDPI and Springer [24]. A lack of relevant articles could not be found due to restrictions imposed by keyword selection and database listings. All the review articles in this study were published after 2014. since the subject matter is still relatively fresh. To make matters worse, only articles published in English were considered for inclusion.

Step 3: Study Selection and Evaluation

The researchers analyzed the text of each manuscript to determine which research was relevant to this subject. They chose articles that cover the supply chain in Industry 4.0-addressed smart factories [25]. Wiley and Semantic Scholar databases were used since there were not enough publications published in this field and further study was required. This study was not found during the search of the chosen papers, but they picked relevant published articles that had possible material related to this research. A total of 57 of the 507 articles were chosen for the study.

Step 4: Analyzing and Synthesizing

Using a set of pre-formulated questions, each individual research was discussed. When looking for trends in existing research on Industry 4.0 and supply chain, we looked for things like how many articles have been published since a certain date and where in the world they were conducted [8]. We also looked for things like research methods like surveying and interviewing as well as how much content was investigated and what previous research had been done on the topic was also taken into account. An analysis of gathered articles was performed to classify them as exploratory, confirmatory, qualitative, or process/technology-level research to provide a response to the second research question. There are many types of papers: exploratory, quantitative, and process-level [6]. In addition, the non-negative matrix factorization (NMF) approach was employed to construct words that correctly characterize each cluster using TM. Human expert and machine-based methodologies are used to provide the following findings and analyses [26].

Step 5: Reporting and Using Results

Researchers provide their findings based on a technique that includes the assessment of chosen articles according to categories described in the previous paragraph and TM. Based on this evaluation, research gaps are identified and suggestions for future studies are offered for improvement. A summary and conclusions are provided at the end of this publication [27].

4 Review of Research Trends

Figure 3 displays the publication schedule from 2016 to 2022, as shown. IoT-based Industry 4.0 supply chain research began in 2016, and the number of studies on this topic is predicted to grow dramatically by the end of 2023. Since just a few articles were approved for publication in July 2022, only half of this year's data reflects this

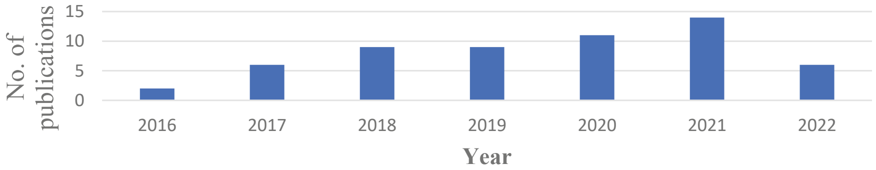


Fig. 3 Number of publications distributed by year

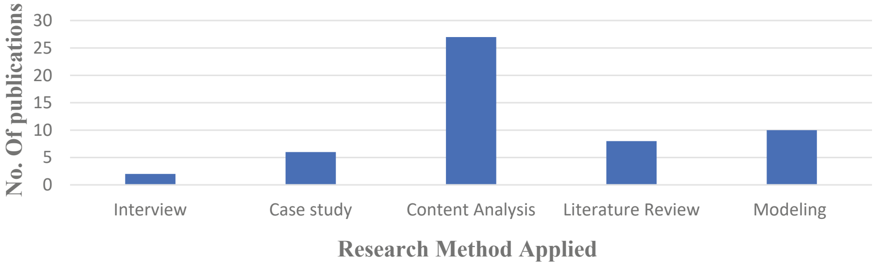


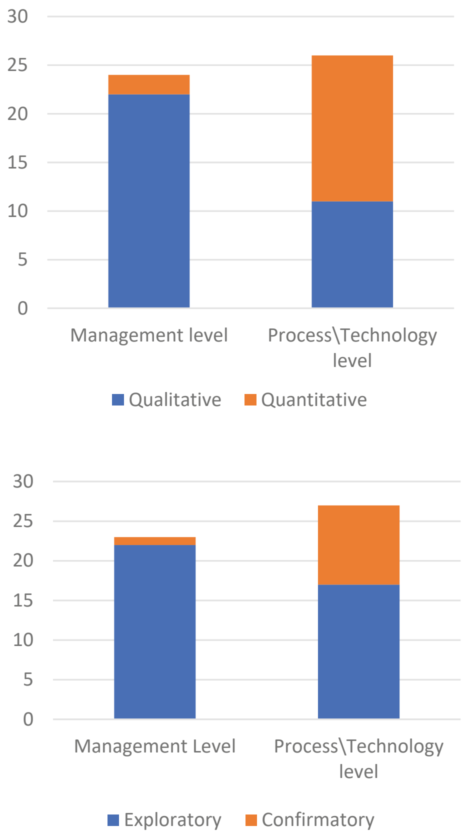
Fig. 4 Research method applied

year's trend. With a total of 8 papers published, IEEE had the most publications (see Fig. 3) followed by Springer with 9 publications. This demonstrates that the main publishers and repositories have prioritized and covered Industry 4.0 in the supply chain. Prior research has mostly used content analysis to study Industry 4.0 with IoT in the SC (Fig. 4), indicating that other empirical approaches like surveys, interviews, case studies, modelling, etc. are hardly used.

5 Content Analysis

Two distinct approaches were used to perform a content analysis to address the study's second goal, which was to identify research-relevant topics and areas. The first method relied on human expert analysis; the second method used traditional machine learning (TM) [28] and to find a solution to the question, we applied both methods. In addition to this, the most pressing problems, current trends, and recent findings are emphasized. Is there a distinction between the conceptualization of Industry 4.0 and its actual implementation in the supply chain? According to the findings of the investigation, there is no clear solution [29]. The authors were not successful in their search for a complete literature classification for the supply chain that was based on IoT based SCM in Industry 4.0. As a direct consequence of this, the authors did a comprehensive review of several different research. The several types of studies, including exploratory, confirmatory, qualitative, and quantitative, as well as management, process, and technology-level research, were each broken down further into three categories [30]. Most of the articles that were selected for this investigation are review papers. These papers are of a format that is comparable to the current investigation; however, they focus on different topics, use different methods, and cover the different subject matter, which disqualifies them from consideration in the management versus technology level categories. Because of this, it was decided that they would not be considered in this part of their investigation [31]. Ben Daya, for instance, examined both the significance of the Internet of Things to the SC as well as its position in the organization [32]. In one of the studies, a framework was developed that might be used to detect I4 in the construction supply chain (CSC). Zhong et al. published another piece of study on Intelligent Manufacturing in the age of 14, which can be found here [33]. The research carried out by Barreto and colleagues investigates the effects that has on the logistics industry [34]. Figure 5 presents the analysis level, which includes a summary of the classified papers that fall under the management and technology level and is based on both qualitative and quantitative analysis.

Fig. 5 Analysis level



Management Level

Several publications at the "management level" pushed for an industrialization and integration framework for the supply chain. Industrialization and supply chains were explored in this research. Researchers predict that the traditional supply chain will be rendered obsolete by emerging technologies such as cyber-physical systems, the Internet of Things (IoT), the iPhone/iPad/iPod Touch, smart manufacturing, and big data. It has been determined that Industry 4.0 is a platform for the exploration of new digital technologies, process optimization, and digitalization.

Process/Technology Level

Industry 4.0 is said to be based on dynamic systems that are driven by real-time data. Some of the research focused on the dynamic SC. Researchers led by Ivanov devised a mechanism for automating manufacturing facilities' intelligent supply chain planning schedules. Supply Networks of the Future's dynamic service scheduling was conceptualized by Vladimirovich Sokolov and his colleagues in 2017 [35]. Research undertaken by Dunke et al. examined how online optimization can be used to address real-time difficulties in Industry 4.0 and SC planning [31].

6 Industry 4.0 and IoT in Supply Chain Research (Future Work)

In the current investigation, the content analysis of the selected publications is discussed in two different ways: first, based on the authors' systematic review, and second, based on the TM that was used to discover research gaps and future research prospects [36]. Both methods are described in detail below.

(a) Most of the technical publications in this review highlight a conceptual or technological framework. More technical papers explaining the approach and technology implementations are needed [37].
(b) Few studies highlight Industry 4.0's influence on the supply chain and evaluate it using before-and-after case studies. It's also important to know what effects Industry 4.0 will have on firm productivity [38].
(c) Few quantitative studies exist, indicating a lack of analytical research and technological know-how [39].
(d) In addition to the three categories (exploratory, confirmatory, and qualitative/quantitative), each topic may be sorted into one of the following three categories: management, process, and technology [40].

7 Conclusion

Using a thorough literature analysis, the authors of this work assessed the current state of research on the use of Industry 4.0 in IoT based supply chain management. The results of this study's literature assessment demonstrate that the SC views Industry 4.0 as an important idea [41]. Using this notion, human contact would be reduced, and efficiency would be boosted in enterprises [42]. Engineers and academics may use these sectors as a starting point for Industry 4.0 deployment and study in these areas. A few of them are worthy of additional study and examination [43]. Industry 4.0 experts are few and far between, which hampers further investigation and conversation on the issue [44].

References

1. Abdirad M, Krishnan K (2020) Industry 4.0 in logistics and supply chain management: a systematic literature review. EMJ Eng Manag J 1–15. http://doi.org/10.1080/10429247.2020.1783935
2. Full article: 5G in digital supply chain and operations management: fostering flexibility, end-to-end connectivity and real-time visibility through internet-of-everything. https://doi.org/10.1080/00207543.2021.2002969. Accessed 31 May 2022
3. Machado CG, Winroth MP, Ribeiro da Silva EHD (2020) Sustainable manufacturing in Industry 4.0: an emerging research agenda. Int J Prod Res 58(5):1462–1484. https://doi.org/10.1080/00207543.2019.1652777
4. (PDF) The Industry 4.0 revolution and the future of manufacturing execution systems (MES). https://www.researchgate.net/publication/306150248_The_Industry_40_revolution_and_the_future_of_Manufacturing_Execution_Systems_MES. Accessed 08 June 2022
5. Mourtzis D, Vlachou E, Zogopoulos V, Fotini X (2022) Advances in production management systems. The path to intelligent, collaborative and sustainable manufacturing, vol 513, pp 354–362. Accessed: 08 June 2022 [Online]. Available: https://doi.org/10.1007/978-3-319-66923-6
6. Fatorachian H, Kazemi H (2021) Impact of Industry 4.0 on supply chain performance. Prod Plan Control 32(1):63–81. https://doi.org/10.1080/09537287.2020.1712487
7. Wamba SF, Queiroz MM (2022) Industry 4.0 and the supply chain digitalisation: a blockchain diffusion perspective. Prod Plan Control 33(2–3):193–210. https://doi.org/10.1080/09537287.2020.1810756
8. Ivanov D, Dolgui A, Sokolov B (2019) The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. Int J Prod Res 57(3):829–846. https://doi.org/10.1080/00207543.2018.1488086
9. Zheng T, Ardolino M, Bacchetti A, Perona M (2021) The applications of Industry 4.0 technologies in manufacturing context: a systematic literature review. Int J Prod Res 59(6):1922–1954. http://doi.org/10.1080/00207543.2020.1824085
10. Álvarez-Rodríguez DA, Normey-Rico JE, Flesch RCC (2017) Model predictive control for inventory management in biomass manufacturing supply chains. Int J Prod Res 55(12):3596–3608. https://doi.org/10.1080/00207543.2017.1315191
11. Mishra A, Jha AV, Appasani B, Ray AK, Gupta DK, Ghazali AN (2022) Emerging technologies and design aspects of next generation cyber physical system with a smart city application perspective. Int J Syst Assur Eng Manag. http://doi.org/10.1007/s13198-021-01523-y

12. Xu H, Yu W, Griffith D, Golmie N (2018) A survey on industrial internet of things: a cyber-physical systems perspective. IEEE Access 6:78238–78259. http://doi.org/10.1109/ACCESS.2018.2884906
13. Da Xu L, Xu EL, Li L (2018) Industry 4.0: state of the art and future trends. Int J Prod Res 56(8):2941–2962. https://doi.org/10.1080/00207543.2018.1444806
14. Yang H, Kumara S, Bukkapatnam STS, Tsung F (2019) The internet of things for smart manufacturing: a review. IISE Trans 51(11):1190–1216. https://doi.org/10.1080/24725854.2018.1555383
15. Sethi P, Sarangi SR (2017) Internet of things: architectures, protocols, and applications. J Electr Comput Eng 2017. http://doi.org/10.1155/2017/9324035
16. Ghasempour A (2019) Internet of things in smart grid: architecture, applications, services, key technologies, and challenges. Inventions 4(1). http://doi.org/10.3390/inventions4010022
17. Wang R, Wang D, Wang L, Dou Y (2021) Design of remote monitoring system for new energy vehicles from the perspective of key technologies. In: 2021 IEEE international conference on advances in electrical engineering and computer applications, AEECA 2021, Aug 2021, pp 463–467. http://doi.org/10.1109/AEECA52519.2021.9574236
18. Petrenko AS, Petrenko SA, Makoveichuk KA, Chetyrbok PV (2018) The IIoT/IoT device control model based on narrow-band IoT (NB-IoT). In: Proceedings of the 2018 IEEE conference of Russian young researchers in electrical and electronic engineering, ElConRus 2018, vol 2018, pp 950–953. http://doi.org/10.1109/EIConRus.2018.8317246
19. Li S, Da Xu L, Zhao S (2018) 5G internet of things: a survey. J Ind Inf Integr 10:1–9. http://doi.org/10.1016/j.jii.2018.01.005
20. Hassan MB, Ali ES, Mokhtar RA, Saeed RA, Chaudhari BS (2020) NB-IoT: concepts, applications, and deployment challenges. In: LPWAN technologies for IoT and M2M applications. Elsevier, pp 119–144. http://doi.org/10.1016/b978-0-12-818880-4.00006-5
21. Elgarhy O, Reggiani L (2018) Increasing efficiency of resource allocation for D2D communication in NB-IoT context. Procedia Comput Sci 130:1084–1089. https://doi.org/10.1016/j.procs.2018.04.160
22. Kusiak A (2020) Open manufacturing: a design-for-resilience approach. Int J Prod Res 58(15):4647–4658. https://doi.org/10.1080/00207543.2020.1770894
23. Buer SV, Strandhagen JO, Chan FTS (2018) The link between industry 4.0 and lean manufacturing: mapping current research and establishing a research agenda. Int J Prod Res 56(8):2924–2940. https://doi.org/10.1080/00207543.2018.1442945
24. Raza U, Kulkarni P, Sooriyabandara M (2017) Low power wide area networks: an overview. IEEE Commun Surv Tutorials 19(2):855–873. https://doi.org/10.1109/COMST.2017.2652320
25. Dolgui A, Ivanov D (2022) 5G in digital supply chain and operations management: fostering flexibility, end-to-end connectivity and real-time visibility through internet-of-everything. Int J Prod Res 60(2):442–451. https://doi.org/10.1080/00207543.2021.2002969
26. Dolgui A, Ivanov D, Sokolov B (2020) Reconfigurable supply chain: the X-network. Int J Prod Res 58(13):4138–4163. http://doi.org/10.1080/00207543.2020.1774679
27. Moghaddam M, Nof SY (2018) Collaborative service-component integration in cloud manufacturing. Int J Prod Res 56(1–2):677–691. https://doi.org/10.1080/00207543.2017.1374574
28. Ivanov D (2021) Supply chain viability and the COVID-19 pandemic: a conceptual and formal generalisation of four major adaptation strategies. Int J Prod Res 59(12):3535–3552. https://doi.org/10.1080/00207543.2021.1890852
29. Sasi A, Subramanian T, Kumar Ravichandran S (2022) Systematic literature review on industry revolution 4.0 to enhance supply chain operation performance, pp 173–179. http://doi.org/10.1145/3512676.3512705
30. Javidroozi V, Shah H, Feldman G (2019) Urban computing and smart cities: towards changing city processes by applying enterprise systems integration practices. IEEE Access 7:108023–108034. https://doi.org/10.1109/ACCESS.2019.2933045
31. Brintrup A et al (2020) Supply chain data analytics for predicting supplier disruptions: a case study in complex asset manufacturing. Int J Prod Res 58(11):3330–3341. https://doi.org/10.1080/00207543.2019.1685705

32. Ben-Daya M, Hassini E, Bahroun Z (2019) Internet of things and supply chain management: a literature review. Int J Prod Res 57(15–16):4719–4742. http://doi.org/10.1080/00207543.2017.1402140
33. Zheng P et al (2018) Smart manufacturing systems for Industry 4.0: conceptual framework, scenarios, and future perspectives. Front Mech Eng 13(2):137–150. http://doi.org/10.1007/s11465-018-0499-5
34. Barreto L, Amaral A, Pereira T (2017) Industry 4.0 implications in logistics: an overview. Procedia Manuf 13:1245–1252. https://doi.org/10.1016/j.promfg.2017.09.045
35. Ivanov D, Sokolov B, Chen W, Dolgui A, Werner F, Potryasaev S (2021) A control approach to scheduling flexibly configurable jobs with dynamic structural-logical constraints. IISE Trans 53(1):21–38. https://doi.org/10.1080/24725854.2020.1739787
36. Rao SK, Prasad R (2018) Impact of 5G technologies on Industry 4.0. Wirel Pers Commun 100(1):145–159. https://doi.org/10.1007/s11277-018-5615-7
37. Ivanov D, Dolgui A (2021) Stress testing supply chains and creating viable ecosystems. Oper Manag Res. https://doi.org/10.1007/s12063-021-00194-z
38. Dolgui A, Ivanov D, Potryasaev S, Sokolov B, Ivanova M, Werner F (2020) Blockchain-oriented dynamic modelling of smart contract design and execution in the supply chain. Int J Prod Res 58(7):2184–2199. https://doi.org/10.1080/00207543.2019.1627439
39. Jouini W, Moy C, Palicot J (2012) Decision making for cognitive radio equipment: analysis of the first 10 years of exploration. Eurasip J Wirel Commun Netw 2012. http://doi.org/10.1186/1687-1499-2012-26
40. Wu D, Rosen DW, Wang L, Schaefer D (2015) Cloud-based design and manufacturing: a new paradigm in digital manufacturing and design innovation. CAD Comput Aided Des 59:1–14. https://doi.org/10.1016/j.cad.2014.07.006
41. Zhong RY, Xu C, Chen C, Huang GQ (2017) Big data analytics for physical internet-based intelligent manufacturing shop floors. Int J Prod Res 55(9):2610–2621. https://doi.org/10.1080/00207543.2015.1086037
42. Kumar Y, Singla R (2022) Effectiveness of machine and deep learning in IOT-enabled devices for healthcare system. In: Internet of things, pp 1–19. http://doi.org/10.1007/978-3-030-81473-1_1
43. Ivanov D (2020) Viable supply chain model: integrating agility, resilience and sustainability perspectives—lessons from and thinking beyond the COVID-19 pandemic. Ann Oper Res. https://doi.org/10.1007/s10479-020-03640-6
44. Chen M, Miao Y, Hao Y, Hwang K (2017) Narrow band internet of things. IEEE Access 5:20557–20577. https://doi.org/10.1109/ACCESS.2017.2751586

A Review on Urban Flood Management Techniques for the Smart City and Future Research

Anil Mahadeo Hingmire and Pawan R. Bhaladhare

Abstract Flooding in cities is a worldwide occurrence that presents a significant problem to municipal administrations and urban planners. The loss of the life, delays in public transportation, damage to public and private property, the interruption of services such as the water supply and power supply are some of the effects of urban flooding which leads to economic losses as well as public health issues. The motive of this research paper is to review the various strategies for managing urban floods and to determine the research scope in terms of smart city development. The flood is one of the most prevalent natural catastrophes that may strike any city. Rainfall, water level, temperature, humidity, drainage water level, water discharge, as well as other parameters are generally viewed in flood prediction models including artificial neural networks (ANN), fuzzy inference processes, regression models, deep learning, gradient boosting decision trees, and self-organizing feature mapping networks (SOM). Real-time flood parameters were considered in the flood detection and warning system. Real-time flood characteristics were considered in the flood detection and warning system, and the system was constructed utilizing IoT. The accuracy of flood prediction of computational intelligence techniques is only 76.48% in average.

Keywords Urban flood · Smart city · Flood forecasting · Flood detection · Flood warning · Artificial neural network · K-nearest neighbor · Machine learning · Internet of technology

1 Introduction

Millions of people across the globe are afflicted by flood disasters, which result in substantial loss of life and property devastation. Urban flooding is becoming more common across the world, posing a serious challenge to city governments and urban

A. M. Hingmire (✉) · P. R. Bhaladhare
School of Engineering, Sandip University, Nashik, India
e-mail: anilhingmire@yahoo.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_23

planners. Urban floods cause difficulties ranging from small occurrences to major occurrences, with cities being inundated for a few hours to many days.

In developing nations, urbanization increased from less than 25% in 1970 to more than 50% in 2006. By 2020, seven of the top 10 economies in the world will be located in Asia. At the same time, Asia is one of the world's most rapidly urbanizing areas. Peak flows accelerate the occurrence of floods, which affect large populations in densely populated areas and result in catastrophic infrastructural and economic losses for business and industry. The climate model of Rafiq et al., predicted that the rainfall in winter would blow-up to the 20–30% by the 2080s.

There are two types of flood causes: direct and indirect. Natural reasons such as global climate change, changes in weather patterns, high rainfall, and so on, as well as man-made ones such as the loss of natural drains and continual urban development owing to population growth, are examples of direct causes. Indirect factors include an insufficient and faulty drainage system, on-street parking, and a poor or non-existent waste management system. Global climate change, urbanization, uncontrolled dam water releases, and inefficient and insufficient drainage systems are among the leading causes of floods, according to the report.

The global climate is made up of the sun, earth, seas, wind, rain, snow, forests, deserts, and other interrelated systems. Due to fluctuating weather patterns, we can't predict rain, snow, storms, cloudbursts, and other weather occurrences. Glaciers are receding on average at a rate of 10–15 m per year, according to a National Intelligence Council report and international studies of glaciers melting due to climate change. If the rate accelerates, river basins fed by these glaciers may flood, resulting in lower flows and water shortages for drinking and cultivation. The amount of impervious surface area in a given site increases due to urbanization, resulting in decreasing of hydrologic response time and the risk of flood. Much of the sewage and drainage system's condition is unknown. They are either overwhelmed by the water or choked with garbage and non-biodegradable plastic bags. Because of unlawful connections, sewers overflow, and the sewage system can no longer handle the increasing volume. The sudden and unexpected release of water from dams and lakes, which occurs without giving the populace the time to react, causes floods in metropolitan areas. Because most urban drainage systems were built to handle lesser amounts of runoff than those encountered today, they are prone to failure during high-intensity rainstorm storms. During a period of severe weather, even the world's brightest cities will be put to the test by floods. As a result, smart city administrations will need to take a comprehensive approach to preventing such tragedies.

Figure 1 depicts the flood management process, which includes real-time risk management, pre-flood measures, and post-flood measures. Pre-flood precautions include things like insurance, flood defense measures, and spatial planning. Risk management in real time encompasses forecasting, warning, and rescue, among other things. Including relief, cleanup, reconstruction, regeneration, etc., are post-flood measures.

Sensors and AI based algorithms can aid in the development of appropriate levels of monitoring and reaction, whether it's for calibrating dam water releases, controlling rivers, sewage network etc. Other options include storm surge obstacles, flood

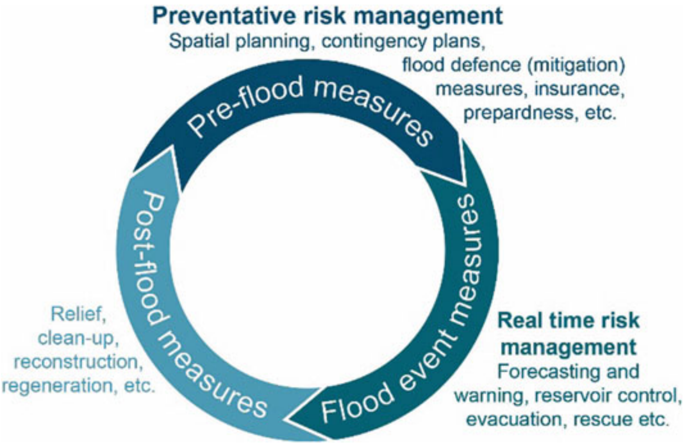

Preventative risk management
Spatial planning, contingency plans, flood defence (mitigation) measures, insurance, prepardness, etc.
Pre-flood measures
Flood event measures
Real time risk management
Forecasting and warning, reservoir control, evacuation, rescue etc.
Post-flood measures
Relief, clean-up, reconstruction, regeneration, etc.

Fig. 1 Flood management

insurance, 3D mapping and simulations, enhanced urban design, and regulatory compliance.

The aim of this paper is to analyze the key urban flood management techniques that have been proposed and implemented to mitigate urban flooding. The study's other goal is to determine the scope of future research in the subject of urban flood management. Engineers, designers, and administrators can use this review to get a broad understanding of flood management approaches for smart cities.

2 Related Work

Munawar et al. [1] focuses on flood detection using classifier model in deep learning with specific implementation of Harr cascade classifier approach on Unmanned Aerial Vehicles (UAVs) images. The 300 UAV images used to classify the flooded and non-flooded regions from the intended area. The image preprocessing, a landmark feature selection for detection, model training, employing image classification to detect floods and measuring the effectiveness of the recommended model. In these experimental results the detection roads and landmark accuracy for buildings were found 94% and 91% respectively. The accuracy observed is 91% for the given UAV input images which can be used for flood relief and rescue operation. The limitations or research gaps in the manuscript is focusing on detection of 2D flood-affected areas. A region's flood severity may be determined by measuring the depth of the floodwaters [1].

Darabi et al. [2], the work in this paper focusing on the spatial prediction of urban flood using hybrid multi-boosting neural network model (MultiB-MLPNN) and testing in the flood prone area. The comparative study of MultiB-MLPNN and

Multilayer Perceptron Neural Network (MLPNN) is performed and observed that the performance of the hybrid multi-boosting neural network model (MultiB-MLPNN) is better than the MLPNN [2–4]. The limitation or research gap in this article is that hydraulic data such as inundation depth, velocity, etc. was not available to hydraulics-based models for the study area.

Pham et al. [5] focuses on the Flood risk assessment using a hybrid model, which comprises deep learning and Multi-Criteria Decision Analysis model (MCDA). The model is trained and tested with the 847 data records of past flood locations in the study area. Deep Neural Networks (DNNs) algorithm is used to create a map of flood susceptibility, and the MCDA technique was utilized to create maps of hazards, exposure, and vulnerabilities. The model uses the geospatial databases which include curvature, distance from the river, elevation, flow accumulation, river density, rainfall, land cover and location of the flood for training purpose [5]. The limitation of this model is that it requires local geo environmental dataset for implementations at different locations.

Samikwa et al. [6] developed a model for flood prophecy using Artificial Neural Network (ANN), Internet of Things (IoT) and edge computing technology. The system leverages temporal correlative information from water level data as well as the real-time rainfall data to anticipate flood water levels ahead of time utilizing long short term memory [6]. The data collected of the water level and rainfall using the distance and rainfall sensors, respectively. The sensors are immediately linked to an Arduino Nano 33 BLE, a low power IoT BLE device. Researchers used a collection of data from Melbourne Water that included hourly rainfall and water level data. The dataset, which has 78,844 rows and was compiled hourly between 2009-12-01 00:00:00 and 2018-12-02 18:00:00, contains statistics on rainfall and water levels for around 9 years. The research gap or limitation of the model is the intensity of urban flood is not determined in the research [6].

Wu et al. [7] designed a model for the prophecy of urban flood depth based on rainfall returns period dataset using deep learning. Gradient Boosting Decision Tree (GBDT) is used in this work to forecast the depth of an urban flooded region using a regression model [7–9]. The model is based on water evaporation, rainfall duration, peak rainfall, catchment area, slope and land use such as roads, buildings, grassland, etc. The model's evaluation revealed that the relative prediction error was 11.52%, demonstrating the method's relevance to the depth prediction of urban floods [7]. The research gap or limitation is it relies on conditional factors such as slope, land use, buildings not on the real time driving factors like heavy rainfall, high soil moisture, water velocity, volume etc.

Yoon [10] focuses on urban flood forecasting based on Numerical Weather Prediction using quantitative precipitation forecasts (QPFs) method. In order to increase the precision of the forecasts, this study used the Adaptive Blending Method, which incorporates radar-based extrapolation and QPFs like the McGill Algorithm for Prediction Nowcasting by Lagrangian Extrapolation (MAPLE), the KOrea NOwcasting System (KONOS), the Spatial-scale Decomposition method (SCDM), Unified Model Local Data Assimilation and Prediction System (UM LDAPS), and Advanced Storm-scale Analysis and (ASAPS) [10–14]. This study uses radar-based

QPF data that have a geographical resolution of 1 km and a temporal resolution of 10 min. The Korea Meteorological Administration (KMA) releases QPF data for the Seoul metropolitan region, which makes it simple to assess the hydrological applicability and accuracy of different QPFs. The necessity to examine more diversified rainfall cases and study locations, as well as optimize weights utilizing various storm events need to be considered.

Yves Abou Rjeily et al. focus on establishing a flooding forecast system (FFS) that can inform the Urban Drainage System authority for probable floods ahead of time in 2017 [15, 16]. For forecasted storm events, the model estimates the depth of water variation in the critical manholes. Proposed FFS and Nonlinear Auto Regressive with eXogenous Inputs (NARX) neural network are used in the strategy to predict flooding episodes. The NARX (Nonlinear Auto Regressive with eXogenous Inputs) neural network calculates its output by fusing exogenous input with recurrent behaviour. The NARX neural network's exogenous input can represent a time series of rainfall intensity [17–20]. The limitation of this research is that it uses the small size of dataset as it considered five years data and few manholes.

Chen et al. [21] proposed a flood inundation index-based urban flood forecasting model which contains a runoff production module that categorizes surfaces as pervious or impervious, as well as a surface runoff routing method that routes surface runoff and calculates the inundation index [21]. This work offers the DPSIR (Driving Pressure Index, Impact Index, Force Index, State Index, and Response Index) urban flood warning approach, which is based on multiple index fuzzy assessment. The urban flood forecasting model includes the precipitation prediction, pipeline runoff routing, runoff production, surface runoff routing, drainage system and grid division. The approach described in the research performs better but does not concentrate on the flood routing model.

Simões et al. [22] represented a simulation work of 1D/1D, 1D/2D approach for urban drainage flood risk assessment and its performance. The paper includes the development of hydraulic 1D/1D, 1D/2D and Hybrid Infoworks of overland and sewer networks for drainage flood forecasting and calculating the computational time [23–26]. Because of the extended simulation duration, this study work's restriction or conclusions are that it is not yet suited for use in real-time flood prediction applications. A new sort of model was given to overcome this flaw. It combines a 1D and a 2D overland network, enabling 2D simulation in the most important locations with a quick simulation time and no information loss in other flooded areas [22].

In 2014, Lohani et al. created a modified fuzzy inference approach for real-time flood forecasting. By incorporating the notion of unusual and common hydrological conditions into the fuzzy modelling system, which consists of a modified Takagi Sugeno (T-S) fuzzy inference system known as the threshold subtractive clustering based Takagi Sugeno (TSC-T-S) fuzzy inference system [27, 28]. To evaluate the effectiveness of a flood forecasting model the peak percent threshold statistics (PPTS) used as a performance criteria. Using hourly rainfall and discharge data, the created model was evaluated for various lead durations [29]. The limitations of this research is that it focuses on flood forecasting and false prediction, not on warning as well as flood routing technology.

Naik et al. [30] implemented and analyzed logistic regression method to predict the flood in Kerala state. In this study, a rainfall dataset of Kerala state for last 115 years was utilized to train and forecast a flood model. The dataset considered is month wise rainfall and is from data.gov.in. The accuracy of the model was found 75% only [30]. The model's shortcoming is that the flood forecast accuracy is weak, and it has to be improved.

Mendoza-Cano et al. developed and tested the water level monitoring and flood alerting system based on IoT-based sensor network in 2021. The project used IoT and a wireless sensor network to monitor and warn about urban flooding in the Mexican city of Colima-Villa de alvarez [31]. During extreme occurrences such as tropical storms, the technique comprises a smart water network that collects real-time hydro meteorological data. For flood monitoring, the approach takes into account the soil moisture, fluvial water level, and meteorological data [32]. From a hydrological standpoint, the interconnections between tributaries and crucial sites for data collecting essential for monitoring and modelling were also taken into account. Each node has the relevant sensors, a data logger, communication module, power module, and local backup store. The system lacks a flood resilience notion while being inexpensive.

Chang and Chang [33] proposes Artificial Intelligence based Urban Flood Warning System using IoT-based Flood Depth Sensors. For continually learning and updating model parameters, the technique combines a regional flooding prediction algorithm with an online correction algorithm. This study combines the recurrent nonlinear autoregressive with exogenous inputs network (R-NARX) model with self-organizing feature mapping networks (SOM) to forecast regional floods [33]. The study focuses on flood prediction rather than flood resistance.

Vinothini and Jayanthy [34] used the Internet of Things to create a real-time Flood Detection and Notification System. The system uses a sensor network to gather data from the ground field, such as temperature, humidity, and water levels, and a Decision Tree classifier to analyse flood information in order to determine if the amount of water is normal or dangerous [34]. The DHT11 sensor is used for temperature and humidity sensor. The probe sensor is used to measure water level with PIC16F877 microcontroller. The raw data is collected from sensor values is 120 flood values. The research reveals that the suggested Decision Tree Algorithm outperforms the existing HyperPipes Algorithm in terms of accuracy. The suggested system's weakness is that it was only tested on a short dataset (only 120 flood values), and it has to be evaluated on a bigger dataset. By comparing the data from anticipated and actual flood occurrences, it is possible to assess the accuracy of the flood forecasting model [35–37].

3 Analysis and Discussion

Analysis of urban flood management techniques has been performed in this section. Table 1 show the analysis of the techniques.

Table 1 Analysis of the flood management techniques

S. No.	Author(s)	Focus of the paper	Key points in the coverage	Techniques used	Parameters analyzed	Research gaps/limitations
1	Hafiz et al. [1]	Flood detection	UAV dataset, deep learning, disaster management, landmarks detection	Haar cascade classifier, CNN classifier	Landmarks were roads, buildings, and bridges	Inclusion of floodwater depth is underlined in the paper; can be used in determination of flood intensity in a region
2	Darabi et al. [2]	Spatial prediction of urban flood	Neural networks, urban planning, GIS, artificial intelligence	Hybridized machine learning multi boosting (MultiB-MLPNN), MLPNN	Elevation, slope, precipitation, distance from river, Curve Number (CN) factor and distance from channel	Inclusion of hydraulic data such as inundation depth, velocity for hydraulics-based models
3	Pham et al. [5]	Flood risk assessment	Multi-criteria decision analysis, food risk assessment, AHP, deep learning	Deep Neural Networks (DNNs), Multi-Criteria Decision Analysis model (MCDA)	Curvature, distance from river, elevation, flow accumulation, river density, rainfall land cover and location of flood	Model depends on local geo environmental dataset for implementations at different locations
4	Samikwa et al. [6]	Flood prediction, early flood warning	Edge computing, flood prediction, artificial neural networks, Internet of things, long short-term memory	Artificial Neural Network (ANN), short-term flood prediction	Rainfall and water level	Not determined the flood intensity

(continued)

Table 1 (continued)

S. No.	Author(s)	Focus of the paper	Key points in the coverage	Techniques used	Parameters analyzed	Research gaps/limitations
5	Wu et al. [7]	Urban flood depth prediction	Depth prediction, urban flood, deep learning, data warehouse	Regression model, deep learning, Gradient Boosting Decision Tree (GBDT)	Water evaporation, rainfall duration, peak rainfall, catchment area, slope, roads, buildings, grassland	Model relies on the conditional factors such as slope, land use, buildings not on the real time driving factors like heavy rainfall, high soil moisture, water velocity, volume etc
6	Naik et al. [30]	Flood forecasting	Flood prediction, machine learning, logistic regression	Logistic regression model	Rainfall	Flood prediction accuracy is poor and need to be improve
7	Yoon [10]	Urban flood forecasting	Hydraulic modelling, short-term forecasting; blended QPF, floods	Adaptive blending method, radar-based extrapolation, real-time urban runoff forecasting using QPFs	Radar-based image data, rainfall, 1059 pipes, 773 manholes, 772 sub basins	Inclusion of diverse rainfall cases and study areas, utilizing a spatial mixing approach, weights are optimized using multiple storm events and criteria to increase forecast accuracy
8	Rjeily [17]	Urban flood forecasting	Urban drainage systems, flooding forecast, NARX neural network, proactivity	Nonlinear Auto Regressive with eXogenous inputs (NARX)	Water depth variation, rainfall intensity	Can test for large dataset for flood prediction accuracy

(continued)

Table 1 (continued)

S. No.	Author(s)	Focus of the paper	Key points in the coverage	Techniques used	Parameters analyzed	Research gaps/limitations
9	Lohani et al. [29]	Real time flood forecasting system	Self-Organizing Map (SOM), flood forecasting, fuzzy system	Fuzzy inference system, modified Takagi Sugeno (T–S) fuzzy inference system	Rainfall and water discharge data	Inclusion of flood warning system
10	Chen et al. [21]	Urban flood forecasting and warning	Fuzzy evaluation, urban flood forecasting model, fuzzy comprehensive evaluation, urban risk warning, radar based precipitation estimation	DPSIR model, DMFEW model	Pressure index, impact index, force index, state index, and response index	For flood forecasting and warning not focused on flood resilience method
11	Simoes et al. [22]	Urban drainage flood forecasting	Pluvial flooding, flood forecasting, hybrid models, urban drainage models	Hydraulic 1D/1D, 1D/2D model	Flow and water depth	It takes long simulation time
12	Mendoza-Cano et al. [32]	Urban flood monitoring and warning	Tropical storms, digital water network, early warning systems, IoT	Network design and sensor locations, smart water network	Fluvial water level, soil moisture and weather parameters	Inclusion of flood resilience method
13	Chang and Chang [33]	Urban flood warning system	Regional flood inundation forecast, Artificial Neural Networks (ANN), spatial–temporal distribution, Artificial Intelligence (AI)	Self-Organizing feature Mapping networks (SOM), Recurrent Nonlinear Auto Regressive with eXogenous inputs network (R-NARX)	Flood depth	Inclusion of flood resilience method

(continued)

Table 1 (continued)

S. No.	Author(s)	Focus of the paper	Key points in the coverage	Techniques used	Parameters analyzed	Research gaps/limitations
14	Vinothini and Jayanthy [34]	Flood detection and notification system	Sensor data, flood detection system, decision tree, IOT	Decision tree classifier	Temperature, humidity and water levels	Can test for large dataset for flood detection accuracy

Many researchers have proposed various computational intelligence algorithms for flood management, such as fuzzy systems, neural networks, machine learning, CNN, SVM, hybrid algorithms are discussed in this paper. Different dataset used in flood forecasting and flood detection by the researcher is images, water level, Temperature, Humidity, soil moisture, rainfall, water depth, water flow, pressure, water discharge data etc. Ultrasonic, radar, and depth sensors are the three most often utilized sensors for IoT water depth measurements. The 2–400 cm non-contact measuring feature is offered by the ultrasonic ranging module HC-SR04, and the ranging precision is up to 3 mm [37]. The QHR104 is a radar water level sensor that measures levels continuously. The gadget measures the water level without coming into touch with the water by using the pulsed radar technique.

It is examined that the most of research focuses on flood forecasting and monitoring from last two decades and proposes solutions based historical data and the real time data from flood forecasting to hydraulic models. The recent work mainly focuses on preparedness of the flood by prediction, monitoring and alerting models. The research gap analyzed of the discussed methodologies was the models tested by the researchers are on the smaller data size and accuracy of flood prediction variable.

Table 2 shows the analysis of various flood forecasting techniques such as ANN, KNN, Logistic Regression, Deep Learning, Gradient Boosting Decision Tree, Fuzzy inference, SVM and hybrid models.

Figure 2 shows that the accuracy of the most of models is between 80 and 90% but it is variable depending upon the cases and parameters. Though the prediction is above 76.48% in average, the rainfall and the flood is still uncertain. Hence the need arises to focus upon the development of flood resilience model which will automatically control or reduce the intensity of flood for the smart city.

4 Conclusion

Building a real-time flood forecasting system requires knowledge from several academic disciplines. They include skills in wireless flood data transfer, sensor data fusion, data sensing at the appropriate time and place, model construction, remote weather station prediction, and flood resilience. New avenues have been opened up by IoT and computational models like the fuzzy model, artificial neural networks (ANNs), machine learning, and hybrid models, permitting the creation of new hardware and software for flood monitoring and forecasting.

In order to reduce the impact of flood disasters through early warning, researchers and practitioners had conducted studies in the fields of collecting flood data, monitoring, forecasting, and detecting floods as well as early warning systems and data visualization. Many flood forecasting methods and hydrological models have been developed globally. The most of the research that have been studied at dealt with flood predicting, flood detection, and flood warning. In order to lessen flood-related fatalities, it is urgently necessary to create an efficient urban flood control system.

Table 2 Performance analysis of flood forecasting techniques

S. No.	Author(s)	Techniques used	Accuracy
1	Hafiz et al. [1]	Haar cascade classifier, CNN classifier	91%
2	Darabi et al. [2]	Hybridized machine learning multi boosting (MultiB-MLPNN), MLPNN	26%
3	Samikwa et al. [6]	Artificial Neural Network (ANN), LSTM model	97.6%
4	Wu et al. [7]	Regression model, deep learning, Gradient Boosting Decision Tree (GBDT)	Relative error of prediction: 11.52%
5	Naik et al. [30]	Logistic regression model	75%
6	Rjeily [17]	Nonlinear Auto Regressive with eXogenous inputs (NARX)	99.8%
7	Lohani et al. [29]	Fuzzy inference system, modified Takagi–Sugeno (T–S) fuzzy inference system, SVM, ANN	TSC-T-S fuzzy model: 49% SC-T-S fuzzy model: 48.6% SOM: 48.5% ANN: 48.1%
8	Chen et al. [21]	DPSIR model, DMFEW model	80%
9	Fotovatikhah et al. [14]	ANN model	89%
10	Fotovatikhah et al. [14]	SVM-LN (linear) model, PL (polynomial), RBF (radial basis function), sigmoid (SIG)	89%
11	Sankaranarayanan et al. [36]	SVM model	85.57%
12	Fotovatikhah et al. [14]	k-Nearest Neighbour (kNN) model	87%
13	Tehrany et al. [35]	Decision Tree (DT)	87%
14	Fotovatikhah et al. [14]	Logical Regression (LR) model	73%
15	Fotovatikhah et al. [14]	Naïve Bayes model	72%

A smart water management solution and an autonomous urban flood control system may be constructed to lessen the effects of urban flooding in the smart city.

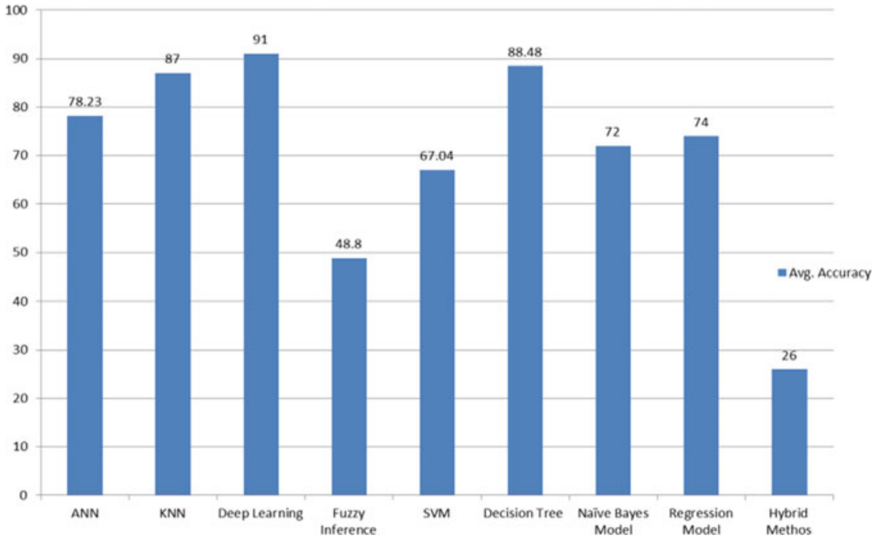


Fig. 2 Accuracy analysis of computational intelligence

References

1. Munawar HS, Ullah F, Qayyum S, Heravi A (2021) Application of deep learning on UAV-based aerial images for flood detection. Smart Cities 4:1220–1242. https://doi.org/10.3390/smartcities4030065
2. Darabi H, Rahmati O, Naghibi SA, Mohammadi F, Ahmadisharaf E, Kalantari Z, Torabi Haghighi A, Soleimanpour SM, Tiefenbacher JP, Bui DT (2021) Development of a novel hybrid multi-boosting neural network model for spatial prediction of urban flood. Geocarto Int. http://doi.org/10.1080/10106049.2021.1920629
3. Feng B, Zhang Y, Bourke R (2021) Urbanization impacts on flood risks based on urban growth data and coupled flood models. Nat Hazards 106:613–627. http://doi.org/11069-020-04480-0
4. Indrasari W, Kadarwati LV. Prototype of water level monitoring system using magnetic sensor and ultrasonic based on Arduino Mega 2560. ICOSTA 2021. J Phys Conf Ser. http://doi.org/10.1088/1742-6596/2193/1/012052
5. Pham BT, Luu C, Van Dao D, Van Phong T, Nguyen HD, Van Le H, von Meding J, Prakash I (2021) Flood risk assessment using deep learning integrated with multi-criteria decision analysis. Knowl Based Syst 219:106899. 0950-7051/2021. http://doi.org/10.1016/j.knosys.2021.106899
6. Samikwa E, Voigt T, Eriksson J (2020) Flood prediction using IoT and artificial neural networks with edge computing. In: 2020 international conferences on internet of things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData). http://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics50389.2020.00053
7. Wu Z, Zhou Y, Wang H, Jiang Z (2020) Depth prediction of urban flood under different rainfall return periods based on deep learning and data warehouse. Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2020.137077
8. Kim HI, Han KY (2020) Urban flood prediction using deep neural network with data augmentation. Water 12:899. http://doi.org/10.3390/w12030899

9. Carlson K, Chowdhury A, Kepley A, Somerville E, Warshaw K, Goodall J (2019) Smart cities solutions for more flood resilient communities. In: IEEE Xplore 2019 systems and information engineering design symposium (SIEDS). http://doi.org/10.1109/SIEDS.2019.8735625
10. Yoon S-S (2019) Adaptive blending method of radar-based and numerical weather prediction QPFs for urban flood forecasting. Remote Sens 11:642. http://doi.org/10.3390/rs11060642
11. Qian K, Mohamed A, Claudel C (2019) Physics informed data driven model for flood prediction: application of deep learning in prediction of urban flood development. https://arxiv.org/abs/1908.10312
12. Kumar N, Agrawal A, Khan RA (2019) Cost estimation of cellularly deployed IoT-enabled network for flood detection. Iran J Comput Sci 2:53–64. http://doi.org/10.1007/s42044-019-00031-4
13. Arshad B, Ogie R, Barthelemy J, Pradhan B, Verstaevel N, Perez P (2019) Computer vision and IoT-based sensors in flood monitoring and mapping: a systematic review. Sensors 19:5012. http://doi.org/10.3390/s19225012
14. Fotovatikhah F, Herrera M, Shamshirband S, Chau K, Ardabili SF, Jalil Piran Md (2018) Survey of computational intelligence as basis to big flood management: challenges, research directions and future work. Eng Appl Comput Fluid Mech 12(1):411–437. http://doi.org/10.1080/19942060.2018.1448896
15. Keung KL, Lee CKM, Ng KKH, Yeung CK (2018) Smart city application and analysis: real-time urban drainage monitoring by IoT sensors: a case study of Hong Kong. In: Proceedings of the 2018 IEEE international conference on industrial engineering and engineering management (IEEM). http://doi.org/10.1109/IEEM.2018.8607303
16. Souza AS, de Lima Curvello AM, dos Santos de Souza FL, da Silva HJ (2017) A flood warning system to critical region. Procedia Comput Sci 109C:1104–1109
17. Rjeily YA, Abbas O, Sadek M, Shahrour I, Chehade FH (2017) Flood forecasting within urban drainage systems using NARX neural network. Water Sci Technol. http://doi.org/10.2166/wst.2017.409
18. Zhang W, Wang X, Liu Y, Zhang T (2016) Simulation of rainstorm waterlogging based on SWMM and visualization module research. In: 2016 IEEE international conference on smart city and systems engineering. http://doi.org/10.1109/ICSCSE.2016.28
19. Rothkrantz LJM (2016) Flood control of the smart city Prague. In: IEEE smart cities symposium Prague 2016. http://doi.org/10.1109/SCSP.2016.7501043
20. Lo S-W, Wu J-H, Lin F-P, Hsu C-H (2015) Visual sensing for urban flood monitoring. Sensors 15:20006–20029. http://doi.org/10.3390/s150820006. ISSN 1424-8220
21. Chen Y, Zhou H, Zhang H, Du G, Zhou J (2015) Urban flood risk warning under rapid urbanization. Environ Res 139:3–10. www.elsevier.com/locate/envres. http://doi.org/10.1016/j.envres.2015.02.028
22. Simões N, Ochoa S, Leitão JP, Pina R, Sá Marques A, Maksimović Č (2011) Urban drainage models for flood forecasting: 1D/1D, 1D/2D and hybrid models. In: 12th international conference on urban drainage, Porto Alegre/Brazil, 11–16 Sept 2011
23. Yusoff A, Mustafa IS, Yussof S, Din NM (2015) Green cloud platform for flood early detection warning system in smart city. In: 2015 5th national symposium on information technology: towards new smart world (NSITNSW). IEEE. http://doi.org/10.1109/NSITNSW.2015.7176406
24. Xu G, Huang GQ, Fang J, Qiu X (2014) An integrated cloud platform for cooperative smart asset management in urban flood control. In: Proceedings of the 2014 IEEE 18th international conference on computer supported cooperative work in design. http://doi.org/10.1109/CSCWD.2014.6846820
25. Narayanan RK, Lekshmy VM, Rao S, Sasidhar K. A novel approach to urban flood monitoring using computer vision. In: Fifth international conference on computing, communications and networking technologies (ICCCNT). IEEE. http://doi.org/10.1109/ICCCNT.2014.6962989
26. Sayers W, Savića D, Kapelan Z, Kellagher R (2014) Artificial intelligence techniques for flood risk management in urban environments. In: 12th international conference on computing and control for the water industry, CCWI 2013. Procedia Eng 70:1505–1512. http://doi.org/10.1016/j.proeng.2014.02.165

27. Sunkpho J, Ootamakorn C (2011) Real-time flood monitoring and warning system. Songklanakarin J Sci Technol 33(2):227–235
28. Bruen M, Yang J (2006) Combined hydraulic and black-box models for flood forecasting in urban drainage systems. J Hydrol Eng. ISSN 1084-0699/2006/6-589
29. Lohani AK, Goel NK, Bhatia KKS (2014) Improving real time flood forecasting using fuzzy inference system. J Hydrol 509:25–41. https://doi.org/10.1016/j.jhydrol.2013.11.021
30. Naik S, Patil SA, Verma A, Hingmire A (2020) Flood prediction using logistic regression for Kerala state. Int J Eng Res Technol (IJERT) 09(03)
31. https://www.smartcitygovt.com/blog/2018/8/21/flooding-and-the-smart-city
32. Mendoza-Cano O, Aquino-Santos R, López-de la Cruz J, Edwards RM, Khouakhi A, Pattison I, Rangel-Licea V, Castellanos-Berjan E, Martinez-Preciado MA, Rincón-Avalos P, Lepper P, Gutiérrez-Gómez A, Uribe-Ramos JM, Ibarreche J, Perez I. Experiments of an IoT-based wireless sensor network for flood monitoring in Colima, Mexico. J Hydroinformatics 23(3):385. http://doi.org/10.2166/hydro.2021.126
33. Chang L-C, Chang F-J (2020) IoT-based flood depth sensors in artificial intelligent urban flood warning systems. In: EGU general assembly 2020. https://doi.org/10.5194/egusphere-egu2020-12523
34. Vinothini K, Jayanthy S (2019) IoT based flood detection and notification system using decision tree algorithm. In: Proceedings of the international conference on intelligent computing and control systems (ICICCS 2019). IEEE Xplore Part Number: CFP19K34-ART; ISBN: 978-1-5386-8113-8
35. Tehrany MS, Pradhan B, Jebur MN (2013) Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. J Hydrol 69–79. http://doi.org/10.1016/j.jhydrol.2013.09.034
36. Sankaranarayanan S, Prabhakar M, Satish S, Jain P, Ramprasad A, Krishnan A (2020) Flood prediction based on weather parameters using deep learning. J Water Clim Change. http://doi.org/10.2166/wcc.2019.321
37. https://cdn.sparkfun.com/datasheets/Sensors/Proximity/HCSR04.pdf

Application of Distributed Constraint Optimization Technique for Privacy Preservation in Cyber-Physical Systems

Manas Kumar Yogi and A. S. N. Chakravarthy

Abstract As the modern world is applying smart technologies in every day and in every sphere of life, the advent of Industry 4.0 standards increase the risk of privacy loss while using the cyber physical systems. There are many popular methods of privacy preservation in CPS but they are too complex to implement and not have high degree of data utility. In keeping all these reasearch challenges, we have applied a novel approach of distributed constraint optimization for minimizing the privacy loss in a cyber physical system. The main advantage of our proposed mechanism is that the balance between local privacy level and global privacy level is maintained so that the data utility is not degraded.

Keywords Cyber physical systems · Privacy · Distributed · Utility · Constrainst

1 Introduction

Network and interoperability have become critical with regards to the digitized business, being liable for associating items, machines, individuals and the climate in a solitary shrewd assembling system. The reconciliation of these with data advancements has driven organizations to accomplish levels of proficiency previously unheard of. Data innovations support the capacity, assortment and investigation of information while working innovations support the production of actual value. It is the consolidation between the two that leads to the much sought-after 'cyberphysical systems', made conceivable by a focal framework that controls and screens tasks at all levels [1]. Cyberphysical systems are a bunch of intuitive systems upheld by clever machines which, composed and constrained by a focal substance, send functional data to qualified laborers. Along these lines, creation tasks run under

M. K. Yogi (✉)
CSE Department, Pragati Engineering College (Autonomous), Surampalem, Andhra Pradesh, India
e-mail: manas.yogi@gmail.com

A. S. N. Chakravarthy
CSE Department, JNTUK, Kakinada, Andhra Pradesh, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_24

restraint while engineers deal with on expected issues and potential arrangements straightforwardly in this robotic copy. In numerous areas of human exercises, CPS has acquired and more consideration, particularly in the limits where physical cycles and physical hardware should have been controlled, arranged and facilitated smartly. Taking into account the intricacy of present day CPS, issues of guaranteeing the security and wellbeing of those systems are of high pertinence [2, 3]. The potential dangers can be connected with a cyber, physical or the two components of CPS and in this manner require complex methodology for recognizable proof and relief of safety and security weaknesses. Distributed constraint optimization (DCOP or DisCOP) is the distributed simple to constraint optimization [4, 5]. A DCOP is an issue where a gathering of agents must distributedly pick values for a bunch of variables to such an extent that the expense of a bunch of constraints over the variables is limited. Distributed Constraint Satisfaction is a framework for depicting an issue as far as constraints that are known and enforced by particular members (agents). The constraints are portrayed on certain variables with predefined spaces, and must be appointed to similar qualities by the various agents. Issues characterized with this framework can be tackled by any of the algorithms that are designed for it. DCOPs are a well known approach to formulating and taking care of multi-specialist coordination issues, for example, the distributed scheduling of gatherings, distributed coordination of automated air vehicles and the distributed designation of focuses in sensor networks [6, 7]. Privacy worries in the scheduling of gatherings and the limit of communication and calculation assets of every sensor in a sensor network makes brought together constraint optimization troublesome. Therefore, the idea of these applications require a distributed methodology.

2 Related Work

With the inescapability of cell phones and situating advances, area based administrations (LBSs) turned out to be progressively normal in practically all friendly and business spaces [8, 9]. Various portable LBS applications, for example, route, long range interpersonal communication and administration suggestion have been formed and incorporated into individuals' day to day exercises, giving accommodating information about their environmental factors. Notwithstanding, when trajectory information are distributed for examination, a pernicious client might have the option to find individual and delicate information of people, regardless of whether any expressly recognizing information is eliminated before distributing [10]. With the information on semi identifiers, i.e., focuses that can be connected to outer information and used to reidentify people, an assailant might have the option to anticipate back the mysterious individual development information [11]. Hence, extreme worries about privacy are raised, since this measure of information gives area information that recognizes people and, possibly, their touchy information like social traditions, strict inclinations and sexual inclinations. In the recent times, the course of information distribution has expanding and more perplexing. The extents of informational indexes have developed exorbitantly to where the adequacy of customary algorithms, which run in concentrated registering conditions, is turning out to be progressively troublesome. It is normal to handle enormous informational indexes with the guide of distributed platforms, for example, the MapReduce framework, to work with distributed information handling for huge scope bunches and accomplish significant performance improvement [12]. Information anonymization approaches proposed for incorporated figuring conditions normally present poor performance to process identifiers and to offer privacy from this information. Furthermore, they have zeroed in on the nature of privacy protection and the utility of the distributed information. Notwithstanding, adaptability is additionally an important issue when the quantity of semi identifiers are excessively high. Therefore, the fundamental inquiry raised while thinking about this multitude of issues is: “Is it conceivable to anonymize huge scope informational collections without losing adaptability?” The research lies mainly in the aspects of dynamical systems like CPS where privacy cannot be treated as a centralised feature [11]. Due tot he privacy requirements of users changing from time to time and the uncertainty in the way they interact with an CPS entity, forced us to think in the aspect of distributed privacy [13]. It is easier to postualte company privacy policies for an organisation but while enforcing the rules these consume more effort and time. Hence the main challenge lies in optimizing the privacy requirements by formulating it as a distributed constraint optimization issue. Many advances in privacy models have been made to tackle the issue of information anonymization. k-secrecy is a notable model zeroed in on social data set conspires, that forestalls character revelation by guaranteeing that for each mix of upsides of semi identifiers, there are basically k records in the distributed informational index [14]. All in all, each record can not be linked to a person by an attacker with likelihood lower than 1/k. K-secrecy has likewise been as of late considered by trajectory anonymization strategies to offer privacy mindful information publishing, although it has been shown that it presents a few impediments and finding an ideal k-anonymization is NP-hard. Speculation and concealment are the most utilized strategies to ensure the k-obscurity property. In this situation, a few anonymization strategies have been proposed to safeguard trajectory against personality revelation. Never Walk Alone (NWA) utilizes (k, δ)-obscurity privacy model to anonymize trajectories by producing round and hollow volumes of sweep δ that contain basically k trajectories [15]. For this reason, trajectories are first isolated into disjoint gatherings grouped utilizing the Euclidean distance. Each group will contain basically k trajectories with Euclidean distance δ. Be that as it may, the algorithm might bring about little gatherings with less than k trajectories, not giving significant privacy ensures, practically speaking. scientists have additionally proposed a methodology in light of information speculation to make anonymization gatherings. They considered timestamps as semi identifiers and expected that every trajectory has its own arrangement of times as its semi identifier. Subsequently, they proposed an idea of k-namelessness for trajectory information by characterizing an attack graph for foes. In the event that each hub in the graph has degree k and the attack graph is symmetric, it implies a moving item fulfills k-namelessness. To accomplish the k-secrecy property for every trajectory, anonymization bunches are not be guaranteed to disjoint. In that work, the authors accept that the information distributer knows the semi identifiers for every trajectory, except they don't specify how this knowledge can be acquired. As of late, one more category of methods, specifically differential privacy, has gotten a developing interest in information anonymization. Most works on differential privacy center around unambiguous information insightful tasks, for example, count inquiry addressing and continuous example mining. The primary thought of these strategies is to deliver loud information and to guarantee that the expulsion or expansion of a solitary record doesn't fundamentally affect the result of any investigation. In this manner, unlike our work, they can't protect information honesty, which is important to safeguard in numerous applications.

Databases assume a critical part in logical movement. However, the actual fuel of databases, open data, behaves like a soil when matched with the meshwork of our security regulations. Since the beginning of reidentification a precarious circumstance has arisen: however we can assemble and handle extraordinarily tremendous measures of data, the legitimate sloughs encompassing sharing this data acts restrictively [16]. Synthetic data offers progress. However not a silver shot, the technique permits us to stop the deidentification-reidentification weapons contest and spotlight on what makes a difference, helpful data. To this degree, we suggest that the security local area acknowledge synthetic data as a legitimate, following stage to the database-protection issue. Our mechanism should be used in conjnction with the features of synthetic data to overcome the privacy leak pertinent in the most popular privacy preservation systems employed in CPS.

3 Proposed Mechanism

We intend to apply the privacy preservation problem as a distributed constraint optimization problem. As the various entities in a CPS ecosystem interact with each other to complete their full functionality, the motivation is suitable. Also the privacy aspect is intended not to be a global issue but distributed local issues. In the sense, we apply the distributed constraint to the privacy of each CPS agent calling it as local privacy. When the agents send messages to other agents in the CPS ecosystem, they are aware of only thier local privacy level which must be optimized. Like this, we formulate the global privacy limit as an optimization problem as shown below.

Consider the following optimization problem

$$\min \sum_{i=1}^{N} P(x) \tag{1}$$

where P(x) represents the minimum privacy leak for the whole CPS ecosystem where the users are from 1, 2, 3… to N. These are legititate users. They want their sensitive information to be hidden without effecting the data utility.

We consider a network of N processors communicating according to a connected and undirected graph G = ((1, 2, 3, …N); E), where E denotes the set of edges. Edge (i, j) depicts the notion that node i sends data to j. Note that, being the graph undirected, for each (i, j) belongs to E, then (j, i) also belongs to E. We denote by [E] as the cardinality of E. Each CPS node i knows only f_i, g_i and X_i, and tries to contributes its privacy leak value at minimum level by acting locally.

Distributed Constraint Algorithm for Privacy Preservation
CPS node states: x_i, α_i, φ_i, μ_{ij} for i, j belongs to 1, 2, …, N.

Where xi is the privacy vector values of a CPS node i.

α_i represents the privacy barrier, i.e. maximum privacy loss that is permissible when two CPS nodes exchange information among themselves

φ_i, μ_{ij} are the auxilliary variables which enter into the privacy preservation scenario when scaled globally in the CPS ecosystem.

Evolution:

Collect μ_{ij} from i, j

Compute(($x_{i,}$, α_i), φ_i) as a primal dual solution pair of

$$\min \sum_{i=1}^{N} P(x)$$

$$\text{subject to } \alpha_i > 0,\ x_i \text{ belongs to E}$$

$$\text{and } \sum_{j=1}^{N} \left(\mu_{ij} - \mu_{ji}\right) \le 1 \tag{2}$$

Collect φ_i from j belongs to E.

Update for all values of j as shown below

$$\mu_{ij}^{t+1} = \mu_{ij}^{t} - \eta\left(\varphi_i^{t+1} - \varphi_j^{t+1}\right) \tag{3}$$

We can observe that, each CPS entity i tries to determine the optimal mechanism by interacting locally with neighbouring CPS entity. In the above proposed algorithm, the CPS nodes interact with each other by exchanging their privacy requirements but they donot reveal any information about their exact degree of privacy release. So, the feature is so robust that it leads to minimum privacy loss in the whole CPS.

We have kept the initialisation of the privacy needs of the nodes at random values, so that the algorithm can handle the dynamic nature of privacy requirements of the inherent CPS nodes. If any CPS node join or leave the CPS ecosystem, the privacy preservation problem will be converted into a new optimization problem. Our novel approach will help in giving a solution which will eventually converse to a solution of the latest optimization problem.

4 Experimental Results

For deploying our proposed approach we have considered a lung cancer dataset downloaded from Kaggle which has 15 features and 284 instances. We have used python language to develop the code applying the distributed constraint optimization algorithm. In Fig. 1 it has been shown that our proposed novel mechanism performs better when compared to other popular methods of privacy preservation in a CPS ecosystem like location privacy, personal differential privacy, differential privacy. From the graph, we can observe that our proposed appraoch maintains a balanced level between data privacy and data utility. The main challenge is the tradeoff between data privacy and data utility and the distributed constraint optimization method tries to balance this tradeoff evenly.

As evident from the graph below, the privacy loss reduces below 4% and data utility is nearly 10% with application of our proposed approach.

From Fig. 2 we can observe how the increase in number of iterations helps in reduction of privacy loss. In this sceanario also, our proposed technqiue outperforms the other three popular methods of privacy preservation in a cyber physical system. As observed fromt the above graph, with nearly 10 iterations the privacy loss can be brought down below 4% but other methods have privacy loss of nearly 5% on avaerage even with 10 iterations of the algorithmic run.

From Fig. 3,we can infer that as the number of interactions between the various CPS nodes increases, the local as well as global privacy level of the entire CPS also increases. By using our novel approach the privacy level increases by nearly 7% which is better than other methods of privacy preservation in comparision. The global level of privacy may remain below the accepted level of privacy which is

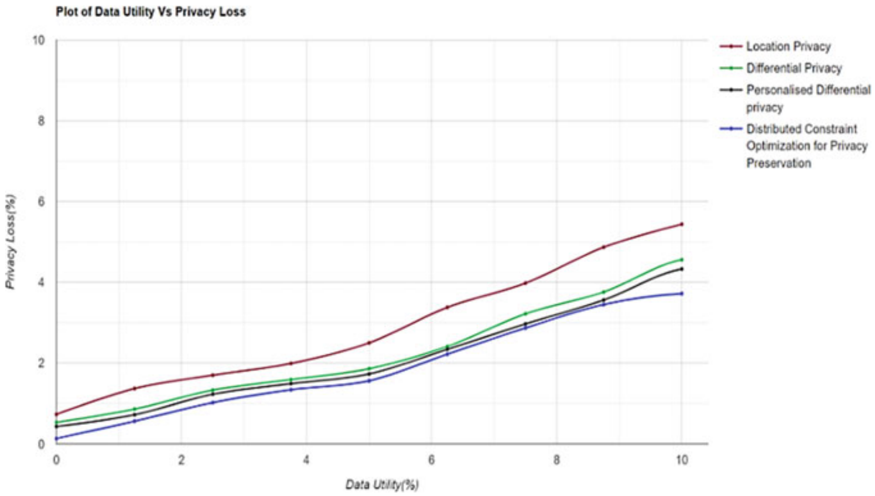


Fig. 1 Plot of data utility versus privacy loss among popular methods for privacy preservation in CPS

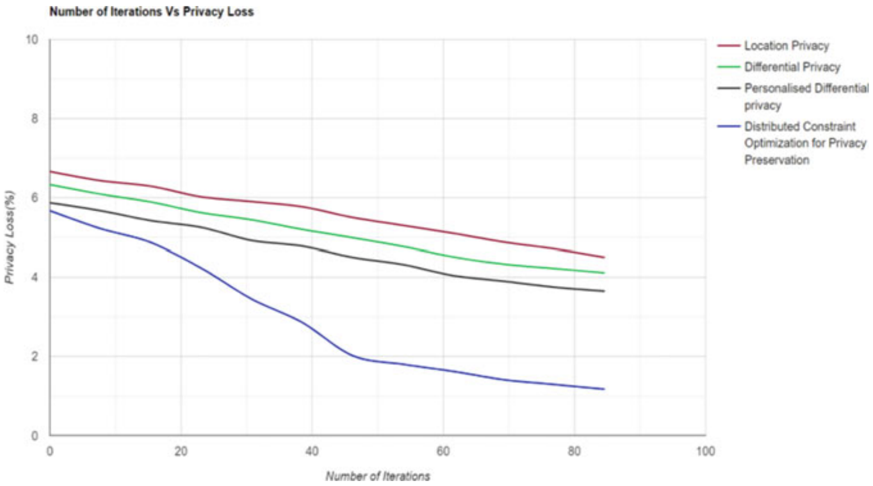


Fig. 2 Impact of number of iterations of the methods against privacy loss (%)

less than the worst case privacy level of sum of all local privacy levels of individual CPS nodes. Whether it is a cyber node or physical node, the privacy leak should be kept at a minimum level else the auxilliary variables values in the algorithm cannot decrease the privacy levels. The auxilliary variables help in keeping the local privacy levels bounded. This helps in scaling the overall global privacy level with a minimum cost function. The local agents donot know what is the global privacy level and this principle of abstraction supports in enhancing the total data utility value at optimum level. This is the most appreciable fact about the proposed technique.

5 Future Work

In problem formulation we have assumed few factors. The first assumption is that each inetracting agent in the CPS ecosystem is aware of a privacy constraint value and factors effecting the privacy constraints. But in reality, it may not be the case. The second assumption is that, each agent interacts with its neighboring agents only by passing a direct signal. We are ruling out indirect communication between the CPS nodes and various agents but again the doesnot hold true in all the cases in real time. In a dynamic system like CPS, there are multiple instances of indirect communication between the cyber components and physical components. In future work, we will try to reduce the scope of this assumptions so that the robustness of our proposed approach will increase and the dependency of the agents on the indirect interactions between the various CPS nodes will not effect much the inherent privacy loss. In future we also want to develop privacy preservation models. In future we also plan to introduce synthetic data as input to the proposed technqiue so that there

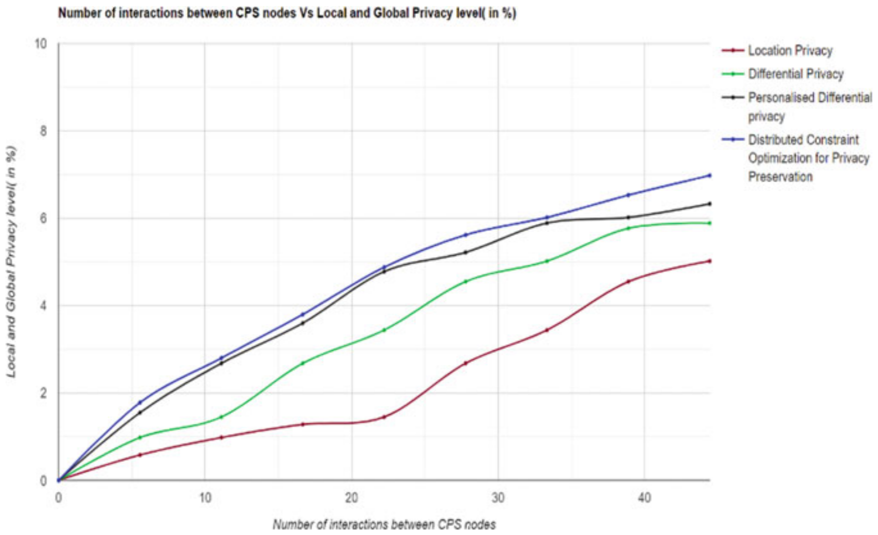


Fig. 3 Plot of number of inetractions between various CPS nodes versus local and global privacy levels (in %)

is a two-fold benefit in privacy level. We want to advocate a factor of privacy barrier which will act as an indicator of how the balance between local privacy level and global privacy levels can be optimized so that data utlity is maximized. Also, there lies oppurtunity to include the adversary knowldege to develop threat models on which the proposed techniques should be deployed so as to know the weaknesses, if any on our novel approach. Overall, we can say that in future lot of potential design issues in privacy can be solved by working in this future direction.

6 Conclusion

Our paper will serve as a guide for researchers who want to explore the constraint satisafction strategy for optimal allocation of privacy budget in each participating node of a cyber-physical system. The distributed constraint optimization helps in maintaining the overall privacy loss. With these approach the balance between data privacy and data utility is also maintained because if it is not maintained then the user will not get satisafction with the functionality of the system. This paper provides a robust framework on which future researchers can build efficient privacy preservation models by extending with other types of distributed constraint satisafaction models.

References

1. Ali W et al (2020) ALPHA: an anonymous orthogonal code-based privacy preserving scheme for ındustrial cyber–physical systems. IEEE Trans Ind Inform 17(11):7716–7724
2. Keshk M et al (2021) Privacy-preserving schemes for safeguarding heterogeneous data sources in cyber-physical systems. IEEE Access 9:55077–55097
3. Kanchan S, Singh G, Chaudhari NS (2022) SPSR-VCP: secure and privacy preserving SignRecryption in vehicular cyber physical systems. J Ambient Intell Humaniz Comput 13(1):1–20
4. Cai Z, Zheng X (2018) A private and efficient mechanism for data uploading in smart cyber-physical systems. IEEE Trans Netw Sci Eng 7(2):766–775
5. Yu J et al (2018) Privacy-preserving data aggregation computing in cyber-physical social systems. ACM Trans Cyber Phys Syst 3(1):1–23
6. Feng J et al (2020) Privacy-preserving computation in cyber-physical-social systems: a survey of the state-of-the-art and perspectives. Inf Sci 527:341–355
7. Bhattacharjee A et al (2021) Vulnerability characterization and privacy quantification for cyber-physical systems. In: 2021 IEEE international conferences on internet of things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData) and IEEE congress on cybermatics (Cybermatics). IEEE
8. Wang H, Fan K, Zhang K, Wang Z, Li H, Yang Y (2021) Secure and efficient data privacy-preserving scheme for mobile cyber physical systems. IEEE Internet Things J
9. Kanagaraj A et al (2022) Differential privacy techniques-based ınformation security for cyber physical system applications: an overview. In: Real-time applications of machine learning in cyber-physical systems, pp 52–64
10. Lu W et al (2021) Edge blockchain assisted lightweight privacy-preserving data aggregation for smart grid. IEEE Trans Netw Serv Manag 18(2):1246–1259
11. Gao C et al (2021) Sampled-data-based fault-tolerant consensus control for multi-agent systems: a data privacy preserving scheme. Automatica 133:109847
12. Li Q et al (2021) Privacy-preserving distributed processing: metrics, bounds and algorithms. IEEE Trans Inf Forensics Secur 16:2090–2103
13. Hefner T, Shani G, Stern R (2022) Privacy preserving planning in multi-agent stochastic environments. Auton Agent Multi-Agent Syst 36(1):1–27
14. Wang Y et al (2021) Privacy-preserving consensus for multi-agent systems via node decomposition strategy. IEEE Trans Circuits Syst I Regul Pap 68(8):3474–3484
15. Mugunthan SR (2019) Security and privacy preserving of sensor data localization based on internet of things. J ISMAC 1(02):81–92
16. Shakya S (2019) Efficient security and privacy mechanism for block chain application. J Inf Technol 1(02):58–67

Grip Assisting Glove for Charcot-Marie-Tooth Patients

Varun Sarathchandran, Jason Vincent, Juel Mathais George, Polu Sathwik Reddy, and R. Ambika

Abstract Charcot-Marie-Tooth is a neurodegenerative disease which causes muscle degeneration and loss of sense in the appendages. Patients affected with the disease find holding an object in their hand difficult, since they have no sensation in their skin and have no idea how well they are grasping it. Over time, it is possible that their grip falls, and the object slips from their hand. To counter this, a device which improves the quality of life of people affected with Charcot-Marie-Tooth has been presented in this paper. The proposed device is in the form of a glove, which uses force sensors to continuously track the force applied by the user on the object they are holding. If it falls below a certain threshold, it warns them through haptic feedback-vibration motors placed in the forearm where the users' sensation of touch is maintained. The sensors are paired with analog signal conditioning circuitry to obtain as good a linearity as possible without sacrificing resolution.

Keywords Assistive technology · Analog signal processing · Force measurement · Haptic feedback

1 Introduction

Charcot-Marie-Tooth (CMT) is a neurodegenerative disease that leads to the degeneration of a neuron's myelin sheath which causes poor sensation in the patient's fingers and toes. Methods to rehabilitate the hand function of patients are scarce since CMT is a rare disease, and studies of how hand function decreases over disease progression is sparse [1, 2]. Charcot-Marie-Tooth affects approximately 1 in 3000 people [3]. It is the most common inherited neuro-muscular disorder [4]. The severity of symptoms can vary greatly from person to person. It is passed on genetically and is caused by mutations in the genes that affect the nerves in your extremities [5, 6]. Daily activities become difficult tasks for those affected-owing to weakness in the

V. Sarathchandran (✉) · J. Vincent · J. Mathais George · P. Sathwik Reddy · R. Ambika
BMS Institute of Technology and Management, Yelahanka, Bangalore 560064, India
e-mail: varunsarath@gmail.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_25

hands along with decreased sensitivity to touch, heat, and cold in the feet and lower legs.

This project aims to design a system that detects and informs the user affected by CMT when their grip on an object recedes. A device is offered in the form of a glove which continuously measures the amount of force the user is applying on something they are holding. When the device senses that the user is gradually losing their grip, a vibration motor triggers in the forearm- indicating to the user that they must reinforce their grip. The system makes use of a set of force sensors that sends signals through an analog signal conditioning circuit which is then fed to the microcontroller. The microcontroller uses the conditioned signal and determines if the received signal crosses the defined threshold required to trigger the vibration motor. The threshold defined is a marker that indicates when the force detected by the force sensor is dropping, that is, it indicates if the person holding an object is losing their grip. If the force signal drops below the threshold, the microcontroller sends the signal that initiates the vibration motor. The vibration motor located at the forearm, indicates the user that their grip is decreasing beyond the set threshold.

2 Literature Survey

An extensive literature survey has been carried out and this section presents some of the existing methods.

2.1 Smart Muscle Strength Assessment Glove for Rehabilitation Purposes [7]

A low cost, muscle strength assessment device was designed. It used five force sensors and a load cell to measure the force application capability of the muscles in the finger. The device then classified the users muscle function using Neural Networks. Real data was used to train the algorithm. Based on the result, the physiotherapist was advised a therapeutic plan.

2.2 Calibration and Evaluation of a Force Measurement Glove for Field-Based Monitoring of Manual Wheelchair Users [8]

A glove was designed to detect loading events and relative changes in the application of force in wheelchair users. The glove comprises of four force sensors in the palm. Each FSR was fitted into the glove using sewed compartments. FSRs were connected to a PCB through a cloth sleeve. An elastic band fastened the PCB housing to the user's forearm. The PCB sampled the sensor, and the data was stored in a 32 GB microSD card for further processing. The glove managed to detect 72.7% of the peaks. A certain calibration condition, which was the average of the 5 conditions was tested, and gave the least error for all metrics and a linear correlation r equal to 0.80. Thus, it can be concluded that the average loading condition located at the palm is significantly aligned with the average of all the testing conditions than any individual condition.

2.3 A Fabricated Force Glove that Measures Hand Forces During Activities of Daily Living [9]

14 flexiforce sensors were used to measure forces applied by different parts of the fingers. Results from the study were used to determine minimal forces of the hand during activities of daily living to appropriately design exoskeletons and prosthetics. The flexiforce sensors—A201, which were attached to the palmar side of the hand, were directly connected to an arduino microcontroller via a resistive voltage divider circuit. The difference in voltage across one of the resistors was used to map the voltage-force ranges. A low pass filter of 10 Hz was then used to remove noise. The study concluded that forces applied during daily activities lie between 9 and 24 N.

2.4 Wearable System with Embedded Force Sensors for Neurological Rehabilitation Trainings [10]

A glove which is compact and energy efficient, thus suitable for daily wearing by patients was designed. It integrated four A301 sensors on the four fingers. The device used a ATmega 328 on Bluno Beetle board capable of wireless transmission of signals. Since the sensor in itself is flexible, implementing it on a soft surface leads to inaccurate results. Thus, two small discs were 3D printed and adhered to both sides of the sensor head. A circuit for analog signal conditioning was designed to compromise in terms of force sensing resolution and linearity of transduction. The glove was designed to be used for neurological rehabilitation training.

2.5 Tactile Sensorized Glove for Force and Motion Sensing [11]

The paper aimed to assess hand related injuries such as thumb tendonitis and carpal tunnel syndrome. It focused on the constant use of smartphones causing thumb and wrist injuries, which left untreated might cause permanent injury. Finger function testing methods are often bulky and not dynamic which restricts hand movements. To improve present devices, a glove system was developed which is tactile and sensorized. The system has several thin-film flexible and stretchable strain gauges and force sensors to independently measure finger movement and pressure. It provides real time accurate results within minutes and causes minimal discomfort to the wearer.

2.6 Smart Glove to Measure a Grip Force of the Workers [12]

The aim of the project is to analyse the hand for musculoskeletal injuries that are a result of everyday activities or from repetitive loading in poor conditions. The prototype of the device contains 14 connected sensors. An electronics control unit on the wrist performed analog signal processing. The sensor current was amplified. Sensitivity of the sensor can be fine-tuned by manipulating the value of the feedback resistor. 12-bit successive approximation ADCs were used for digitization. Communication to the PC was done through a Bluetooth module.

2.7 Force-Sensing Glove System for Measurement of Hand Forces During Motorbike Riding [13]

The design includes flexi-sensors on a glove, a chipboard, a Bluetooth trans-receiver and a smartphone. Two Flexi-Force A401 sensors were placed on the index and middle fingers to monitor the two-finger braking. Another two sensors were positioned on the thumb and palm to observe the rotation of the throttle. The microcontroller is the MCF51JM128VLH which is a 32-bit RISC device, and an ultra-low power MC. Bluetooth technology was used for real-time data communication. An algorithm was designed to identify the specific hand movements observed during riding a bike. Taking into account the force sensing elements, using the algorithm it was capable of differentiating between the actions of clutch, wheel, brake and throttle.

2.8 Smart Tactile Gloves for Haptic Interaction, Communication and Rehabilitation [14]

These gloves utilise different types of sensors to measure the bending angle, pressure, and/or different orientations of the hand and/or fingers during interaction. The three main categories of smart gloves:

Gesture-based: These use cameras to capture hand movement. The videos are then processed using computer vision to determine function. Heavy cameras cause poor mobility and distress the wearer.

Touch-based: Information is either sent or received by the user through touch using touch sensors.

Gesture and Touch based: Uses a combination of both aforementioned methods.

3 Working/Methodology

The basic block diagram of the proposed method is shown in Fig. 1.

The device is considered to be made of three separate blocks.

1. Sensing and analog signal condition block
2. Continuous force monitoring block
3. Haptic feedback block.

3.1 *Sensing and Analog Signal Condition Block*

Two A101 [15] sensors (sensor specifications in Fig. 2) are placed on the thumb, and the ring finger to continuously monitor the force applied by the hand onto an object held in it. The force sensors are the smallest offered by the firm FlexiForce and are of dimensions 7.6 mm × 15.6 mm. The sensing area is a circular disk of diameter 3.8 mm. The sensor can measure a maximum force of up to 10 pounds. It is a standard piezoresistive force sensor with a typical performance curve.

The challenge is to design an analog signal conditioning system to accurately measure the force applied, with reasonable linearity in the requisite range. A standard potentiometer arrangement is found to be unusable because of extreme nonlinearity

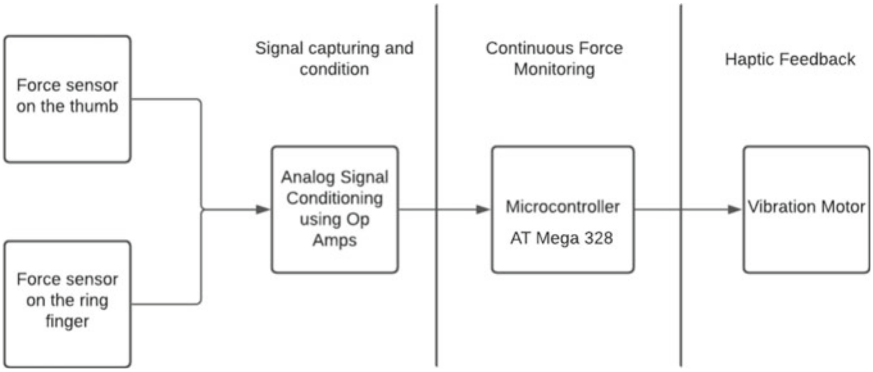

Force sensor on the thumb
Force sensor on the ring finger
Signal capturing and condition
Analog Signal Conditioning using Op Amps
Continuous Force Monitoring
Microcontroller
AT Mega 328
Haptic Feedback
Vibration Motor

Fig. 1 A representation of the basic block diagram of the proposed model

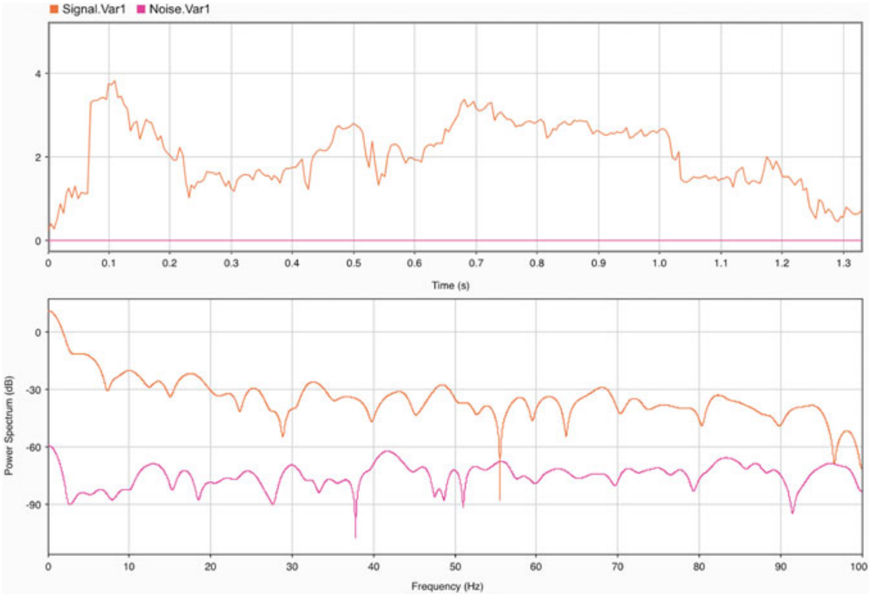


Fig. 2 Signal and noise samples from the ring finger, and power spectrum

in the range. Thus, the given circuit is used. The circuit is a combination of the op amp MCP6004. The op amp is used in the inverting amplifier mode with a capacitor as feedback to perform noise cancellation. The feedback resistance, capacitor value and V reference values are calculated to specific values for the need. Sensor specifications are mentioned in Table 1.

Circuit Design

The value of Vref is maintained as −3.7 V, the output of a 3. 7 V LiPo battery which is used to provide for an isolated supply. The value of the feedback resistor is chosen to be the same value of that of the sensor resistance at maximum force applied when a user grabs onto an object. This ensures that the input to the microcontroller will be about 3.7 V when an object is normally held, ensuring both a leeway of 1.3 V for unpredictable signal swing, as well as sufficient resolution when the force on the object begins to fall.

The capacitor in the feedback acts as a single pole low pass filter of cut off frequency

$$1/2\pi RC \tag{1}$$

The spectrum of noise is studied on an oscilloscope and MATLAB to determine the requisite cut off frequency. The capacitor value is then calculated. An additional single pole LPF is considered at the output of the op amp, but is not implemented

Table 1 Sensor specifications from manufacturer

	Typical performance	Evaluation conditions
Linearity (Error)	< ±3% of full scale	Line drawn from 0 to 50% load
Repeatability	< ±2.5%	Conditioned sensor, 80% of full force applied
Hysteresis	<4.5% of full scale	Conditioned sensor, 80% of full force applied
Drift	<5% per logarithmic time scale	Constant load of 111 N (25 lb)
Response time	<5 µs	Impact load, output recorded on oscilloscope
Operating temperature	−40 to 60 °C (−40 to 140 °F)	Convection and conduction heat sources
Durability	≥3 million actuations	Perpendicular load, room temperature, 22 N (5 lb)
Temperature sensitivity	0.36%/°C (0.2%/°F)	Conductive heating

Source From [15], Tekscan Sensor Data Sheet for Flexiforce A101

since the signal to noise ratio is satisfactory, considering the wires used are of short lengths.

After a reasonable tradeoff between linearity and resolution is achieved, the force sensor can be interfaced to an arduino nano microcontroller board with an ATmega328 processor.

Hardware components used

1. Arduino Nano board with ATmega328 processor
2. MCP6004 Op Amp
3. 33 kΩ Resistors
4. 47 pF and 1 nF Capacitor
5. Two FlexiForce A101 Force Sensors.

MATLAB Signal Analyser Test

An object is held on the glove and the voltage values are polled at a frequency of 500 Hz. The power spectrum of noise as well as the signal when an object is held is plotted and analysed using the Signal Analyser on MATLAB©. The signal to noise ratio on the ring finger is found to be 60.993 dB, and that on the thumb is 50.1913 dB.

SNR calculated using the SNR function in the signal processing toolbox = 60.993 dB. Signal and noise samples obtained from the thumb, and power spectrum are shown in Fig. 3.

SNR calculated using Signal Processing Toolbox in MATLAB = 50.1913 dB.

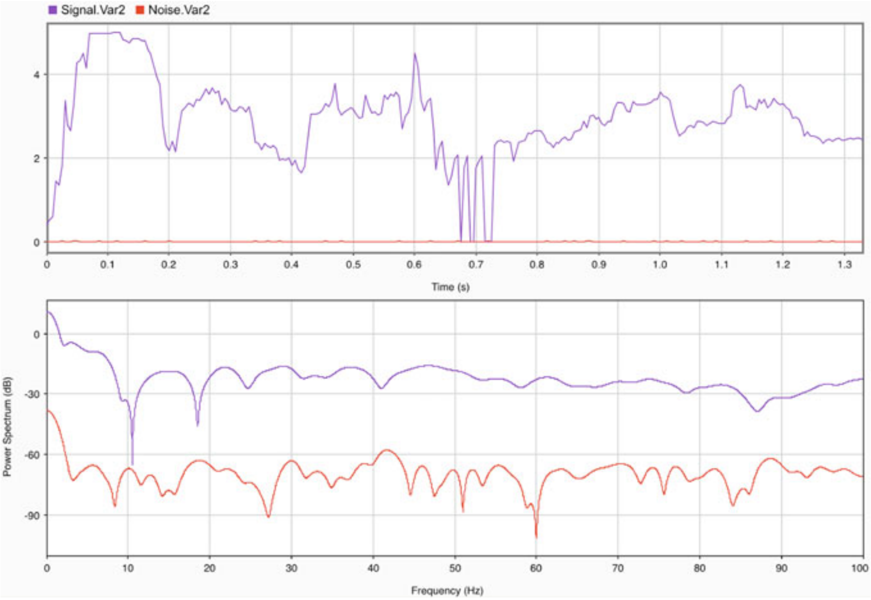


Fig. 3 Signal and noise samples obtained from the thumb, and power spectrum

3.2 Continuous Force Monitoring Block

An algorithm is implemented to detect when the user holds an object, and a 1034 Mobile Phone Vibrator Motor is triggered to vibrate in the upper arm of the user to warn when they are about to drop the object being held. In case the user drops the object, the microcontroller senses the same and stops the vibration. If the user reinforces their grip, and the object is held firmly, the vibration is ceased.

The Arduino Nano has a 10-bit ADC, which maps the input voltage to values between 0 and 1023. Instead of calibrating the force sensor and dealing with values in Newtons, extensive testing is conducted to find the values of the ADC output in three different conditions: reading when no object is held, reading when an object is held firmly, and when the object starts to slip. The values are found to be 10,150 and 80 respectively.

3.3 Haptic Feedback Block

Initially, when the value goes high in both the sensors, the device assumes that an object is being held. The force acting on the object is continuously measured. When the value decreases, the microcontroller triggers the vibration motor to warn the user that they might be losing their grip. If they do reinforce their grip, the vibration is stopped. If the grip continues to fall and the object falls (or is set down), the microcontroller detects the same and stops triggering and goes back to the first state (no object is held) as in Fig. 4 and its performance curve of FSR is shown in Fig. 5.

Two force sensors are used to make sure that the user is actually grabbing onto something and not simply pressing their fingers against a rigid object.

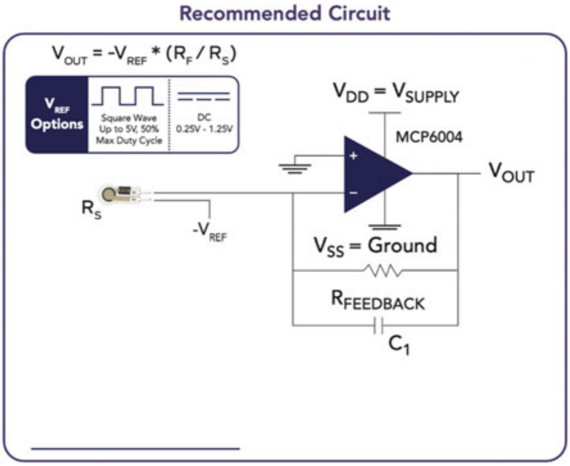


Fig. 4 Circuitry used for analog signal conditioning

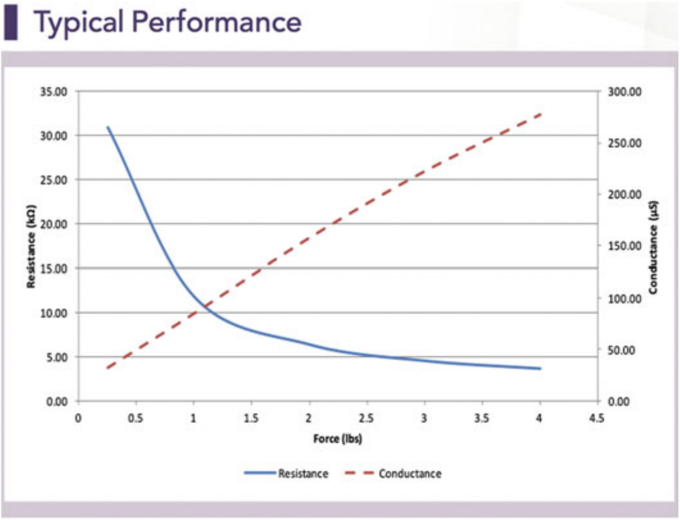


Fig. 5 Typical performance curve of the FSR. *Source* From [15], Tekscan Sensor Data Sheet for Flexiforce A101

4 Design of the Glove

The force sensors are sewn into the thumb and ring finger of the glove as in Fig. 6a–c. The glove is a low-cost product made of wool. The material provided a considerable amount of passive help to the user to maintain their grip. The entire circuitry is placed inside a box, which is placed on top of the wrist. The vibration motor for haptic feedback is placed under the wrist, assuming that the user would have the sense of touch intact here. It could be relocated to another site if the area is affected by the disease. Components of the glove prototype are shown in Fig. 7.

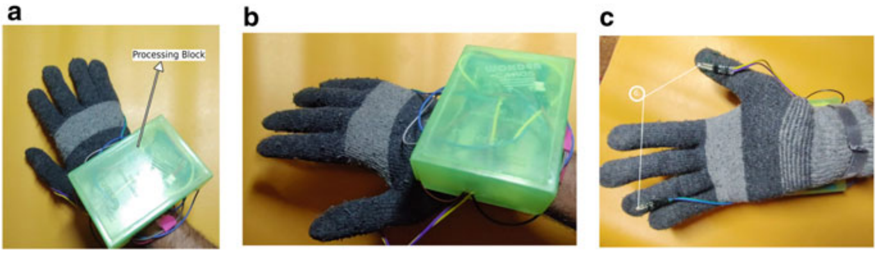


Fig. 6 **a**–**c** Pictures of the proposed glove prototype

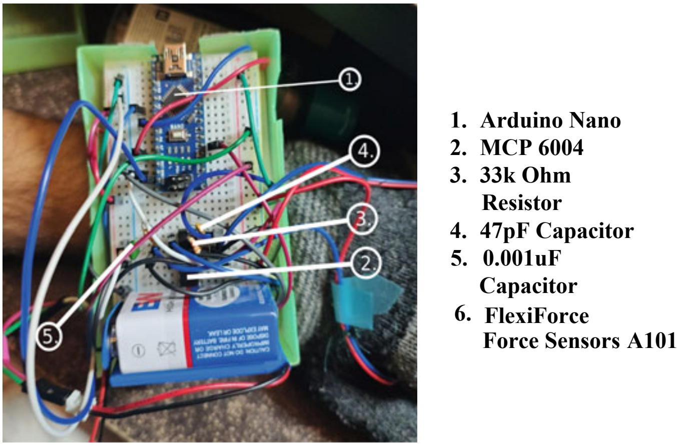


Fig. 7 Components of the glove prototype

5 Comparison

5.1 Smart Muscle Strength Assessment Glove for Rehabilitation Purposes [7]

This is an AI integrated device to classify the strength of the muscle or injury level into 6 different categories. It utilised a collection of 5 force sensors on each finger and a single load cell placed on the palm to capture the strength of the muscle in terms of maximum force applied. The data was then fed to an ANN to perform the classification. The device was mainly used to aid physicians to track how well recovery after a muscle injury is progressing and has no direct feedback to the person affected by the disability or injury.

The proposed device on the other hand is a specific assistive device for people with disabilities followed by CMT. It directly aids the user by helping them monitor their grip level as they grasp objects in daily life. It assists the user to perform daily tasks without having to constantly worry about dropping things. It utilises only two force sensors to detect grip level, along with a microcontroller and vibration motor for direct feedback to the user.

5.2 Calibration and Evaluation of a Force Measurement Glove for Field-Based Monitoring of Manual Wheelchair Users [8]

The project provided an out-of-laboratory sensing device to monitor and store the forces applied on the palm while moving a wheelchair. Four force sensors were knitted into the palm of the glove, with each sensor slightly overlapping over each other. A PCB placed in a 3D printed case sampled the force sensors at 20 Hz and recorded it into a 32 GB microSD card for further processing offline. MATLAB was used to analyse the force values recorded. The performance of the arrangement was compared with Smart Wheel, a wheelchair designed to continuously measure force values exerted on the wheel by the user. The glove signal managed to record 72.7% of the peaks recorded by the Smart Wheel.

The objective of the study is to come up with a comparable measurement device to those used in the laboratory, not to provide a direct application to an individual affected with a disability. The proposed device, though not as accurate as a laboratory device, aims to assist people with CMT.

5.3 A Fabricated Force Glove that Measures Hand Forces During Activities of Daily Living [9]

The objective of the study was to achieve an in-detail quantitative analysis of the force exerted by different parts of the hand when an object is held. The study used fourteen force sensors, on different parts of the thumb and four fingers. The study intended to measure the least amount of force that an actuator needs to apply at every joint when modelling hand prosthetic devices. The study focussed on a quantitative analysis of forces to enable future prosthetics to be designed with these specifications in mind. The study itself did not implement an application for the end user, it rather provided information for further applications to use.

This study is a direct application specifically designed to aid people with CMT, complete with a feedback loop to the user to indicate when they are about to lose grip. However, this study does not dive into the anatomy of the hand as well as the aforementioned study which uses 14 force sensors-providing high precision and accuracy from different parts of each finger. This study uses only two force sensors to quantitatively determine whether the user is about to drop an object or not.

5.4 Wearable System with Embedded Force Sensors for Neurologic Rehabilitation Trainings [10]

The device was designed with the objective of providing electronic measurements of physical parameters to aid in neurological rehabilitation. A glove which did not hinder comfort of day-to-day activities of the wearer was designed. Four force sensors were placed on the fingertips and paired with an ATMega 328 on a Bluno Beetle board—which provided wireless functionality. The analog condition part involved a voltage divider circuit paired with an op amp. To improve the force sensors' ability to uniformly capture forces, a small 3D printed support was placed on either side of the circular sensor surface. This avoided poor measurements due to the soft material of the glove. Although the device implements signal conditioning and capturing very accurately, no end-user application was implemented in the same device. It can be used to externally capture force application patterns in the user's day-to-day activities so that a neurologist would be able to assess the progress of the patient's recovery.

The proposed device aids the end-user by providing for a feedback mechanism directly in the device-along with able signal capturing and conditioning. It directly improves the quality of life of the user.

5.5 Tactile Sensorized Glove for Force and Motion Sensing [11]

The objective of the device was to assess hand related injuries such as thumb tendonitis and carpal tunnel syndrome. Finger movement and pressure was assessed by several thin-film flexible and stretchable strain gauges and force sensors. The project essentially assessed the extent of the aforementioned hand related injuries by conducting a number of tests and evaluating the data collected from these tests. The tests included thumb rotation, bending and reaction. The work is currently limited to the thumb movements only.

This glove system is primarily used to assist patients suffering from CMT. Two force sensors are used to measure the grip strength, an algorithm that automatically checks if the received force value drops below a certain threshold and triggers a vibration motor accordingly. It assists the patient to perform their day-to-day task normally.

5.6 Smart Glove to Measure a Grip Force of the Workers [12]

The aim of the project was to analyse the hand for musculoskeletal injuries. It achieved this by accumulating data and by observing a worker in the long term a study on the impact of repetitive workload on his health can be made. 14 force sensors whose sensitivity can be adjusted by using the feedback resistor, an electronic control system which performs signal processing, a 12-bit ADC and a Bluetooth module were used.

This proposed device assists users suffering from CMT perform their day-to-day activities. The force sensor measures the user's grip strength. This signal from the force sensor is conditioned to accurately measure the force applied, with reasonable linearity in the requisite range. The Arduino Nano uses a 10-bit ADC, which maps the input voltage to values between 0 and 1023. The data is then compared with the defined threshold and if it drops below said threshold, it triggers the vibration motor.

5.7 Force-Sensing Glove System for Measurement of Hand Forces During Motorbike Riding [13]

The paper focused on assessing the hand performance during motorbike riding. It aimed to discriminate user force while braking, steering and applying clutch. The glove system used 2 FlexiForce A401 sensors to simulate the force applied during braking. The microcontroller along with Bluetooth was used to communicate the analysed data to a tablet or monitor. A two-way user interface was developed to establish communication. The algorithm identified the specific hand movements observed during riding a bike.

This system was designed specifically to tackle the day-to-day issues faced by patients diagnosed with CMT. 2 FlexiForce A101 sensors are used to measure the grip strength. A microcontroller along with the algorithm designed, receives signals from these sensors after signal conditioning, compares and sends an output signal to the vibrator motor to notify the patient when their grip strength drops.

Prior literature described either focus on measuring of force applied by different parts of the palm and hand in a laboratory testing, and some do not implement any kind of feedback. A glove designed specifically for patients with CMT, with feedback designed in a way that the haptic block is modular and can easily be placed from one part of the arm to another is not common. The reason behind implementing modularity in the feedback device is to allow the user to switch the placement of the vibration motor wherever they wish to; patients with CMT gradually lose their sense of touch, and thus would require the device to be placed in different parts as the disease progresses.

This proposed device is also tested in a laboratory setting, because any device actually helpful to an end user must hold up in the outside world. Testing has been performed using real world objects like a glass, remote and a bottle, to cover a wide range of weight and size of real-world objects.

6 Testing

The glove is tested in daily life environments while holding different objects. The object is gradually released and the point at which the object dropped is noted. The algorithm is designed to the same values.

Post design, the device is tested to see how well it warns the user. A total of 100 trials are carried out with 30 on a glass, 30 on a metal water bottle and 40 on a remote. These objects are chosen to test the everyday performance of the glove. The results are tabulated in Table 2.

The lowest accuracy is observed on the lightest object, the remote. The heavier objects performed better. The reason behind this is presumed to be the fact that since the dimensions and weight of the object is low, it does not make good contact with the force sensor. A work-around would be to include more force sensors along the finger, albeit with the added cost.

Table 2 Tests conducted with everyday objects

Object	Glass	Water bottle	Remote
Accuracy	28/30 (93%)	27/30 (90%)	32/40 (80%)

The errors are observed due to the smaller size of the force sensor and the fact that sometimes, the actual point of contact between the object and user is not exactly on the force sensor head.

It is also observed that at times, the warning came in earlier than expected. Although from the perspective of a CMT patient, it is better to be warned earlier; the problem is that the user would find it inconvenient to keep holding it harder than they normally would. A quick and straightforward solution would be to invest in a force sensor with a slightly larger head, and to adjust the threshold to the specific users' preference.

7 Conclusion

Charcaot-Marie-Tooth is a disease which causes a deterioration in the quality of life over time. People suffering from the disease must constantly be aware of what they are holding and how well they are gripping it. The proposed device allows the user to go about their daily tasks without worrying as much about dropping things. The device is able to recognize when users' grip is slacking, and when the object would soon fall. They are warned by the vibration motor fixed in the upper arm.

CMT causes degradation of the sensation of touch in different parts of the limbs. Even in the arm, the vibration motor can be placed in different positions, varying from user to user so that they can get maximum haptic feedback. The areas where the force sensors are placed too can be modified from user to user depending on where they have muscle and touch degeneration. Over extended use of the device, users will be more accustomed to trusting the glove to warn them, giving them ease of mind and the ability to engage in social interactions.

A further extension of the project can be an implementation of an exoskeleton which holds onto the object the user is holding, if they do not reinforce their grip, even after the warning from the vibration motor.

Acknowledgements We would like to express our gratitude to Professor Saneesh C T, Associate Professor, Department of Electronics and Communication Engineering for aiding us with the circuit design and testing to achieve an acceptable level of signal conditioning.

References

1. https://charcot-marie-toothnews.com/2019/10/22/engineered-gloves-can-detect-worsening-hand-function-in-people-with-cmt-study-reports/
2. https://www.mayoclinic.org/diseases-conditions/charcot-marie-tooth-disease/symptoms-causes/syc-20350517
3. MATLAB© (2021) 9.7.0.1190202 (R2021b). The MathWorks Inc., Natick, Massachusetts
4. Pareyson D, Marchesi C (2009) Diagnosis, natural history, and management of Charcot-Marie-Tooth disease. Lancet Neurol 8(7):654–667. ISSN 1474-4422

5. Barreto LC, Oliveira FS, Nunes PS et al (2016) Epidemiologic study of Charcot-Marie-Tooth disease: a systematic review. Neuroepidemiology 46(3):157–165. https://doi.org/10.1159/000443706
6. Banchs I, Casasnovas C, Albertí A, De Jorge L, Povedano M, Montero J, Martínez-Matos JA, Volpini V (2009) Diagnosis of Charcot-Marie-Tooth disease. BioMed Res Int (Article ID 985415): 10 p. http://doi.org/10.1155/2009/985415
7. Makableh YF, Ghabashneh E, Harahsheh T, Khwaileh F (2019) Smart muscle strength assessment glove for rehabilitation purposes. In: 2019 IEEE Jordan international joint conference on electrical engineering and information technology (JEEIT), pp 648–652. http://doi.org/10.1109/JEEIT.2019.8717478
8. Anderson A et al (2020) Calibration and evaluation of a force measurement glove for field-based monitoring of manual wheelchair users. In: 2020 IEEE 20th international conference on bioinformatics and bioengineering (BIBE), pp 1004–1007. http://doi.org/10.1109/BIBE50027.2020.00170
9. Austin EF Jr, Kearney CP, Chacon PJ, Winges SA, Acharya P, Choi JW (2022) A fabricated force glove that measures hand forces during activities of daily living. Sensors (Basel) 22(4):1330. Published 2022 Feb 9. http://doi.org/10.3390/s22041330
10. De Pasquale G, Mastrototaro L, Pia L, Burin D (2018) Wearable system with embedded force sensors for neurologic rehabilitation trainings. In: 2018 symposium on design, test, integration and packaging of MEMS and MOEMS (DTIP), pp 1–4. http://doi.org/10.1109/DTIP.2018.8394187
11. Yeo JC, Lee C, Wang Z, Lim CT (2016) Tactile sensorized glove for force and motion sensing. IEEE Sens 2016:1–3. https://doi.org/10.1109/ICSENS.2016.7808596
12. Borik S, Kmecova A, Gasova M, Gaso M (2019) Smart glove to measure a grip force of the workers. In: 2019 42nd international conference on telecommunications and signal processing (TSP), pp 383–388. http://doi.org/10.1109/TSP.2019.8768848
13. Ye Q, Seyedi M, Cai Z, Lai DTH (2015) Force-sensing glove system for measurement of hand forces during motorbike riding. Int J Distrib Sens Netw. https://doi.org/10.1155/2015/545643
14. Ozioko O, Dahiya R (2021) Smart tactile gloves for haptic interaction, communication, and rehabilitation. Adv Intell Syst 4:2100091. https://doi.org/10.1002/aisy.202100091
15. Tekscan. FlexiForce Standard Model A101. Tekscan, Inc., Boston, MA, USA. DS Rec J 062821 Datasheet

Accident Detection in Surveillance Camera

A. P. Adil, M. G. Anandhu, Jeovan Elsa Joy, Twinkle S. Karethara, S. Anjali, and B. R. Poorna

Abstract Road accidents are a major cause of death, and many victims die as a result of not reporting such events to the appropriate authorities. Because the event was not reported, there is a lack of emergency medical assistance, which leads to deaths. A computer vision-based traffic observing and revealing strategy can help with giving health related crises continuously, perhaps saving many individuals. Conventional traffic systems, which are outfitted with IP cameras and sensors, are currently set up all around the city to supervise and control traffic. In this paper, we present a better traffic checking framework that perceives and distinguishes moving items like vehicles, cruisers, etc. in live camera, takes care of, identifies accidents of these moving articles, and promptly sends crisis admonitions to the fitting authorities. An innovative architecture for detecting road accidents is given in this paper. The suggested framework uses YOLO to locate accurate objects, followed by accident detection for surveillance data. The nearest police station is notified of the observed accident. On commonplace street traffic CCTV reconnaissance film, the proposed framework gives a reliable procedure to accomplish a high Detection Rate and a low False Alarm Rate.

Keywords Vehicle detection · Deep learning · Convolutional neural network · Python · Opencv · Computer vision · Yolo V4

1 Introduction

Vehicle accidents are a profoundly critical and high need general wellbeing worry, since insights show that more than 1.25 million individuals pass on because of street crashes each year [1]. Speeding, driving, a perilous vehicle, policing, most urgently,

A. P. Adil (✉) · M. G. Anandhu · J. E. Joy · T. S. Karethara · S. Anjali · B. R. Poorna
Mar Baselios College of Engineering and Technology, Mar Ivanios College Rd, Nalanchira, Trivandrum, Kerala 695015, India
e-mail: adilap2000@gmail.com

S. Anjali
e-mail: anjali.s@mbcet.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_26

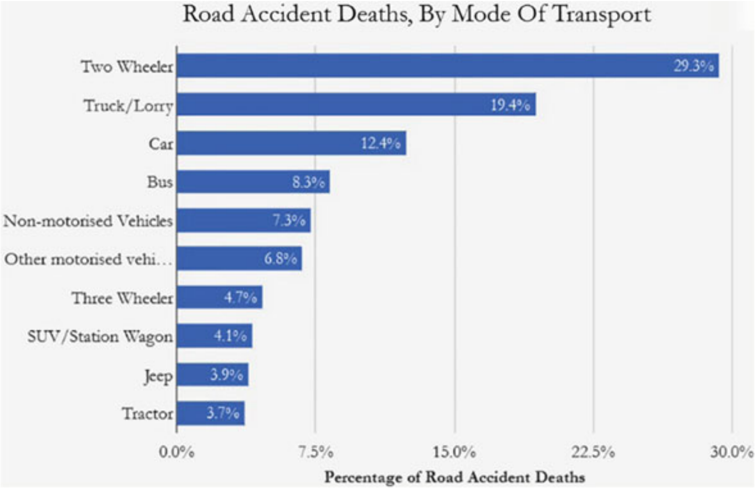


Fig. 1 Graphical representation of reported cases of road accidents by modes of transport

lacking post-crash crisis care are all chance elements. Speeding, driving, a perilous vehicle, policing, most urgently, lacking post-crash crisis care are all chance elements. Ease the number of accidents. We can make our systems smarter and more efficient with advances in Artificial Intelligence and Machine Learning. Computer vision is the investigation of how computers can repeat the human cerebrum. A part of man-made reasoning gathers and deciphers data from computerized photographs. The vehicular populace is becoming quicker than the economy and populace. Accidents and passings on the road, particularly including bikes, are expanding at a disturbing rate. Most mishap passings happen inferable from an absence of speedy clinical treatment on courses like express motorways. An office for conveying brief clinical treatment to the mishap site can assist with forestalling setbacks. Subsequently, the idea of a ready framework emerges, which recognizes the mishap and its seriousness to inform a close by place for the sending of a rescue vehicle or clinical guide at the mishap area. Here, CCTV will decide if an accident happened and distinguish the incident. Following the choice of mishaps, the framework will send a caution to specialists, police headquarters, and advise them of the incident [2]. The salvage group can get to the area right away. The captured image will be emailed to the rescue workers via the technology. Yolo V4 is used for detection. Graphical representation of reported cases of road accidents by modes of transport are shown in Fig. 1.

2 Literature Review

CCTV observation camera creating Spatio-Temporal Video Volumes (STVVs) Autoencoders [3] produce an oddity score while at the same time recognizing moving articles, following the items, and afterward tracking down the convergence of their tracks. This approach may be useful in determining vehicle accidents at crossings with typical traffic flow and adequate lighting. The disadvantages include evaluating incidents under low visibility situations, considerable occlusions in vehicle accidents, and large fluctuations in traffic patterns.

Farneback For motion detection, optical flow was employed, and for accident detection, a statistic technique was applied [4]. The purpose of threshold alterations is to respond to changes in illumination and traffic congestion. The disadvantage is that only highway and expressway traffic patterns are considered. The database was too tiny to make a comparison. The time it takes to process video is longer.

Another work is based on LSTM (Long Short-term memory) using automated systems to detect unusual events [5]. This methodology leads to better security and broader surveillance. The drawbacks are that they take long time to train and require more memory to train. The dropout is much harder to implement. It is sensitive to different random weight initialization.

Gaussian Mixture Model (GMM) recognize vehicles and afterward vehicles are followed utilizing the mean shift calculation. This calculation handles impediments during mishaps very well [6]. The downsides are their dependence on restricted boundaries in situations where there are unpredictable changes in rush hour gridlock design and extreme weather patterns.

Mask RCNN performs well in accident detection. It detects position and type of target objects [7]. It also extends into field of semantic segmentation. The false alarm rate is decreased here. The disadvantages are that computation time is high and the accuracy is less.

3 Problem Statement

Traditional traffic monitoring systems are simply meant to monitor or regulate traffic, but they do not give any solution to reduce the deadly unintentional human damage rate that occurs due to a lack of medical treatment in real time. Consider the following scenario: an accident occurs but no one is present to report it; the sufferer is in critical condition and every second counts; any delay might result in disability or death. We cannot completely eliminate accidents, but we can make improvements. providing post accidental care just-in-time. The existing strategies for creating criteria for accident detection are ineffective. The false alarm rate has grown, and they do not work well in poor weather situations, such as limited visibility.

4 Implementation

The proposed system given in Fig. 2. We represent how the structure is acknowledged to perceive vehicular impacts. Our fundamental objective is to foster a straightforward yet powerful strategy for identifying car crashes that can work rapidly and convey basic data to specialists.

The proposed framework realizes its intended purpose via the following stages:

Vehicle Detection
Accident Detection
Send Alert and Location to Rescue team.

The strategy is planned by getting data from a picture/video source, like a CCTV framework or other tantamount. The arrangement of the result from the source will be either .mp4 or .avi [8]. OpenCV, which has implicit capacities for this, will be utilized to take care of the contribution to the application. The main issue is that the handling will be done in BGR variety space, which might be settled by changing the video's variety space a while later. If there are any noise or undesired artifacts in the video stream, they will be removed. This may be accomplished with the use of filters such as average filters. This data generated will be passed into an object detector, which will look for vehicles in the data. We adopted YOLOv4 as the detector for the system because it is extremely fast compared to other detectors like Faster R-CNN and Mask R-CNN [9] while maintaining high accuracy, making it a great choice for real-world applications.

COCO Dataset

MS COCO is an enormous scope dataset for object acknowledgment, division, central issue discovery, and captioning [10]. The dataset contains 328K pictures. There are 164K pictures altogether, isolated into three sets: preparing (83K), approval (41K), and test (41K). In 2015, a new test set of 81K pictures was delivered, which incorporated all past test pictures as well as 40K new pictures. In light of local area input, the preparation/approval split was altered from 83K/41K to 118K/5K in 2017.

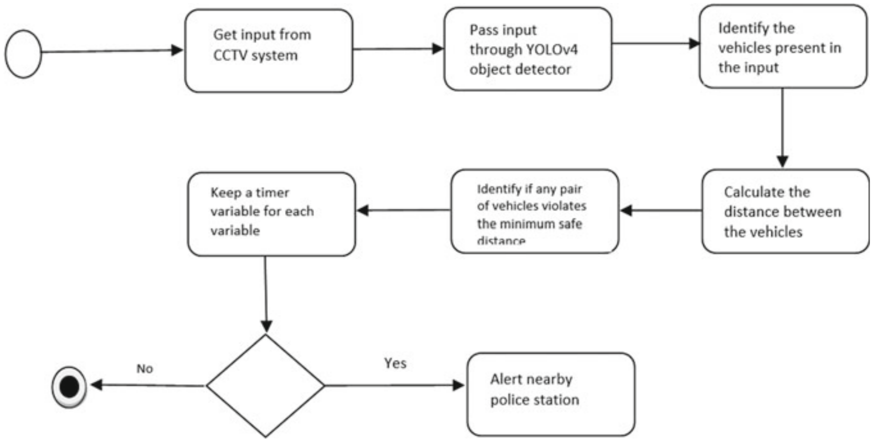

Get input from CCTV system
Pass input through YOLOv4 object detector
Identify the vehicles present in the input
Calculate the distance between the vehicles
Identify if any pair of vehicles violates the minimum safe distance
Keep a timer variable for each variable
No
Yes
Alert nearby police station

Fig. 2 Architectural diagram

The pictures and comments are safeguarded in the new division. The photographs and comments are safeguarded in the new division. The 2017 test set is a subset of the 2015 test set's 41K pictures. A new unannotated dataset of 123K pictures is remembered for the 2017 variant. an enormous scope dataset for object distinguishing proof, division, and inscribing There are north of 200,000 pictures of the all-out 330,000 pictures are marked which are true information.

4.1 *Vehicle Detection*

YOLO: **You Only Look Once**

The Yolo approach employs a neural network to evaluate an image. The picture is partitioned into S × S grids and has bounding boxes [11]. The working of the YOLO algorithm are shown in Fig. 3. There are 24 convolutional layers in this approach, followed by two completely connected layers. By alternating 1 × 1 convolutional layers with preceding layers, the feature space is decreased. The object identification problem is a prediction model with the purpose of separating geographically bounded boxes and identifying related classes inside the bounding boxes. In just one evaluation, a single neural network can predict the bounding boxes and class probabilities from the input images.

In this project, we will use the yolo algorithm to recognize things in a live feed or an image. Because Yolo's operation is based on regression, it is simple. Yolo

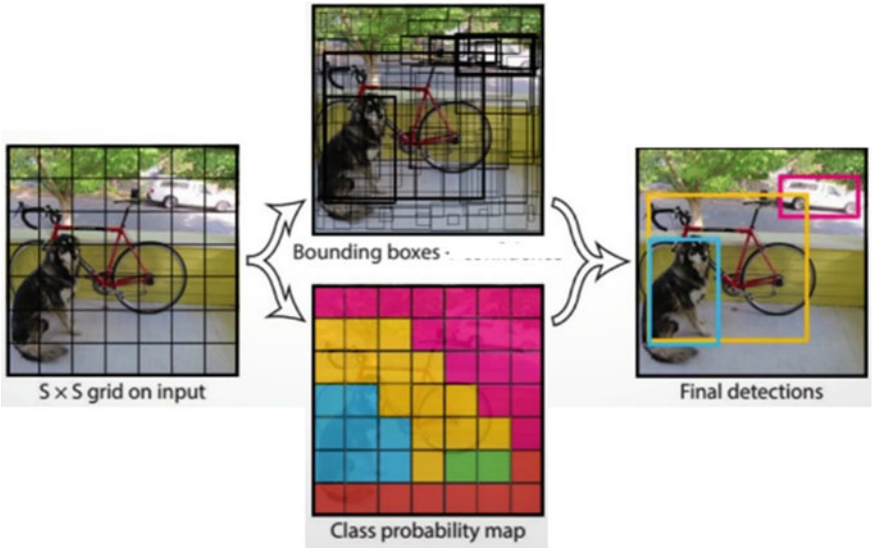


Fig. 3 Working of YOLO

anticipates the class and bounding boxes for the entire image in a single run of the algorithm, unlike CNN, which selects interesting areas of an image.

To use this technique, we need to know what we're trying to predict, i.e., the objects we're likely to be interested in, so we can train our algorithm to look for object classes and bounding boxes.

4.2 Accident Detection

Accident detection is accomplished by CCTV cameras [12], which provide video footage to the system. The framework will switch such video over completely to casings and feed them individually to YOLO for recognition. For Accident Detection, all vehicles seen in a casing are saved and verified whether they cross-over.

The process is started by passing it through the object detector. As we have already discussed, the detector we used here is YOLOv4. YOLOv4 is pre-prepared to perceive very nearly 83 thing groupings, but we are just inspired by vehicles spotted within the supplied input frame at any given time. Thus, in the accompanying stage, we slender down the discovery to just those that have the name 'vehicle.'

We happen to the following stage provided that there are in excess of zero location. The recognition yield incorporates the x and y arrangements, as well as the width and level, of the midpoint of the bouncing box conformed to the perceived vehicle. We get the directions of the jumping box's corner from these focuses since we'll require them later to vary the bouncing box around the recognized vehicles, which will be green assuming that there's no accident and red in the event that there is.

The distance between the vehicles is then determined utilizing the Euclidean Distance metric by deducting the x and y headings of the midpoint of the skipping boxes of the vehicles in thought. The following stage is to check whether the distance is not exactly the suggested least safe distance. For this situation, the base safe distance edge is definitely not a general amount and will differ in light of the circumstance where the framework will be utilized. This turnover happens when we utilize Euclidean Distance to gauge the distance between the vehicles, we are really finding the absolute number of pixels between them, which will move.in perspective on the video quality, the region of the camera at the hour of catch, and various components. When the framework has been arranged and the distance edge is still up in the air. Further accident location will happen naturally. On the off chance that the accident is distinguished, it sends an email caution to the specialists with the area of the event. We have previously resolved the fundamental directions for drawing the rectangular box, and we have entered these bearings into a limit given by the OpenCV library, alongside different boundaries, for example, line thickness, line tone, whether it is red for crash and green for no recognized accident.

The output window also displays a test at the top corner whether an accident is detected or not.

4.3 Alert System

Alert after accident detection is sent through email with the help of STMP protocol [13]. For sending 'get' requests, we use Python's requests module, and for sending email notifications, we use Python's 'smtplib' package. The Simple Mail Transfer Protocol (SMTP) is a set of protocols for sending electronic mail over the internet. It can send a single message to one or many recipients. The authorities are notified. When an accident is recognized concurrently for roughly 20 frames, an email is sent out with the location and an image of the accident site.

5 Results

Certain aspects should be addressed before discussing the findings. The typical casings each second, or FPS, which alludes to the quantity of edges handled in one moment, was used to access the framework. This is on the grounds that in true circumstances, we gave outline rate more significance contrasted with different measures on the grounds that relying on the edge rate by which the info video is being handled, the result can be either quickly enough of genuine use case circumstances or be slow sufficient that it is unimaginable to expect to involve it for getting continuous output [14]. One more variable to consider is the framework which is being utilized for running the venture. The rate of processing movies differs significantly depending on the system's CPU, the availability of accelerators like GPUs, and the type of GPUs utilized0, with more powerful GPUs being the preferable option. Table 1 shows the setup of the system we utilized for testing.

We were able to attain an Average Frame Rate of 64 FPS using a system with the following specifications, which is outstanding for real-time processing when compared to the standard of 30 FPS for real-time processing. Figure 4 shows the detection of accidents from the validated results.

Table 1 Specification of system used for testing and evaluation

Processor	Intel i5-10300H
Number of processing cores	4
RAM	16 GB DDR4 3200 MHz
GPU	NVIDIA RTX 3060
Number of CUDA cores	3840

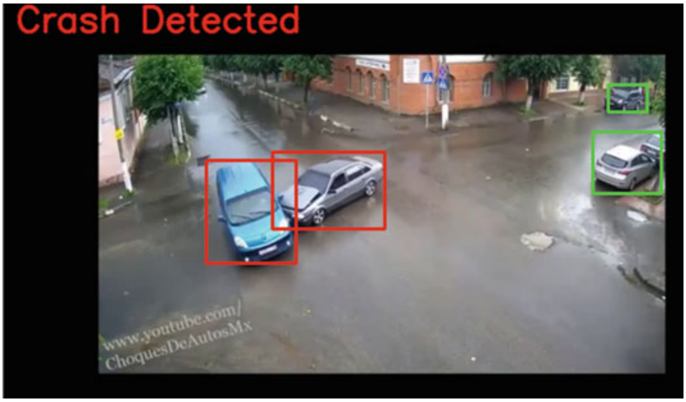


Fig. 4 Accident detected

6 Conclusion

After recognition, the framework will send the mishap picture alongside the mishap area to the Rescue group (Police Station) through email. Further the group will illuminate the close by clinic one the premise of the seriousness of the mishap. The framework productively worked with invariant lighting and camera area conditions and camera quality. The proposed framework is quicker than other item identification strategies and predicts the article better compared to other item discovery calculations, for example, Faster-CNN, Fast CNN or Mask RCNN. The info can likewise be enhanced and give improved results. The proposed framework will assist with keeping away from any defer in distinguishing and giving crisis care which can prompt the expanded seriousness of the mishap to a degree. This architecture has been proven to be effective, paving the way for the creation of real-time general-purpose vehicle accident detection systems. The proposed framework is quicker than other item discovery strategies and predicts the article better than other item identification calculations, for example, Faster RCNN, Fast RCNN or Mask RCNN. The information can likewise be advanced and give improved results. The proposed framework will assist with staying away from any postponement in recognizing and giving crisis care which can prompt the increased seriousness of the mishap to a broader extent. This design has been demonstrated to be powerful, preparing for the making of continuous broadly useful vehicle mishap identification frameworks.

References

1. WHO. Global status report on road safety. Accessed: 2018 [Online]. Available: https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/
2. Wang Y, Zhang D, Liu Y, Dai B, Lee LH (2018) Enhancing transportation systems via deep

learning: a survey. Transp Res Part C Emerg Technol 99(2019):144–163

3. Singh D, Mohan CK (2019) Deep spatio-temporal representation for detection of road accidents using stacked autoencoder. IEEE Trans Intell Transp Syst 20(3):879–887
4. Asimenia D (2017) Adversarial autoencoders for anomalous event detection in images. Ph.D. thesis
5. Mahdyar R, Enver S, Moin N, Nicu S (2019) Training adversarial discriminators for cross-channel abnormal event detection in crowds. In: Conference on applications of computer vision (WACV). IEEE, pp 1896–19
6. Ramos S, Gehrig S, Pinggera P, Franke U, Rother C (2017) Detecting unexpected obstacles for self-driving cars: fusing deep learning and geometric modeling. In: Proceedings of IEEE intelligent vehicles symposium (IV), pp 1025–1032
7. Ijjina E, Chand D, Gupta S, Goutham K (2019) Computer vision-based accident detection in traffic surveillance. In: 2019 10th international conference on computing, communication and networking technologies (ICCCNT), pp 1–6
8. Zhang S, Chen J, Lyu F, Cheng N, Shi W, Shen X (2018) Vehicular communication networks in the automated driving era. IEEE Commun Mag 56(9):26–32
9. Wu G, Zhu X (2019) Using the visual intervention influence of pavement markings for rutting mitigation
10. https://cocodataset.org/
11. Redmon J, Divvala S, Girshich R, Farhadi A (2016) You only look once: unified real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 779–788
12. Robles S (2021) Automatic detection of traffic accidents from video using deep learning techniques
13. Gour D, Kanshar A (2019) Automated AI based road traffic accident alert system: YOLO algorithm. Int J Sci Technol Res 8
14. Qu T, Zhang Q, Sun S (2017) Vehicle detection from high-resolution aerial images using spatial pyramid pooling-based deep convolutional neural networks. Multimedia Tools Appl 76(20):21651–21663

Wheeled Robots for Isolation Wards

U. Sahana and N. Rajesh

Abstract Today's confined individuals and patients who have been infected by viruses can profit from robotic and IoT technology. The world has recently been affected by the Covid-19 epidemic. The virus-affected and quarantined individuals feel helpless since caregivers, medical professionals, and other individuals are scared of the dangerous sickness. This project will produce a robotic IoT agent that will help nurses and doctors monitor patient conditions and deliver meals and medications to them. A robot with a camera transmits data to a remote server through video; the droid cam app will be utilized for this. Sensors assist in retrieving body temperature from a remote location; a machine learning model trained on the image of the patient detects them and maintains the patient's name and body temperature in a centralized database.

Keywords Artificial intelligence · Face recognition · Robotics · Internet of things

1 Introduction

To prevent medical staff and doctors from contracting COVID-19, this theory proposes the use of robotic agents to carry meals and medications to COVID-19 patients or persons suffering from the disease. It continues to be difficult for hospitals to feed and treat people who have the coronavirus. People therefore rely on machines to help them. Food and medications can be delivered to the desired position using a fixed tray that is part of the robot arms. The equipment can be controlled remotely using the Blynk app. To measure the body temperatures of the patients, the robot also has an infrared temperature sensor. A robot is a device that can be programmed by a computer and is equipped to carry out a broad range of autonomous functions. An

U. Sahana (✉) · N. Rajesh
Department of Information Science and Engineering, NIE, Mysuru, Karnataka, India
e-mail: sahanajankal1998@gmail.com

N. Rajesh
e-mail: rajeshn@nie.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_27

external control device or an internal control mechanism can both be used to control a robot. They may also be autonomous or partially autonomous. The bulk of robots are task-oriented machines that put basic usefulness over expressive aesthetics, despite the fact that some of them are made to resemble humans.

Bio-parameters are measurements of biological factors in human body. Temperature of the body of humans, pulse, heart rate, and blood pressure are examples of bio-parameters. Temperature is the bio-parameter we're looking at. An infrared temperature sensor will be used to measure this temperature. Facial recognition is a form of computer software that recognizes people's faces in digital photos. Face identification software works by applying machine learning and algorithmic computations to discover human faces within larger photographs. Other than faces, these larger photographs could show landscapes, buildings, and human body parts, among other things (e.g., shoulders, arms and legs). Detecting face is a technique for recognizing or verifying a person based on his or her appearance in a picture, video, or other audiovisual representation. Commonly, this code is used to log into a programme, system, or service. A person's facial biometric pattern and data are examined as part of the biometric identification process to confirm their identity. The method captures a set of distinctive biometric information related to a person's face and expression in order to recognize, confirm, and/or authenticate them. Blynk is an easy-to-use android app for controlling and communicating with development boards, it has various widgets support with which user interface could be easily created and helps to mask digital and virtual pins with real time development board. Over various media such as Wi-Fi, USB, Bluetooth helps to control actuators from remote place.

2 Literature Survey

Jyoti et al. [1] shows a wireless gesture control robot for people with physical disabilities who are also virus infected. It may function remotely using data and track the movements of the hand or other organs. Some of the important features include Covid 19, Robot, Wireless Technology, and Gesture Controlled System. Antony et al. [2] explained the creation of a medical application-specific remotely controlled autonomous guided vehicle. By detecting temperature, this also helps with health monitoring. Hospital, drugs, monitoring, waste, COVID-19, automatic guided vehicle are things to keep in mind (AGV) Pavithra et al. [3]. The phone and the car are connected via wireless technology. By tapping or pressing on the screen of an Android phone, a manipulator may send commands to the Arduino microcontroller in the car through Wi-Fi and watch as actuators carry them out in real time.

3 Methodology

The several units represented in the proposed project are shown in the following diagram. The procedure makes use of a variety of items, such as Chassis with wheels and motors (Fig. 1).

- Motors are driven by a driver circuit
- Tray to carry items
- Temperature sensors
- Blynk app is to control robot orientation
- ESP32 Module is to operate and coordinate numerous elements
- Smart phone and droid cam app
- Batteries to power the robot module
- Web server to stream footage recorded via Droid cam app.

Procedure

Step 1: Turn on the robot and control units.
Step 2: Enable Wi-Fi.
Step 3: Check the status LED to see if the ESP32 module is connected to Wi-Fi.
Step 4: Open the Blynk application.
Step 5: Activate the Robot's movements by pressing the Blynk app's Start button.

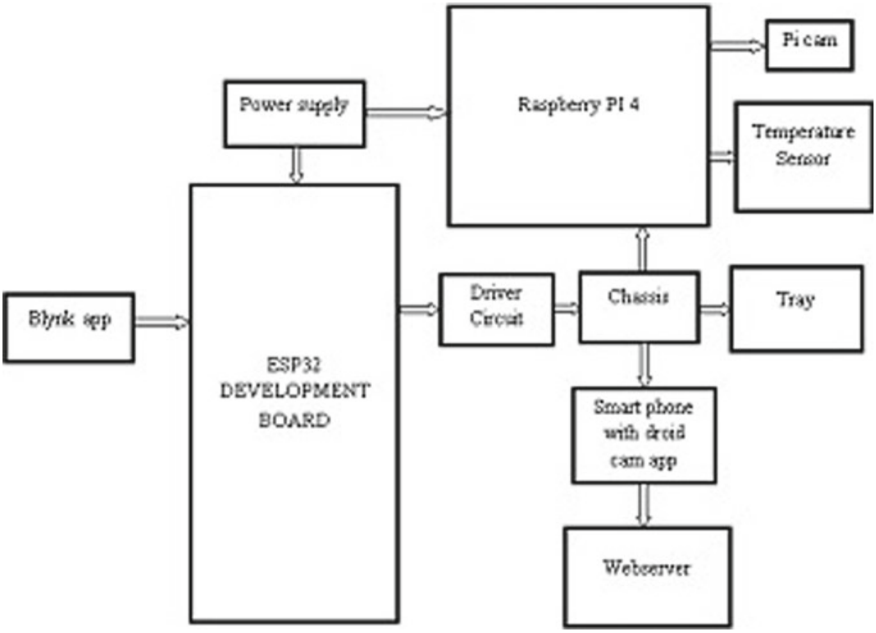


Fig. 1 Architecture diagram

Step 6: Use the Blynk app's other buttons to control the direction.
Step 7: Open a web browser and input the IP address given by the droid Cam software to view the video being broadcast over a smart phone.
Step 8: Check temperature of the person in the log file.

Algorithm for Face Recognition

See Fig. 2.

Skin Detection Algorithm

To detect the face in a digital image, a skin detection method is used. The quality of the image has a direct relationship with recognition. There are a variety of methods for extracting the facial portion of a shot, including simple and challenging methods such as edge based, geometry based, and with a controlled background. We extract the facial region of the supplied input image using one of the ways. As shown in Figure, the input image can be made up of a variety of colors and is not color specific. In this method, color information will be crucial in recovering the photos. In recent days, this method has gotten a lot of attention. When image segmentation identifies the diseased area, analysis can begin. The image may be used to extract a lot of information. The lowering of dimensionality is an important step to take. In order to eliminate model ambiguity and conflict. It's also crucial to think about all the vital details and prevent making any mistakes. One of the most important processes in machine learning is feature extraction. The importance of extracting fundamental traits cannot be overstated. To avoid overfitting and underfitting, features must be chosen carefully. Various visual properties, such as randomness, mean, entropy, and the standard deviation of the colored image, are extracted in this research [4–9].

Morphological Operations

To fill the gaps in the input image, we must conduct morphological procedures. The major morphological procedures are dilation and erosion8. Both techniques are aimed towards pixel processing. The greatest value is dealt with by dilation, whilst the minimum value is dealt with by erosion. We get bounding box for the provided input image using binary distance technique, which restricts recognizing the image's boundary. Face can be tracked using the limitations, and it is projected in the shape of a square in the given image.

Grey Level Co-occurrences Matrix for Face Detection

The next step is to extract the face from the image and determine who the individual is. One of the first methods for extracting texture features was GLCM9. The attributes of the face will be taken from the extracted faces. Following detection, the extracted attributes are projected in the shape of a square in the group shot.

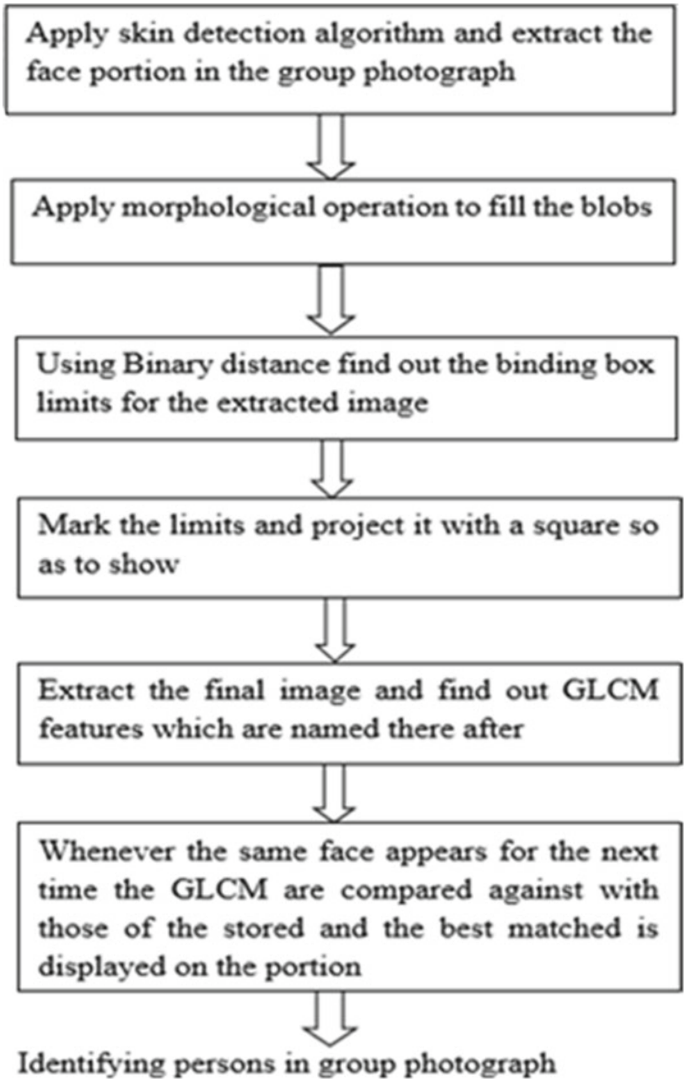

Apply skin detection algorithm and extract the face portion in the group photograph
Apply morphological operation to fill the blobs
Using Binary distance find out the binding box limits for the extracted image
Mark the limits and project it with a square so as to show
Extract the final image and find out GLCM features which are named there after
Whenever the same face appears for the next time the GLCM are compared against with those of the stored and the best matched is displayed on the portion
Identifying persons in group photograph

Fig. 2 Flow chart for face recognition

4 Technology Overview

The face-based car security and ignition system may have trouble with facial recognition. In order to offer a secure environment for starting and exiting the car, a Haar-like element was used. Its purpose was to recognize and detect the authenticated client's essence. Face identification and highlight extraction using a recognition technique could be a remarkable fundamental indication of human faces. Ada-help computation is used to distinguish the face.

A small number of weak classifiers are combined using ada-boost learning to create a strong classifier that can determine whether a given image is a face. The chosen faces are then commonly perceived by comparing the present face's Haar-Classifer to the HaarClassifer of identifiable individuals, as calculated using Haar Classifier. Haar Cascade is an artificial intelligence (AI) item placement computation that can be used to recognize objects in a photo or video.

In order to provide a secure environment for stepping into and out of the car, a haar-like component was used to recognise and detect the essence of the confirmed client. The following is an example of a typical rectangular haar-like element (Fig. 3).

It must first detect the face and then it must label a rectangle for the face discovered, as well as a grayscale image. Before beginning the face recognition process, it is necessary to finish training with a number of different photographs and should be saved in a yml file. The face recognized in the web camera is compared to the training photographs saved in the yml file and popped up with a perfect match. When the face is recognized as that of a specific person, a signal is sent to the microcontroller.

Training an AdaBoost:

- Used the example images (x (1), y 1)…, (x (n), y n) where y 1 = 0, 1 for negative and positive cases.
- Load weights for y 1 = 0, 1 are w (1, i) = 1/(2 m), 1/2 l, where m and l are the number of positive and negative examples, respectively.
- For t = 1,…, T:

  (1) Normalize the weights, $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$

  (2) Choose the best weak classifier based on the weighted error:

$$\varepsilon_t = \min_{f,\, p,\, \theta} \sum_i w_i |h(x_i,\ f,\ p,\ \theta) - y_i|$$

  (3) Describe $h_t(x,\ f_t,\ p_t,\ \mathrm{P}_t)$ where f_t, p_t and θ_t are the reducers of ε_t.

  (4) Upgrade the weights:

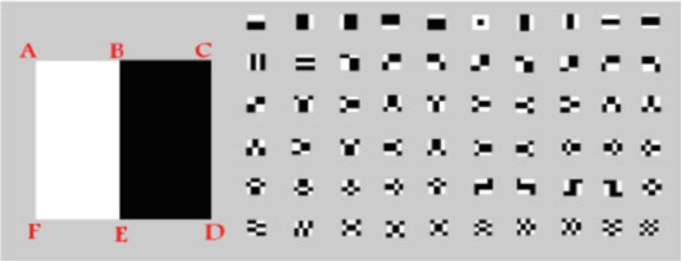


Fig. 3 Haar-like features

$$w_{t+1,\,i} = w_{t,\,i}\beta^{1-e_i}$$

where e_i = zero if instance x_i is classified precisely and $e_i = 1$ otherwise, and $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$

- The final strong classifier is:

$$c(x) = \begin{cases} 1 \; if \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 \qquad otherwise \end{cases}$$

where $x\alpha_t = \log \frac{1}{\beta_t}$

Training the Haar cascade:

The Minimum Acceptable Detection Rate per Layer and the Maximum Acceptable False Positive Rate are both set to d and f, respectively, when training the Haar cascade.

- The goal value for the overall false positive rate, F target, is used by the user.
- P is a collection of success stories.
- Q is a list of negative instances.
- $F_0 = 1.0$; $D_0 = 1.0$
- $i = 0$
- While $F_i > F_{target}$
  - i ← i + 1
  - $n_i = 0$; F_{i-1}
  - While $F_i > f \times F_{i-1}$

$n_i \leftarrow n_i + 1$

Utilize P and Q to teach an AdaBoost classifier with n_i capabilities.
To determine F_i and D_i, evaluate a modern cascaded classifier on the validation set.

- Q ← ∅
- If $F_i > F_{target}$, Compare the findings of the cascaded detector to the collection of non-face photos, and add any false positives to the set Q.

Hardware Components Used

ESP32, which is integrated with 802.11b/g/n Wi-Fi and Bluetooth in dual mode (BT). Jumper wires is used for connecting purposes. The battery with cap is used to power BO motors so they may continue to move. Additionally, it serves as a power supply for the submersible motor that activates the sanitizer dispenser. It can hold 9 V. BO Motors are 100 RPM DC motors that aid in moving the ROBOT model in the desired directions. Wireless temperature sensor MLX90614, it is used to measure body temperature of the patients, it is interfaced to ESP32 board, soon after face of patient is scanned using pi cam, temperature is displayed on the monitor. The Raspberry Pi 3 Model B+ has been used with a camera module attached to it (Figs. 4, 5 and 6).

Fig. 4 Robot

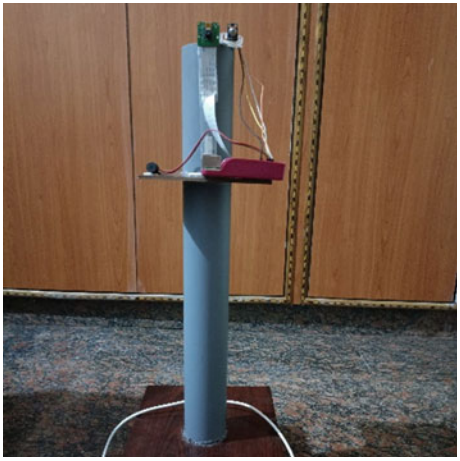
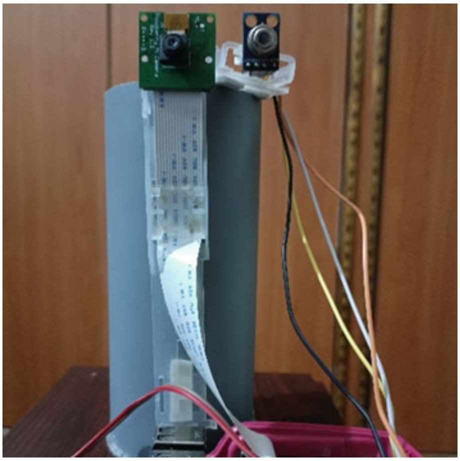
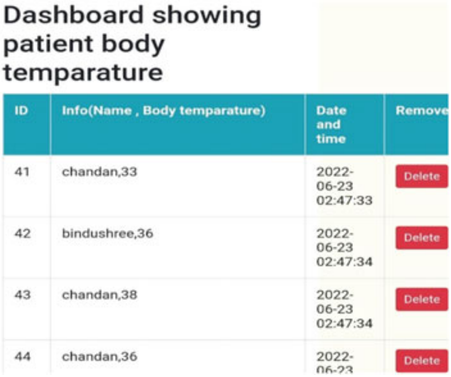


Fig. 5 Webcam and temperature sensor





Dashboard showing patient body temparature

ID	Info(Name , Body temparature)	Date and time	Remove
41	chandan,33	2022-06-23 02:47:33	Delete
42	bindushree,36	2022-06-23 02:47:34	Delete
43	chandan,38	2022-06-23 02:47:34	Delete
44	chandan,36	2022-06-23	Delete

Fig. 6 Dashboard showing body temperature of patients

5 Results and Discussion

6 Conclusion

The proposed model is an attempt to leverage the robot and internet of things concepts to create an innovative product that is useful and useful to society to tackle situations like Covid 19, and it is aimed at frontline health warriors.

The prototype has been evaluated for a variety of test scenarios and has functioned well. However, it does have some limitations, which are stated in the document's drawbacks section.

The prototype could be further enhanced and made usable in real-time applications by considering the restrictions indicated in the preceding sections.

Applications

- It can be used in isolation wards, intensive care units, and quarantine centers to keep track of patients or infected people from afar.
- Helpful for supplying food and medicine to pandemic victims or suspects with little or no human intervention.
- It allows for the retrieval of a patient's temperature from a remote location without the need for human intervention.
- A machine learning model obtains the patient's identity and enters the patient's name and body temperature into a database after scanning an image through a camera.

References

1. Akhund TMNU, Jyoti WB, Siddik MAB (2020) IoT based low-cost robotic agent design for disabled and covid-19 virus affected people. In: 2020 Fourth world conference on smart trends in systems, security and sustainability (WorldS4)
2. Antony M, Parameswaram M, Mathew N, Sajithkumar VS, Joseph J, Jacob CM (2020) Design and implementation of automatic guided vehicle for hospital application. In: Proceeding of the fifth international conference on communication and electronics systems (ICCES 2020)
3. Kalaiarasi D, Pavithra S, Pratheeba S, Priyaadharshini RL (2018) IoT based motion control system of a robotic car. Int Res J Eng Technol (IRJET) 05(03). e-ISSN: 2395-0056
4. Durga Devi S, Mounasri P, Mounika S (2017) IoT based remote controlled robot using android. IJESC 7(3)
5. Shyla R, Megha Shree AC (2017) Intelligent wheel chair based on internet of things. Int Res J Eng Technol (IRJET) 04(08). e-ISSN: 2395-0056
6. Duseja L, Deshmukh Y, Karmuse S, Ohol SS (2021) Autonomous RFID controlled assisting robot for isolation wards. In: Proceedings of 6th international conference on intelligent technologies (ICIT-2021), vol 17, p 19
7. Prabhakar M, Paulraj V, Dhanraj JA, Nagarajan S, Kannappan DAK, Hariharan A (2020) Design and simulation of an automated guided vehicle through webots for isolated COVID-19 patients in hospitals. In: 2020 IEEE 4th conference on information and communication technology (CICT). IEEE, pp 1–5
8. Choi HK, Cui C, Seok H, Bae J-Y, Jeon JH, Lee GE, Choi WS, Park M-S, Park DW (2022) Feasibility of ultraviolet light-emitting diode irradiation robot for terminal decontamination of coronavirus disease 2019 (COVID-19) patient rooms. Infect Control Hosp Epidemiol 43(2):232–237
9. Yogesh S, Prasanna B, Parthasarathi S, Ganesh MA (2020) Open MV-micro python based DIY low cost service robot in quarantine facility of COVID-19 Patients. In: Congress on intelligent systems. Springer, Singapore, pp 519–531

A Survey on Various Crypto-steganography Techniques for Real-Time Images

R. Tanya Bindu and T. Kavitha

Abstract Information is the wealth of every organization and in the modern-day when information is shared digitally and via the internet, protecting this treasure has become a top issue. Private photographs need to be protected from unwanted access due to security concerns raised by internet photo transfers. Nowadays, practically everyone shares their personal information online, including photographs, either with other users or in a database that attracts cyber criminals who can use it to their benefit. Steganography can be as a security tool to safely transmit secret information because it is one such technique where the presence of a confidential message cannot be detected. This article compares various steganography techniques, including AES, LSB, DCT, DWT, etc. are compared them with each other and also more advanced techniques involving Cryptography and Steganography i.e., sharing secret data using counting-based s and matrix-based to increase security.

Keywords Steganography · Cryptography · Least significant bit · Advanced encryption standard · Discrete cosine transform · Discrete wavelet transform · Counting based secret sharing · Matrix based secret sharing

1 Introduction

Image processing is translating a physical image to a digital format and performing different operations on it in order to improve the image or extract confidential information that has to be delivered to a particular recipient. It's a form of distributed signal in which the input is an image, such as a video frame or the picture itself, and the output is another image or image-related features. Graphic processing may also be defined as a method for enhancing raw pictures captured by cameras or sensors on satellites, spacecraft, and aircraft, as well as photos acquired during a range of everyday chores. Over the past five-decade, several approaches have been created in graphic processing. In the rapidly expanding field of computer science, digital

R. Tanya Bindu (✉) · T. Kavitha
AMCEC, Bangalore, Karnataka 560083, India
e-mail: tanyapro2797@gmail.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*, Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_28

imaging, computer processors, and multimedia equipment have all advanced in recent years. Due to their versatility and accessibility, fields that formerly used analog images are now migrating to digital systems. Medicine, video creation, photography, remote sensors, and security monitoring are all good examples. In addition to what would otherwise be studied, these and other sources generate significant volumes of digital photographic material on a daily basis. Digital image processing's main goal is to extract useful information from photographs. This would be done entirely by computers in an ideal world, with little or no human involvement [1].

It is common knowledge that today, data security has grown to be a top priority. The development of contemporary communication technologies necessitates the use of unique security measures, particularly when it comes to data networks. As the amount of data being transmitted over the Internet grows daily, network security is becoming increasingly crucial. Cryptography and steganography are the two key methods for ensuring security. Both approaches are well-known and often used in information security.

Steganography is the practice of concealing information in digital media by embedding secret messages in such a manner that only the sender and the intended recipient(s) can recognize their presence. By using certain cryptographic methods, it is possible to transfer data over the Internet safely while making it impossible for an adversary to intercept or steal private or secret data [2].

2 Application of Image Processing

Image processing has become relatively ubiquitous in many different disciplines as vision, image sensors, and computer technologies have evolved. Picture enhancement for better visualization, image compression and transport, and automatic image recognition representation are some of the other applications of digital image processing. The most common uses of digital image processing in military and security applications are low target detection and tracking, arrow direction, vehicle navigation, wide area surveillance, and automatic/aided target detection. One of the goals of image processing in defense and defense systems is to reduce the amount of work that human analysts have to do in order to deal with the growing number of picture data. The process of Steganography System is shown in Fig. 1.

The second most challenging aim of image processing is to develop algorithms and methods that will greatly be dealt with the development of autonomous systems which help in decision-making based on all sensory inputs. Recognition system, authorization techniques, multi-dimensional image processing, image processing, video analysis, customizable DSPs for video coding, high-quality color representation, super high resolution image processing, and other applications are examples of digital photo processing applications [3–5].

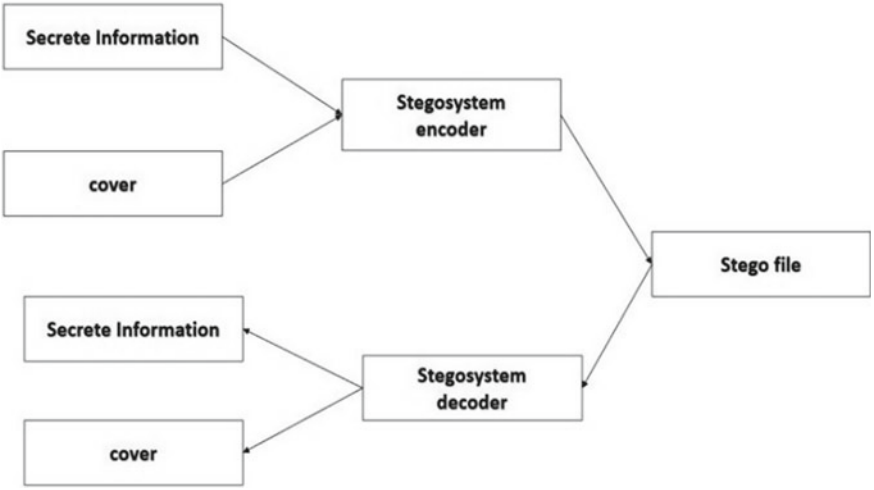


Fig. 1 Steganography system

3 Issues in Image Processing

3.1 Image Compression

The use of digital technology in image storage and transfer is a current trend. Creating a digital television transmission necessitates a large amount of bandwidth. In systems such as teleconferencing, a 64-km per the second route is preferred. Even low channel bandwidths (e.g., up to 1 kb per second) are suitable for some applications, such as videophone and video mobile.

3.2 Image Enhancement

One of the goals of development is to increase the quality of photos by processing them. The image may be of poor quality due to low contrast, noise, or blurriness, among other factors. Many algorithms have been developed to combat this type of corruption. The most difficult task is removing the degradation without causing signal damage. Sound reduction algorithms, for example, frequently include location, scale, or slide, which, unfortunately, blurs the image's boundaries. Methods of orientation have been considered, such as slightly sliding along the edges. However, they are most effective when the harm is modest. How to improve the most corrupted photos is a difficult challenge [6].

3.3 Image Recognition

The recognition system should, in most situations, categories the unknown input pattern into a set of categories. When the number of classes is small and all members of a class are almost identical, this technique becomes easier. The problem can get much worse if the number of classes is excessive, or if members of the same class have dramatically varied looks. As a result, the most difficult task is learning how to perceive commonplace things. How do you make a system that identifies the word "book," for example?

3.4 Image Visualization

The term "visualization" is frequently used to refer to a component of computer graphics. The major goal is to use three-dimensional object and event models as the foundation for creating visuals or picture sequences. Using sensitive pieces to create dynamic scenes is difficult (like clothes, hair, trees, waves, clouds, etc.). Models must be realistic, yet computational expenses must be affordable [6].

3.5 Image Security

Modern photos are sent through the internet to a variety of applications, including police enforcement, military websites, educational, bank data, and private documents. Because these photographs may include sensitive and secret information, their security is of fundamental importance. That means they must be protected while transmitting from attackers or hackers [7]. To put it another way, these photos should be resistant to leaks. There are a number of methods for safe transmitting data which can be used, including picture and data encryption utilizing standard methods.

4 Application of Cryptography in Image Processing

An attacker is highly likely to get access to or even attack sensitive real-time visual data transmitted across insecure connections. By encrypting communications to render them unrecognizable, the art of cryptography enables security. By using certain cryptographic techniques, it is possible to communicate data securely over the Internet while making it impossible for an adversary to intercept or steal private or secret data. This is known as cryptography. Encryption and decryption are two fundamental concepts in cryptography; encryption is the transformation of plain text into cypher text, while decryption is the opposite process as shown in Fig. 2. In contrast to cypher text, which is the text that has been encrypted and is ready to be shared, plain text is the text that contains the real message or data that is not encrypted. To encrypt and decode the data, you'll need a key. Cryptography is being used to turn sensitive information in real-time photographs into incomprehensible data to avoid unauthorized access. Schemes are usually proposed with a high level of security. Slower rates, on the other hand, are always a problem due to their great complexity, which makes them useless in real-time imaging applications.

Steganography is the practice of concealing the presence of a communication, whereas cryptography is the practice of concealing the contents of a message. They're both utilized to keep things safe. However, none of them are capable of meeting basic security requirements such as resilience, undetectability, or capacity. As a result, a new approach called Crypto-Steganography was developed as in Fig. 3, which combines cryptography with steganography to overcome each other's faults, making it more difficult for attackers to assault or steal critical data [2].

The above-mentioned approaches namely, cryptography and steganography are coupled to give stronger picture protection than either one alone. The use of two layers of security to safeguard embedded information in a combined cryptography and steganography picture security system complicates steg analysis [8].

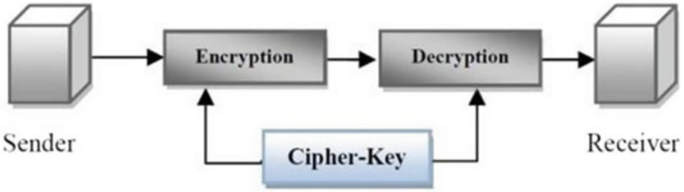


Fig. 2 Cryptography system

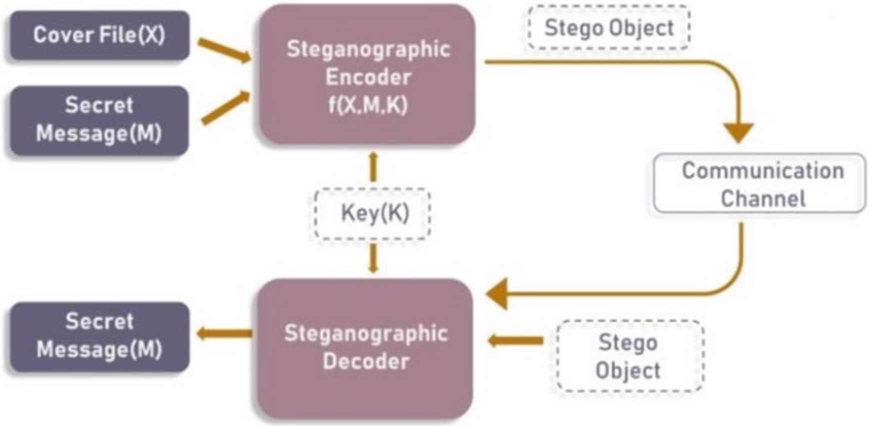


Fig. 3 Crypto-steganography system

5 Literature Survey

Data security is a major concern in today's world that must be handled. Substantially all digital services including internet communication, imaging systems for the military and healthcare, and multimedia systems demand trustworthy security for the transmission and storage of digital pictures. There is a need for security in digital photographs since internet, mobile, and multimedia technology is all advancing more quickly. Therefore, picture encryption methods are required in order to conceal images from such assaults. To conceal images in this system, we employ **Advanced Encryption Technique** (AES). AES is symmetric block encryption that is designed to replace DES for commercial uses. AES is symmetric block encryption developed to counter DES for commercial reasons. It features a 128-bit block size and a 128, 192, or 256-bit key size [9]. The plain text, which is 128 bits long, is divided into 16 bytes, which are organized in a 4 * 4 matrix with four rows and four columns, each representing a single byte (8 bits) [10]. The AES encryption i.e., cipher key after performing AES algorithm also serves as the key for the image in the cryptosystem. Because AES is extremely secure, the picture under test is also safe. The image-based on AES is also faster than some image cryptosystems based on chaotic systems, according to simulation data [11].

Because it uses distinct keys for encryption and decryption, the RSA algorithm is asymmetric. In terms of encryption and decryption, the RSA method has three main phases. Two keys are created during this stage. The term "public key" refers to a cryptographic key that can be used in public. The receiver can only use this key to decrypt the message [12]. Because of RSA's asymmetric encryption technique, encryption is safer, and the receiver is less hesitant to give each sender a unique key to ensure communication. Another advantage of the RSA algorithm is that it is difficult to decipher since it includes the factorization of tough-to-factor prime numbers [13]. A statistical metric that illustrates the strength between two variables is the correlation measurement. The correlation coefficient is utilized in this approach to assess how the algorithm and encryption quality are related. The tighter the association and hence the better the image encryption, the closer the coefficient is to zero. In paper specified as [12], the AES algorithm's correlation coefficient is closer to zero, suggesting higher levels of correlation. The AES technique also provides a higher picture encryption quality since it has more convergent columns in the histogram. Finally, the results of this experiment demonstrated that the AES approach surpasses the RSA algorithm in photo encryption [12].

To prevent assaults, one paper first encrypted a message with the AES algorithm and hashes the key with SHA-2. Following that, they tweaked the **LSB** method by adding a key to make the hiding process non-sequential [14]. LSBS involves overwriting the bit with the lowest arithmetic value, as the name implies. The end outcome of this method somewhat modifies the original output. The modification is done in such a way that it is unlikely to be detected by human eyes [15]. When using LSB methods to create a grayscale image with a record larger than the message content, each pixel can be encoded with three bits. When it comes to switching over the picture pixels and secret messages, we'll only encourage the mystery message to be two-fold bits with the two bits of the LSB, we will receive the key. This key will be stego keyed between the transmitter and receiver. Without this stego key, the recipient will be unable to read the secret data. This framework cannot function without this stego key. This key is known as a Dynamic Symmetric key [16]. While this approach works well with 24-bit color picture files, owing to color variations restrictions and the usage of a color map, it hasn't performed as well with 8-bit color image files. A text message is submerged in the final compressed picture after compression during its implementation. This buried information can only be retrieved by utilizing proper decoding procedures [17]. Another LSB picture steganography technique XOR's each LSB of the red channel with the secret key to embed the secret data, and then chooses whether to replace the secret data in the green or blue channel. The secret key must then be converted into ten-bit stream arrays. In order to retrieve hidden data, they must split the stego picture into the three matrices RGB.

For picture compression, the **Discrete Cosine Transform** (DCT) is often used. Because of its capacity to compress data with a high level of power, multiplication is an important and basic step in computing the discrete cosine transform (DCT) of a picture [18]. It's best to use a 2D Discrete Cosine Transform (DCT) with a convenient carry look-ahead adder. By providing a variable approximation at the lowest power and delay in its most accurate mode, the approximate booth multiplier compensates for the bulky hardware required to provide adjustable approximation using individual approximate multiplier modules in 2D-DCT [19].

The **Discrete Fourier Transform (DFT)** is used in many signal/image processing systems to convert input signals/images from one realm to another. The DFT's hardware is complicated, and different researchers have proposed several implementation solutions. Distributed arithmetic is one of the most attractive and useful approaches for implementing any discrete orthogonal transform. This research proposes an efficient and effective approach for implementing DFT utilizing distributed arithmetic. In comparison to existing group distributed arithmetic for 8-point DFT, the proposed approach uses a recurring pattern of coefficients, stores effectively in memory, and reduces space by 75% [20].

The **Discrete Wavelet Transform** (DWT) is a potent technique for obtaining compressed images at greater compression ratios with higher PSNR values. It uses wavelet transforms for image compression. Wavelets handle data discontinuities better than the DCT since it is not Fourier-based, in comparison to the DCT. The DWT is used to illustrate the signal's dynamic sub-band breakdown. Sub-band analysis is possible without the restriction of dynamic decomposition due to DWT generation in a wavelet packet [21]. The wavelet transform decomposes the signal's time-domain components using a high pass and low pass filter, resulting in low pass (LL) and high pass (LH) filters for the low pass version and low pass (HL) and high pass (HH) filters for the high pass version [22]. One example of DWT in real-time is the design and chip implementation of 2D-DWT using VHDL programming. Level 1 decomposes the original 2D image (64 × 64) into 32 × 32 size LL, LH, HL, and HH sub-bands. Level 2 decomposes the image into 16 × 16 size sub-bands [23].

The study on **Counting Based Secret Sharing (CBSS)** created two different modeling variations for the generation of secret shares, namely the 1-bit and 2-bit methods. The CBSS key research explains the trade-offs involved in determining the number of secret shares versus the target key depending on the application's security requirements. The biggest disadvantage is that as the level of security grows, the proportion of security decreases [24]. They later addressed the flaws in the original CBSS system by proposing a new architecture for secured share distribution that would make pooling and secret restoration easier [25]. Another variant, known as the optimized CBSS, was proposed, which assumed that all phases, including the share generation phase and the share reconstruction phase, were conducted with varied huge sizes of secret keys. In level of security, capacity, and robustness, this experiment used a benchmark of three images of various sizes used to access shares using steganography [26].

We can present a computationally straightforward threshold secret sharing technique in **Matrix Based Secret Sharing (MBSS)**. The method increases the security of already-existing equivalent mechanisms by generating more shares for little secrets at a lower cost than increasing share size [27]. By converting the target key into a matrix, which adds a considerable number of zeros and hence creates a significant number of shares, the matrix- based secret sharing scheme improves upon the secret sharing scheme using counting-based by getting past the zero limitation of the target key [27–30]. To strengthen the security when sending the data, MBSS may be used with any steganography technique like LSB, DWT, etc.

6 Conclusion

LSB is one of the above-discussed stenographic methods that are quick and easy, but it has the drawback that if the message is large, it will affect the stego image and the encrypted data can be seen. To overcome the previous drawback Discrete Wavelet Transform can be used. DWT method is simple, fast, and generates the least amount of visibility and picture distortion. Compared to the other strategies mentioned above, DWT approaches are more quickly discovered and offer more security. In terms of quality, DWT has outperformed other steganography methods. The matrix-based secret sharing method using picture steganography, or the DWT, may be encrypted using a variety of encryption algorithms. The stability and security of the system may be enhanced as a result.

References

1. Soni M, Khare A, Jain S (2014) A survey of digital image processing and its problems. 4:2250–3153
2. Rahmani MKI, Arora K, Pal N (2014) A crypto-steganography: a survey. 5(7)

3. Rao RM, Arora MK (2004) Overview of image processing. In: Advanced image processing techniques for remotely sensed hyperspectral data, pp 51–85
4. Billingsley FC (1970) Applications of digital image processing. Appl Opt 9(2):289–299
5. Saxena E, Goswami N (2014) Automatic object detection in image processing: a survey. 2:2321–8169
6. Huang TS, Aizawa K Image processing: some challenging problems. 21:9766–9769
7. Ahmad SS (2016) Steganography for inserting a message on the digital image using least significant bit and AES cryptographic algorithm. 1–6
8. Bhardwaj R, Khanna D (2015) Enhanced the security of image steganography through image encryption. 1–4
9. Deshmukh P (2016) An image encryption and decryption using AES algorithm. 7:2229–5518
10. Aparna VS, Rajana A, Jairaja I, Nandita B, Madhusoodanana P, Remya AAS (2020) Implementation of AES algorithm on text and image using MATLAB. 978-1-7281-4213-5
11. Zhang Y, Li X, Hou W (2017) A fast image encryption scheme based on AES. 978-1-5090-6238-6
12. Alsaffar DM, Sultan Almutiri A, Alqahtani B, Alamri RM, Fahhad Alqahtani H, Alqahtani NN, Mohammed alshammari G, Ali AA (2019) Image encryption based on AES and RSA algorithms. 978-1-5386-9439-8
13. Jain A, Sharma S (2019) A novel digital image encryption method based on RSA algorithm. 11. ISSN: 0973-7383
14. AL-Shaaby AA, AlKharobi T (2017) Cryptography and steganography: new approach. 5. ISSN: 2054-7420
15. Siper A, Farley R, LomBardo C (2005) The rise of steganography. Corpus ID: 110715828
16. Amarendra K, Mandhala VN, Gupta BC, Sudheshna GG, Anusha VV (2019) Image steganography using LSB. 8. ISSN: 2277-8616
17. Aggarwal A, Sangal A, Varshney A (2019) Image steganography using LSB algorithm. 11. ISSN: 0974-2255
18. Zhou Y, Wang C, Zhou X (2018) DCT-based color image compression algorithm using an efficient lossless encoder. ISSN: 2164-5221
19. Kumar YS, Kumar R, Kumar S (2020) 2D-Discrete cosine transform based dynamically controllable image compression technique. 978-1-7281-8911-6
20. Kumar SRS, Veeramachaneni S, Sk NM (2019) An efficient DFT implementation using modified group distributed arithmetic. 978-1-7281-1381-4
21. Kour P (2015) Image processing using discrete wavelet transform. 3. ISSN: 2321-5984
22. Thakral S, Manhas P (2019) Image processing by using different types of discrete wavelet transform. 499–507
23. Bisht A, Kumar A (2019) DWT chip design and FPGA synthesis for image processing. 8. ISSN: 2277-3878
24. Gutub A, Al-Juaid N, Khan E (2017) Counting-based secret sharing technique for multimedia applications. 5591–5619
25. Gutub A, Al-Ghamdi M (2019) Image based steganography to facilitate improving counting-based secret sharing. https://doi.org/10.1007/s13319-019-0216-0
26. Gutub A, Al-Ghamdi M (2020) Hiding shares by multimedia image steganography for optimized counting-based secret sharing. 7951–7985
27. Porwal S, Mittal S (2020) A threshold secret sharing technique based on matrix manipulation. 2214:020020
28. Al-Shaarani F, Gutub A (2021) Securing matrix counting-based secret-sharing involving crypto-steganography. https://doi.org/10.1016/j.jksuci.2021.09.009
29. Vinothkanna MR (2019) A secure steganography creation algorithm for multiple file formats. J Innovative Image Process (JIIP) 1(01):20–30
30. Chatterjee R, Chakraborty R, Mondal JK (2019) Design of lightweight cryptographic model for end-to-end encryption in IoT domain. IRO J Sustain Wireless Syst 1(4):215–224

A Lightweight Image Cryptosystem for Multimedia Internet of Things

V. Panchami, Arjun Rajasekharan, and Mahima Mary Mathews

Abstract The popularity of social media websites tends to boost the volume of multimedia material. This results in the development of the brand-new industry known as the Multimedia Internet of Things. Lightweight image encryption methods are required to protect multimedia data on these resource-constrained devices. Since chaos theory has grown more common in modern multimedia cryptography, it serves as the foundation for the suggested lightweight encryption technique. 2D augmentation models are used in this chaotic-based multimedia encryption method to provide secure data transit. The suggested method has minimum residual clarity and key sensitivity while, simultaneously maintaining the excellent encryption quality of chaotic maps. The simplified image encryption technique employs the chaotic map model, which has the properties of confusion and diffusion. We have also put forth a novel key generation algorithm for use with the Logistic map and the Rubik's cube transformation. Using the Elliptic-Curve Cryptography (ECC) Key Algorithm, the initial values are produced. To analyze the computational complexity, the suggested method is applied to medical (binary) and colored images. The histograms of the encrypted images are flat and distributed across all the pixel values, according to the security analysis. These images have an entropy of 7.86046675 with average correlation values of 0.0010575 (horizontal), 0.013994 (vertical), and 0.00235 (diagonal) (encrypted image). The suggested lightweight image encryption hence displays a high level of security.

V. Panchami (✉)
CSE-Cyber Security Department, Indian Institute of Information Technology, Kottayam, Kerala, India
e-mail: panchamam036@iiitkottayam.ac.in

A. Rajasekharan
Department of Computer Science and Engineering, Design and Manufacturing, Indian Institute of Information Technology, Kancheepuram, Tamil Nadu, India

M. M. Mathews
Department of Computer Science and Engineering, Indian Institute of Information Technology, Kottayam, Kerala, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_29

Keywords Multimedia Internet of things · Image encryption · Chaotic maps · Elliptic-curve cryptography key algorithm · Arnold's cat map · Henon map · Logistic map

1 Introduction

The tremendous growth in the Internet of Things and multimedia data arise a situation that multimedia data is particularly susceptible, and its protection is more important. Multimedia data is being disseminated at an alarming pace across physical media and the Internet, making data security a critical issue. Data security may prevent data breaches and ensure privacy by encrypting the audiovisual aspect. Therefore, lightweight approaches have to be adapted to secure multimedia data which offers low computational complexity.

The conventional encryption techniques cannot be applied to secure communications in the Multimedia Internet of things as it has high computation cost and complexity. Moreover, Asymmetric encryption techniques have a higher computational cost than symmetric encryption algorithms, in which the image is represented by a vector of real numbers. As a result, the vector is rather lengthy due to the image's enormous sampling coefficients. In computing, a digital image is a 2D vector with pixels ranging in value from zero to two hundred fifty-five. These numbers can be used to represent any geometric shape (circles, curves, or lines) in an image. As a result, safeguarding digital images is critical. Furthermore, the field of information technology is continuously evolving. As a result, cloud computing has grown widely across many industries, including manufacturing, railways, commerce, and government, where security is paramount. As a result, the security architecture is required in these industries to prevent unauthorised individuals from accessing the cloud. Surveillance system applications have developed dramatically during the previous two decades. These systems may be found all over the world (public or private). The cloud, surveillance systems, railways, specialised networks, and applications in the medical field may all benefit from the lightweight image encryption technique. A highly effective image encryption approach is critical for safeguarding content from illegal access and dissemination. Data privacy and security are significant concerns in today's Internet world. By virtue of its low computational cost, the lightweight image encryption approach creates an intriguing framework. This study makes the following significant contributions:

- This approach is developed in order to ensure that neighbouring pixels of the encrypted picture have lower correlation coefficients and requires minimal time and utility for key modification.
- To encrypt and decrypt the images using a highly secure cipher.
- To design a novel Key generation algorithm based on Rubik's cube transformation and Logistic Map.
- To analyse the security of the proposed system.

2 Chaos-Based Encryption and Decryption

2.1 Chaos Maps

Chaotic systems are a subset of nonlinear dynamical systems [1] that are easy to comprehend. They may have a small number of interacting elements and operate according to the basic principles, yet all of these systems exhibit a strong sensitivity to their baseline conditions [2]. In the context of chaos, a map (or function) is an equation or a rule that describes the evolution of a dynamic system through time.

2.2 Chaos Maps for Encryption

Traditional encrypting algorithms such as AES and RSA have many disadvantages and shortcomings [2] when it comes to the encryption of digital images and high processing power:

- Excessive computational time for big images.
- Excessive computational power for large images.

As a result, there may be more advanced approaches for image encryption. A few chaos-based algorithms provide an excellent trade-off between speed, security complexity, and computational overhead.

2.3 Elliptic-Curve Cryptography Key Algorithm

The ECDH (Elliptic Curve Diffie-Hellman Key Exchange) is a mutual authentication system that enables the sender and receiver will generate a secret key on both sides while utilizing the public–private key pairs. ECDH is quite similar to Diffie-Hellman Key Exchange, except that it utilises ECC point multiplication instead of modular exponentiations. The ECDH is centered on this property:

$$(p * G) * q = (q * G) * p$$

Consider: There may be exchange of the secret number $(p * G)$ and $(q * G)$ (the public keys of $P1$ and $P2$) across an unsecured channel and derive a shared secret if there are two secret numbers p and q (two private keys belonging to $P1$ and $P2$) and an ECC elliptic curve with generator point G.

3 Proposed Lightweight Image Encryption Algorithm

The proposed lightweight image encryption algorithm uses Henon Map, XOR, to achieve diffusion property and Arnold Cat Map (ACM) [3, 4] for confusion property. These operations have two inputs, the image which is represented as the $M \times M$ matrix and the keys $K1$, $K2$ and $K3$. The initial values of the keys are generated using Elliptic Curve Cryptography (ECC) key generation algorithm by utilizing the private key of the sender and the public key of the receiver which will be 32 values as in Fig. 3. These 32 values are compressed into 24 values using a compression function while applying the XOR operation. The 24 values are then represented as a $(2 \times 2 \times 2)$ three dimensional matrix for the generation of $K1$, $K2$ and $K3$. The 3-D key generation matrix will be undergone through a transposition operation based on Rubik's Cube. The output of this transformation will be treated as the first key $K1$. The key $K2$ will then act as an input to the logistic Map to produce the second key $K2$. The third key will be the result of the XOR operation between the first key $K1$ and second key $K2$. Figure 1 depicts the overall architecture of the system.

3.1 Key Generation Algorithm

The ECC based key generation algorithm will generate the initial key IK which has 32 values ($32 \times 4 = 128$ bits). This 32 values will be input to a compression function using XOR which is will be represented as follows:

$$IK_i \leftarrow IK_i \oplus IK_{i+8}$$

The output from this function will be 24 values. As the algorithm show below, this 24 values are represented in 3D key generation matrix.

Chaotic systems are a subset of nonlinear dynamical systems [1] that are easy to comprehend. They may have a small number of interacting elements and operate according to extremely basic principles, yet all of these systems exhibit a strong sensitivity to their baseline conditions [2]. In the context of chaos, a map (or function) is an equation or a rule that describes the evolution of a dynamic system through time. Notations and Explanation are shown in Table 1.

3D Generation Matrix and Rubik's Cube Transformation The 3D key generation matrix has 6 sides, F (face) which is represented as RED, R (right) as ORANGE, L (right) as BLUE, U (up) as YELLOW, B (bottom) as GREEN and B (bottom) as white. Each face is again divide into 4 equal parts TL (Top left), TR (Top right), BL (Bottom left) and BR (Bottom right). The Rubik's cube transformation is done by applying six conditions for each side of the cube. For example, if the number is "zero" in the Top left, TL position of the face F of the cube (red, R) then the Matrix should be rotated in a Zig Zag form as: "R2 F2 R2 U2" otherwise it should be rotated

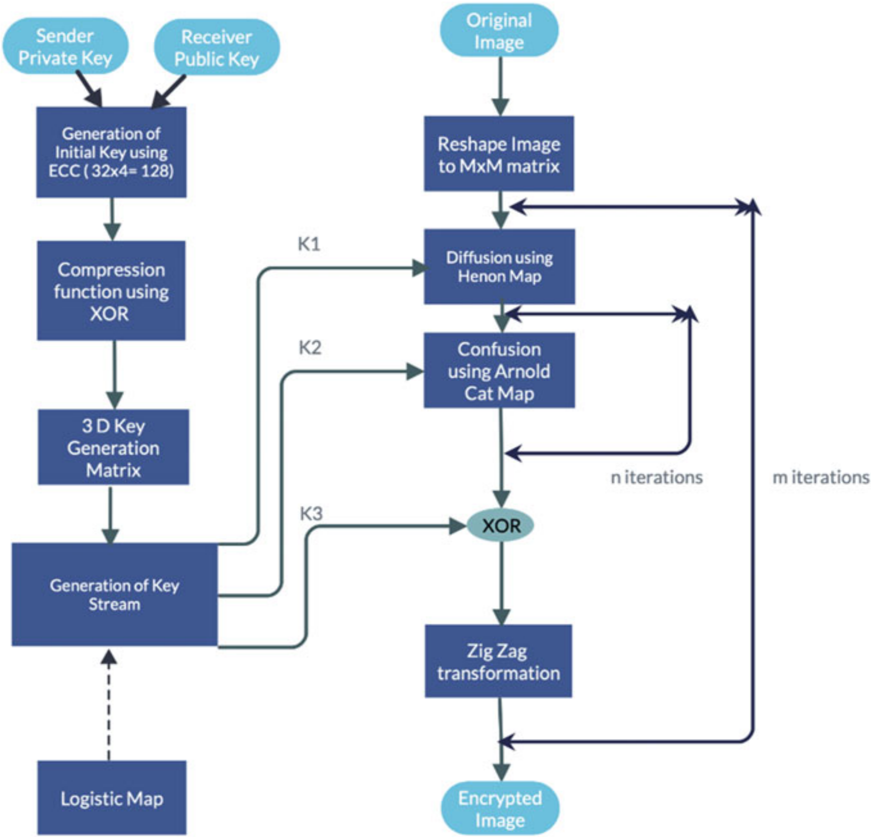

Sender Private Key
Receiver Public Key
Generation of Initial Key using ECC (32x4= 128)
Compression function using XOR
3 D Key Generation Matrix
Generation of Key Stream
Logistic Map
Original Image
Reshape Image to MxM matrix
K1
Diffusion using Henon Map
K2
Confusion using Arnold Cat Map
n iterations
m iterations
K3
XOR
Zig Zag transformation
Encrypted Image

Fig. 1 Proposed chaotic based image encryption scheme

Table 1 Notations and explanation

Notation	Explanation
R	R Orange face of the cube
F	F Red face of the cube
U	U Yellow face of the cube
D	D White face of the cube
L	L Blue face of the cube
B	B Green face of the cube
Mx	Mx Matrix which consists of elements of face X
X	X Face X of the cube will be rotated in clockwise direction
X′	X′ Face X of the cube will be rotated in anti-clockwise direction
Xn	Xn Face X of the cube will be rotated “n times” in clockwise direction
Xn′	Xn′ Face X of the cube will be rotated “n times” in anti-clockwise direction

as Quarter's form, "U F2 U2 R2 U". The result of the final Matrixes are concatenated with the last 6 values of the initial key, to form the first key K1 having 30 values.

Logistic Map based Second Key Generation The second key is generated using Logistic Map. The first key K1 will be the input of the Logistic Map. Each values in K1 is treated as Blocks B_1, $B_2 \ldots\ to\ B_n$. Block means the binary values of each digit. X0 will the initial condition from Logistic map. X01 is the sum of first 3 Blocks B1, B2 and B3 then divide it by 2^{24}. The next value X01 is the sum of next 6 Blocks, B4, B5, B6, B7, B8 and B9 which is divided by 16×6. Then calculate $X_0 \leftarrow X_{01} + X_{02}$. For the next cycle X01 is generated as $X_{01} \leftarrow B_{Previous\,cycle} + 3$ to $B_{Previous\,cycle}$ and X02 as $X_{02} \leftarrow B_{Previous\,cycle} + 6$ to $B_{Previous\,cycle}$. The output as the Logistics Map operation is treated as the second key K2. The XOR operation is applied between the first key K1 (30 values) and the second key K2 (30 values) to generate the third key K3.

3.2 Algorithm for Key Generation

Input: Initial Key (IK: 32 digits)
Output: Keys K1, K2, K3

```
1.    i ← 1
2.    if (i ≤ 26)
2.1.  $IK_i \leftarrow IK_i \oplus IK_{i+8}$
2.2.  Print $IK_i$ and store the value to M, 3D matrix (2 × 2 × 2).
2.3.  Increment i (i++)
3.    if ($M_R$(TL) == 0) then
3.1.  $M_R \leftarrow Zig - Zag$: $R2\,F2\,U2$
3.2.  else
3.3.  $M_R \leftarrow Quarters$: $U\ F2\,U2\,R2\,U$
4.    if ($M_o$ (BL) == 1) then
4.1.  $M_o \leftarrow (R'D'R) * 4U'(R'D'RD) * 2$
4.2.  Else
4.3.  $M_o \leftarrow (R'D'R) * 2U'(R'D'RD) * 4$
5.    if ($M_B$ (TR) == 0) then
5.1.  $M_B \leftarrow$ Cubic: $R\ F\ U'R2\,U\ F'R\,U\ F2\,R2$
5.2.  else
5.3.  $M_B \leftarrow$ Square: $U\ R\ F2\,U\ R\ F2\ R\,U\ F'R$
6.    if ($M_G$(BR) == EVEN NUMBER) then
6.1.  $M_G \leftarrow$ Pillar: $U\ R\,U'R2\,U'R'F'U\ F2\,R\ F'$
6.2.  else
6.3.  $M_G \leftarrow$ Spiral: $U2\,F2\,R2\,U2\,R$
7.    if ($M_W$ (TL) == ODDUMBER) then
7.1.  $M_W \leftarrow RDB2D'R'$
```

```
7.2.   else
7.3.   $M_W \leftarrow D'R'D$
8.     if ($M_Y$ (BL) == EVEN NUMBER) then
8.1.   $M_Y \leftarrow URU'L'UR'U'L$
8.2.   else
8.3.   $M_Y \leftarrow FLFL'D'L'D$
9.     First Key, $K_1 \leftarrow M_R\, v\, M_o\, v\, M_B\, v\, M_G\, v\, M_w\, v\, M_Y\, v\, Last\, 6\, digits$
10.    Compute Second Key, $K_2 \leftarrow LogisticMap(K_1)$
10.1.  $X_o \leftarrow$ Initial condition from LogisticMap
10.2.  $B_i \leftarrow$ Blocks from First Key
10.3.  $X_i \leftarrow X_j,\ where\ j = 1, 2$
10.4.  $X_{01} \leftarrow (B_1 + B_2 + B_3)/2^{24}$
10.5.  $X_{02} \leftarrow (B_4 + B_5 + B_6 + B_7 + B_8 + B_9)/16 \times 16$
10.6.  $X_0 \leftarrow X_{01} + X_{02}$
10.7.  For the next cycle
10.8.  $X_{01} \leftarrow B_{Previous\, cycle} + 3$ to $B_{Previous\, cycle} + 5$
10.9.  $X_{02} \leftarrow B_{Previous\, cycle} + 6$ to $B_{Previous\, cycle} + 5$
11.    Third Key, $K_3 \leftarrow K_1 \oplus K_2$
12.    Print $K_1, K_2, K_3$.
```

3.3 Encryption Process

The input of the lightweight image encryption is the original image. This original image will undergone through a reshaping mechanism in which it composed of two main stages: preparation and pixel modification. The image dimension is rectangular and the image colour model is RGBA (red, green, blue, alpha). For instance, a 400 × 600 image becomes a 600 × 600 image. It will be updated with new pixels with an alpha coefficient of 254, rather than 255. A pixel's transparency is controlled by the RGBA colour model's Alpha channel. To distinguish them from the original image pixels, the subsequently inserted pixels are assigned the value of 254. The new pixels range from 0 to 255. Pixel manipulation has two stages: confusion and diffusion. Both stages need beginning values. These results are calculated using the keys K1, K2 and K3.

3.4 Confusion Process Utilising ACM

This image encryption system's first-pixel alteration is called confusion [5]. The picture is shuffled or randomized such that it can no longer be recognised. In this architecture, ACM [6] is employed to calculate pixel displacement. ACM [7] needs three positive parameters: p, q, and the amount of repetitions. Its values originate from the hidden key element values. The ACM formula is [8]:

$$\begin{bmatrix} x_{n'} \\ y_{n'} \end{bmatrix} = \begin{bmatrix} 1 & p \\ q & pq+1 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix} \bmod N$$

The above formula yields a matrix with x_n and y_n as the new position. The pixels in position (x_n, y_n) will be relocated to $(x_{n'}, y_{n'})$. This computation is repeated until all pixels are shuffled. The mesh grid function of Python NumPy [9] may be used to efficiently move pixel coordinates in a matrix. First, create a N × N mesh grid for x and y to represent the original pixel location in the picture. Two mapping matrices are generated for the new pixel position using the aforementioned algorithm. That is, the [x, y] pixel value shifts according to the matrix. This process is performed for the stated number of times. To make the output matrix less predictable [10], the values of p and q are varied by two integers from the hidden key.

3.5 Diffusion Process Utilising Henon Map

This diffusion technique works by employing bitwise OR gate logical operation on a matrix of image pixels created by a chaotic map matrices [11]. The matrix is Henon Map. The matrix requires two starting variables, x_n and y_n. Both starting values come from hidden key values, much as the confusion stage. The values of x_n and y_n are 0–1 in this scheme. This implies that slight changes [12] in values may have big effects on the diffusion. The chaotic map is built by bit-wise XORing the image pixels produced by the confusion phase [13]. A new vector is created using the diffusion findings.

In order to remove the high correlation between neighboring pixels, the zigzag approach was employed to scramble the pixel positions of the image [14]. The process began with the first pixel of the image matrix, and then proceeded to traverse the consecutive pixels in two dimensional zigzag mode, converting the two-dimensional matrix into a one-dimensional sequence. The starting position for a matrix of size 3 × 3, for example, begins at element 1. Figure 2 shows the original matrix and one-dimensional sequence after a 2D zigzag scan.

3.6 Decryption Process

Decryption is the inverse process of encryption, decryption includes inverse of diffusion and confusion as well as crop border processes. The altered image pixels are used to restore their native values and placements. Both phases need initial values. The hidden key is used to get these attributes' data. To decode the image back to its original form, the inverse of diffusion and confusion [15] is done with the starting value parameters which should be matched with those used in the encryption procedure. Value differences may prevent decryption. After completing the inverse of diffusion and confusion steps, the crop border procedure must be finished. The original image may be seen at this phase, however the image proportions have changed owing to

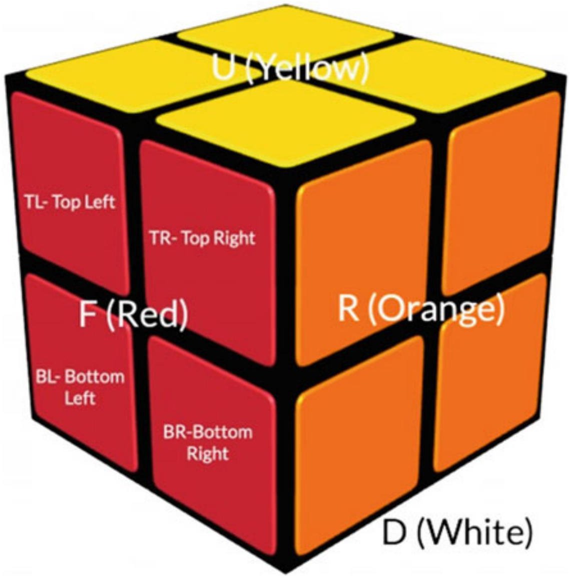


Fig. 2 3-dimensional key generation matrix

the decryption process. So, to restore the image's original proportions, a crop border is used to eliminate the pixels created during the decryption process. The original image may be viewed here, however the decryption process has modified the image dimensions. Then this operation is utilised to remove the pixels formed after the decryption process. Original, Encrypted and Decrypted Images- RGB are shown in Fig. 4. Original, Encrypted and Decrypted Images (Binary) as shown in Fig. 5.

4 Results and Discussions

The security analysis performed in this paper are Intensity Histogram analysis, Adjacent Pixel Auto Correlation analysis and computational complexity is evaluated to evaluate the performance.

4.1 Input and Output Cases

The experiment has been performed in Python 3.9.0 using Visual Studio Code, an environment with a computer of 6 core i7, 2.2 GHz and 32 GB RAM. This test included three RGB images (512 × 512), three medical (binary) images (256 × 256), and a grayscale image (256 × 256).

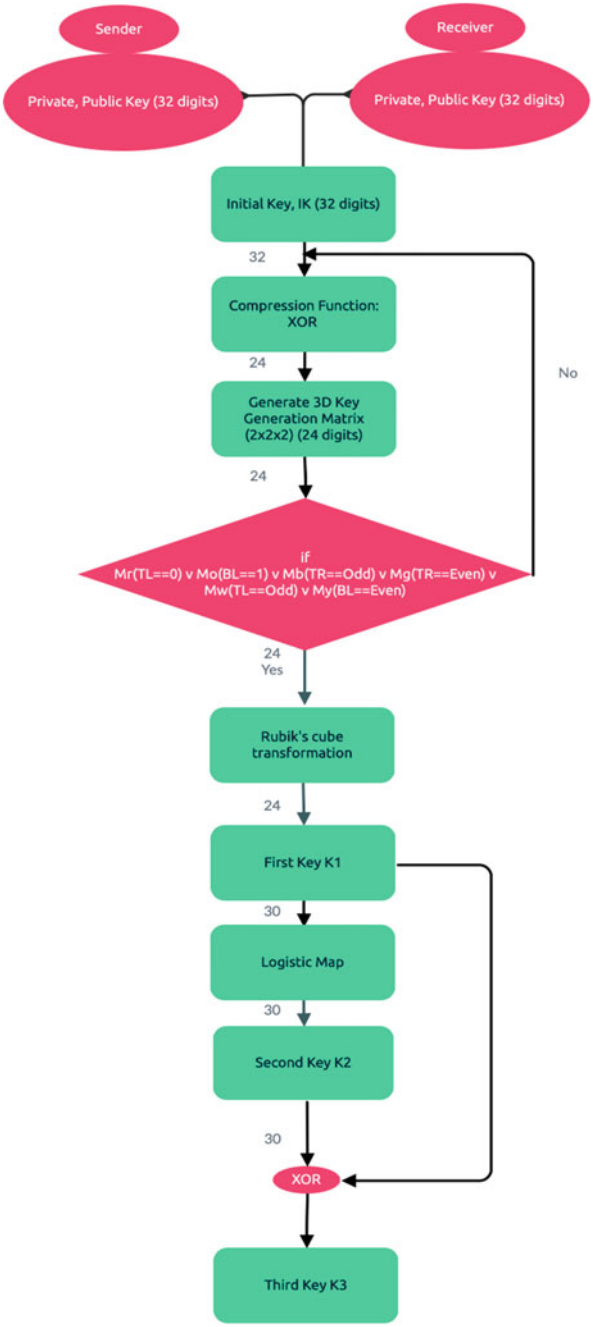

Sender
Private, Public Key (32 digits)
Receiver
Private, Public Key (32 digits)
Initial Key, IK (32 digits)
32
Compression Function: XOR
24
Generate 3D Key Generation Matrix (2x2x2) (24 digits)
24
if Mr(TL==0) v Mo(BL==1) v Mb(TR==Odd) v Mg(TR==Even) v Mw(TL==Odd) v My(BL==Even)
No
24
Yes
Rubik's cube transformation
24
First Key K1
30
Logistic Map
30
Second Key K2
30
XOR
Third Key K3

Fig. 3 Key generation algorithm

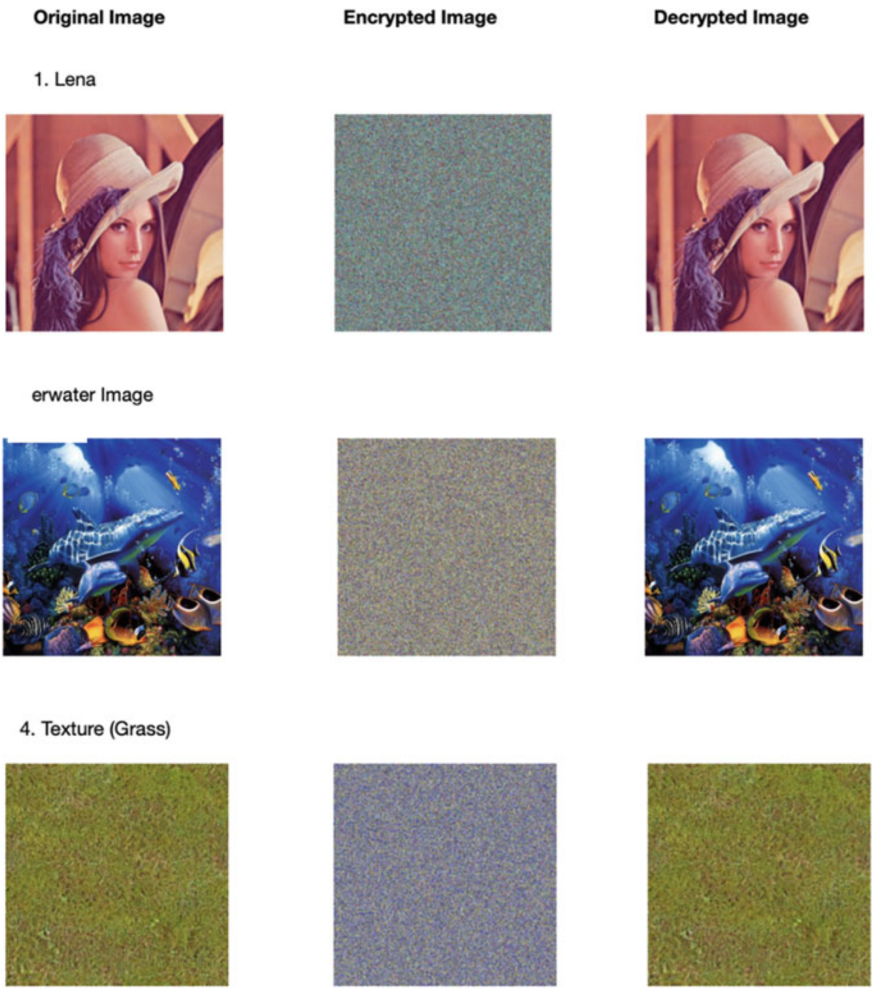
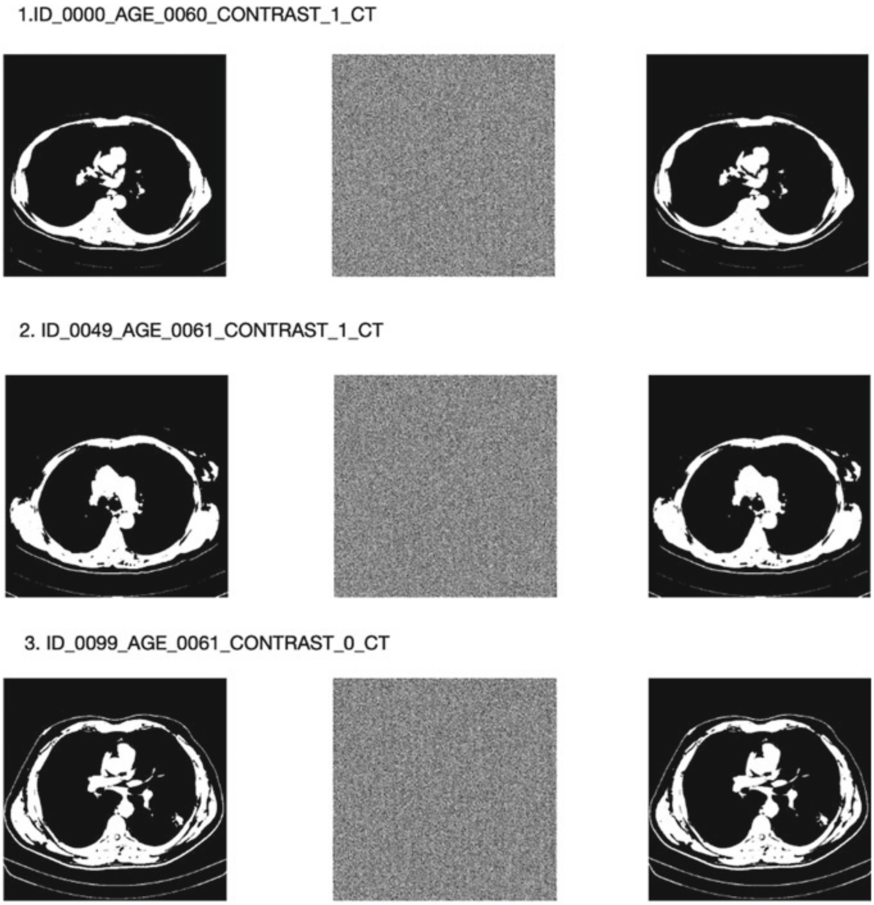


Fig. 4 Original, encrypted and decrypted images (RGB)

4.2 Intensity Histogram Analysis

Image histogram analysis is a simple way to show image encryption quality. Image encryption converts plain text images into random, unintelligible forms. A decent image encryption method produces a cipher image with a consistent intensity histogram. Figures 6 and 7 show the histogram analysis of the coloured and grayscale images. The uniform histogram indicates the best effective encryption technique.



1.ID_0000_AGE_0060_CONTRAST_1_CT

2. ID_0049_AGE_0061_CONTRAST_1_CT

3. ID_0099_AGE_0061_CONTRAST_0_CT

Fig. 5 Original, encrypted and decrypted images (binary)

4.3 Entropy Analysis

Information entropy gauges an encryption system's unpredictability. a typical encryption algorithm's major criteria More entropy usually means a stronger encryption technique. The tables present statistical data for several image groups of varying sizes. For analysis we have considered the encrypted image's entropy, NPCR (number of changing pixels per second), UACI (unified averaged changed intensity), PSNR (peak signal-to-noise ratio), and MSE (mean squared error). Observations from Tables 2 and 3:

i. As the image size grows, the NPCR, UACI, and entropy values grow as well. This is an efficient image encryption method.
ii. General (mixed) images have the least entropy.

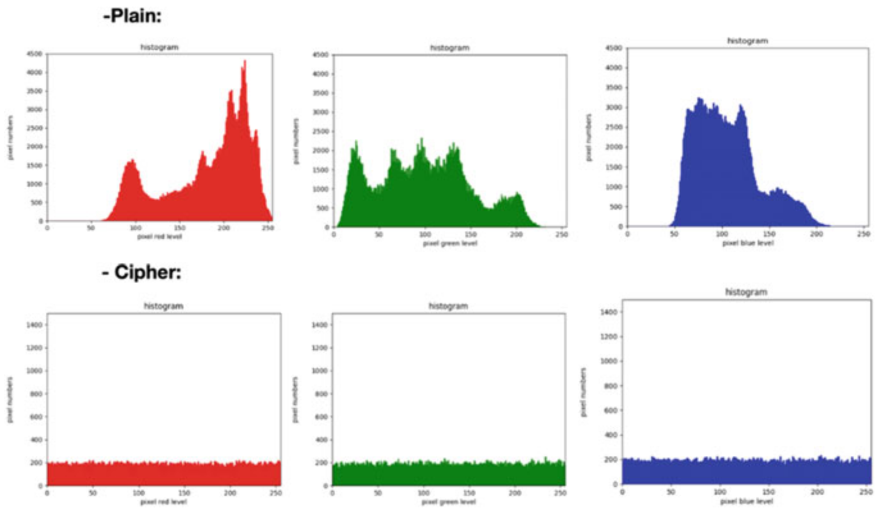


Fig. 6 Histogram of Lena (RGB)

4.4 Computational Complexity Analysis

Various image groupings were studied for computational complexity [16] in terms of time. We need an optimization technique to find the optimal beginning value. The encryption and decryption times are image-independent. However, these time frames vary with image size. The computational performance of encryption and decryption are similar. The dimensions of the image has the highest impact over time complexity [17]. As per the analysis results given in Tables 4 and 5, the proposed system has very less computation complexity for coloured images and it is very efficient for grayscale images.

4.5 Adjacent Pixel Auto Correlation Analysis

The correlation of neighbouring pixels [18, 19] are utilized to measure encryption performance (Horizontal, Vertical or Diagonal). The correlation coefficients [20] of encrypted images are clearly near to zero, indicating that our suggested approach can efficiently eliminate correlations between neighbouring pixels and withstand statistical intrusions [21]. Correlation coefficients between adjacent pixels for different images on RGB and Binary are shown in Tables 6 and 7.

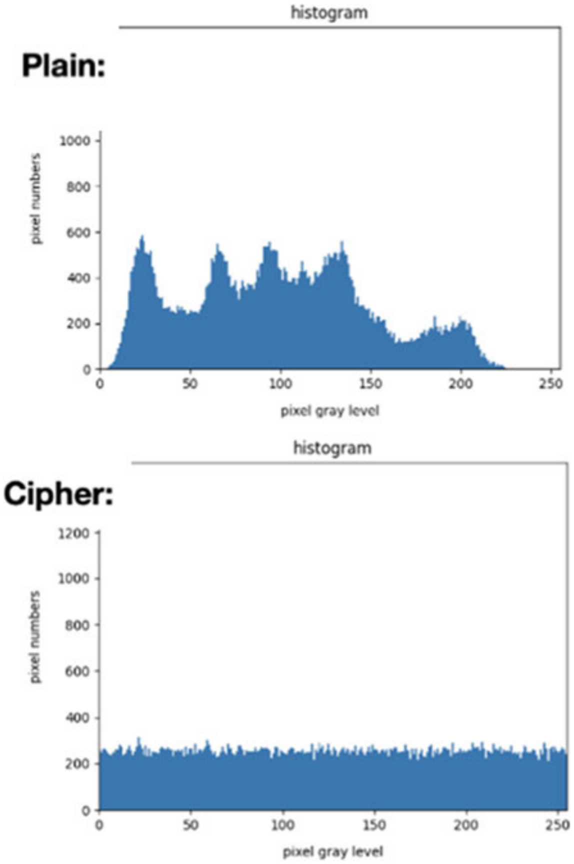


Fig. 7 Histogram of medical image (binary)

Table 2 Statistical analysis for different groups of images (RGB)

Image (512 × 512)	NPCR (%)	UACI (%)	Entropy (plain image)	Entropy (cipher image)
Lena	99.636296	23.517091	7.271856	5.988000
Underwater image	99.962997	34.376602	2.188325	5.736782
Texture (grass)	99.60289	23.826393	6.809031	5.980303

Table 3 Statistical analysis for different groups of medical images (binary)

Medical image (256 × 256)	NPCR (%)	UACI (%)	MSE	PSNR (dB)	Entropy (plain image)	Entropy (cipher image)
MI_1	99.858856	48.823907	21,062.81	7.89564	1.397531	7.997295
MI_2	99.818039	48.911065	21,117.63	7.88435	1.884436	7.997107
MI_3	99.821472	48.097165	20,596.95	7.992773	2.532671	7.997207

Table 4 Analysis of obtained computational complexity in terms of time for different images (RGB)

Image (512 × 512)	Size	Encryption (s)	Decryption (s)	Total time (s)
Lena	512 × 512	1.9602	2.0887	4.0489
Underwater image	512 × 512	1.9255	2.0951	4.0206
Texture (grass)	512 × 512	1.9119	1.9998	3.9117

Table 5 Analysis of obtained computational complexity in terms of time for different medical images (binary)

Medical image (256 × 256)	Size	Encryption (s)	Decryption (s)	Total time (s)
MI_1	256 × 256	0.4694	0.4952	0.9646
MI_2	256 × 256	0.5183	0.6097	1.1280
MI_3	256 × 256	0.5158	0.5538	1.0696

Table 6 Correlation coefficients between adjacent pixels for different images (RGB)

Image (512 × 512)	Plain image			Cipher image		
	Horizontal	Vertical	Diagonal	Horizontal	Vertical	Diagonal
Lena	0.979578	0.988129	0.972011	−0.00011	0.0024	−0.0012
Underwater image	0.942922	0.943511	0.901062	−0.002	−0.0025	−0.0017
Texture (grass)	0.781247	0.800600	0.726798	−0.0019	7.2259×10^{-4}	−0.0021

Table 7 Correlation coefficients between adjacent pixels for different groups of medical images (binary)

Medical image (256 × 256)	Plain image			Cipher image		
	Horizontal	Vertical	Diagonal	Horizontal	Vertical	Diagonal
MI_1	0.959446	0.940174	0.913983	0.007529	0.002932	−0.005537
MI_2	0.954145	0.934506	0.909612	0.008509	0.002016	−0.001271
MI_3	0.923748	0.887473	0.846889	0.007303	0.002962	0.000146

5 Conclusions

The results of this study lead to many conclusions in which the proposed Image encryption algorithm is very efficient for both grayscale images and coloured images. The histogram distribution of cipher images are uniform. Moreover, using the ACM in the confusion phase with varying p and q values renders pixel placements unpredictable and difficult to anticipate. The starting values of the Henon map are at least 10^{14} sensitive. Therefore, changing the beginning settings of the ACM, Logistic, or Henon map may significantly alter the encryption and Key generation results. A 128 bit numeric key provides great security since it can produce 2128 variations. Therefore it has high resistance against brute force attack. The biggest dimension of an image (length/width) has the greatest impact on performance and time. This technique uses a single confusion layer and numerous diffusion layers. Chosen chaotic map is exhibits complex non-linearity and choosing a high-dimensional chaotic system expands the key space.. The diffusion function changes pixel values but does not repeat permutations. The new technique has been subjected to several security evaluations, and the results demonstrate that the proposed algorithm exhibits very strong cryptographic properties, less computation, efficient and adaptability to implement in lightweight applications such IoT devices.

References

1. Desai AP, Crasto J (2012) Chaos-based system for image encryption, pp 4809–4811
2. Fadhel S, Shafry M, Farook O (2017) Chaos image encryption methods: a survey study. Bull Electr Eng Inf 6(1):99–104
3. Soleymani A, Nordin MJ, Sundararajan E (2014) A chaotic cryptosystem for images based on henon and Arnold cat map. Sci World J
4. Kocarev L (2001) Chaos-based cryptography: a brief overview. IEEE Circuits Syst Mag 1(3):6–21
5. Kumar Y, Mamta S (2014) A review paper on image encryption techniques. Int J Res Appl Sci Eng Technol 5(4):169–172
6. Abbas NA (2016) Image encryption based on independent component analysis and Arnold's cat map. Egypt Inf J 17(1):139–146
7. Hariyanto E, Rahim R (2016) Arnold's cat map algorithm in digital image encryption. Int J Sci Res (IJSR) 5(10):1363–1365
8. Elzaher MFA, Shalaby M, Ramly SHE (2016) An Arnold cat map-based chaotic approach for securing voice communication. In: The 10th international conference on informatics and systems, Giza, pp 329–331
9. Elshamy AM, Abdelghany MA, Alhamad AQ, Hamed HFA, Kelash HM, Hussein AI (2017) Secure voip system based on biometric voice authentication and nested digital cryptosystem using chaotic Baker's map and Arnold's cat map encryption. In: 2017 international conference on computer and applications. ICCA 2017, Doha, pp 140–146, 2017.8079739
10. Saha C, Hossain MF (2019) Mri watermarking technique using chaotic maps, NSCT and DCT. In: 2nd international conference on electrical, computer and communication engineering. ECCE 2019, IEEE, Cox's Bazar
11. Dyson FJ, Falk H (1992) Period of a discrete cat mapping. Am Math Mon 99(7):603
12. Munir R (2019) A secure fragile video watermarking algorithm for content authentication based on Arnold cat map. In: 2019 4th international conference on information technology (InCIT), IEEE, Bangkok, pp 32–37 (2019)
13. Raghava NS, Kumar A (2013) Image encryption using henon chaotic map with byte sequence. Int J Comput Sci Eng 3(5):11–18
14. Lin J, Si X (2009) Image encryption algorithm based on hyperchaotic system, pp 153–156
15. Afifi A (2019) A chaotic confusion-diffusion image encryption based on Henon map. Int J Netw Secur Appl 11(4):19–30
16. Sekar JG, Arun C (2020) Comparative performance analysis of chaos based image encryption techniques. J Crit Rev 7(9):1138–1143
17. Kamali SH, Shakerian R, Hedayati M, Rahmani M (2010) A new modified version of advanced encryption standard based algorithm for image encryption. In: Kyoto J (ed) Proceedings of the international conference on electronics and information engineering, pp 141–145

18. Zhou Y, Panetta K, Agaian S (2009) Image encryption using binary key images. In: Proceedings of the IEEE international conference on systems, man and cybernetics. Hyatt Regency Riverwalk, San Antonio, TX, USA, pp 4569–4574
19. Zhang X, Han F, Niu Y (2018) Chaotic image encryption algorithm based on bit permutation and dynamic DNA encoding. Comput Intell Neuro Sci 11(2017):1–12
20. Bisht A, Dua M, Dua S, Jaroli P (2019) A color image encryption technique based on bit-level permutation and alternate logistic maps. J Intell Syst 342:1–15
21. Chatterjee R, Chakraborty R, Mondal JK (2019) Design of lightweight cryptographic model for end-to-end encryption in IoT domain. IRO J Sustain Wirel Syst 1(4):215–224

A Study on Parking Space Allocation and Road Edge Detection for Optimizing Road Traffic

H. Varun Chand, Seema Sabharwal, Anil Carie, and S. Arun Kumar

Abstract Road traffic has been increasing rapidly for the past few decades, which has resulted in air pollution, long waiting times in road traffic, frustration, wastage of effective time on the road etc. The driver's poor vision in identifying road edges, especially during the night, may result in major or minor accidents, which all lead to traffic congestion. Finding a parking space during peak hours also results in congestion. With the emergence of machine learning, IoT, computer vision etc., these problems can be minimised. Road edge detection helps to prevent vehicles that are running off the road and parking space allocation helps to allocate unoccupied space for vehicles even at peak times. The paper addressed a study on different methods through which road edge detection and parking space allocation is possible. This literature survey helps researchers to understand the importance of various cutting edge techniques in road edge detection and parking space allocation. Based on the study, a model is being proposed to provide a solution to edge detection of roads and allocate parking slots, thereby optimising the road traffic.

Keywords Vehicular ad-hoc network · Road edge detection · Parking space allocation · Computer vision · Machine learning

H. Varun Chand (✉)
Department of Computer Science and Engineering, College of Engineering Perumon, Kollam, Kerala, India
e-mail: varunchandh@perumonec.ac.in

S. Sabharwal
Department of Computer Science, Government P.G. College for Women, Panchkula, Haryana, India

A. Carie
School of Computer Science, VIT-AP University, Amaravathi, India
e-mail: anil.carie@vitap.ac.in

S. Arun Kumar
Department of Computer Science and Engineering, Bethesda Institute of Technology and Science, Gwalior, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_30

1 Introduction

Nowadays, everyone has their own vehicles and it is found to be difficult to travel along the road during rush hours [1]. Due to the massive growth in vehicle numbers, it becomes very difficult to park vehicles in the city premises during peak hours. This resulted in driver frustration, wastage of time, air pollution etc. Also, while traveling along roads in unfavourable circumstances like foggy, rainy etc., it becomes more difficult for the drivers to identify the edges of roads, which results in driver frustration [2]. If the driver cannot identify the edge of the road, it may sometimes result in accidents.

During rush hours, it becomes more difficult to park a vehicle even if some vacant parking slots are available. This is due to the lack of proper knowledge about that parking space [3]. So, identifying all available vacant slots and informing the drivers about those may help to solve this problem. But unfortunately, so many factors need to be considered if a registered system allocates parking space, like if two or more drivers are allocated the same vacant slot to whom that space needs to be allocated, if the vacant space is allocated by a non-registered user, etc. Even the size of a vehicle also needs to be considered before allocating the space and all these make parking space allocation a challenging task [4, 5]. The road edge detection will identify off road regions and on road regions. During unfavourable driving conditions like heavy rain, fog, night time etc., the driver feels difficulty in identifying the road edge and sometimes may result in accidents. The second largest number of counts for road accidents in the world is caused by road departure accidents, which is due to lack of identifying the road edge in unfavourable weather conditions [6]. This paper included a review on such techniques through which road edges can be detected.

These two problems have been discussed in this paper and provided a detailed literature review on various image processing, computer vision, and machine learning techniques to solve such problems.

2 Study on Parking Space Allocation

Due to lack of proper knowledge about vacant lots, drivers find it difficult to allocate vehicles during rush hours. Hence, this results in air pollution, frustration, excess fuel consumption, wastage of effective time etc.

Context: A parking space reservation system that can allocate space based on auction. Here the V2G concept is used with the parked vehicle to sell their electricity through electric discharge devices [7].

Objective: An auction based reservation system for parking space allocation with focus on V2G technology is proposed.

Method: The user can determine the parking place and time in advance, thereby reducing the waiting time for parking. Here, V2G technology is also used for the storage of electricity for electric vehicles. The parking systems are not based on traditional methods, first come first serve, but based on the willingness of users who pay more money. Hence, it's called an auction-based system. The system invites reservations from users with a deadline, including their willingness to pay per unit of time, scheduled time of arrival, and scheduled departure time. Some parking lots are equipped with electric discharge devices, where there is a facility for electric vehicles to sell their power. The allocation will be made at the scheduled deadline.

Result: The method focused on auction-based systems where the user who pays more money will be allocated the space. Since the space allocated has the facility for electric discharging devices, each of these parking lots containing electric vehicles may act as a power generating source.

Significance: V2G is an emerging technology that focuses on storing electricity from electric vehicles, which has been incorporated into allocating parking spaces.

Drawback: Since V2G technology is used, if any non-electric vehicles are allocated to parking lots with electric discharge devices, the intended benefit may not be received.

Context: CERP-IoT, a project proposed by the European Union to promote various technologies ITS, IoT, etc. This cloud based system can be used for smart parking mechanisms.

Objective: A cloud based mobile application for parking systems is implemented with three layers- sensor layer, communication layer and application layer [8].

Method: This cloud based system used three layers. The sensor layer deals with technologies like RFID, infrared, microwave, CCTV etc. for vehicle parking access control. The communication layer provides connection between the sensor layer and the application layer. It uses 3G, 4G, ZigBee, WiFi, V2X etc. for communication. The application layer uses the cloud tier and mobile app tier. Mobile apps are used by the registered drivers and data processing and optimisation is done at the cloud tier. With the help of sensor layer technologies, vacant and occupied parking lots are identified. The cloud platform will process the space allocation and is informed through the registered app.

Result: A cloud-based application for parking space allocation that processes all information in three layers. Information is stored and processed at the centralised cloud storage unit.

Significance: All data has been processed on the cloud platform, from where the vacant space is allocated for the requested vehicle.

Context: A cloud-based smart vehicle parking system (SVPS) is used to allocate parking space for requested vehicles [9].

Objective: Cloud infrastructure is utilised for allocating vacant parking slots for the requested driver and hence overcomes the problem of parking space allocation on time.

Method: SVPS architecture has a cloud based parking slot management system and traffic management bureau. This bureau is connected by communication terminals, roadside units, parking side units and vehicular nodes. Here, each of these parking slot units is connected with the roadside unit, which is linked with a cloud server. Each PSU collects information about occupied and vacant slots, which is updated on the cloud server. If a specific user is requesting parking based on availability, the slots will be allotted. If a specific parking lot is not available for booking, the system will find and recommend the nearest lot for parking. The architecture of cloud based SVPS is shown in Fig. 1.

Result: It minimises parking search time and improves the performance of parking resource utilisation, thereby reducing fuel consumption.

Drawback: It does not provide privacy and security for the data during vehicular communication. The data is prone to attacks.

Context: A micro-controlled based system that uses image processing techniques to identify parking locations for different vehicles [10].

Objective: To allocate parking areas using a smart system based on microcontroller and image processing methods for registered users.

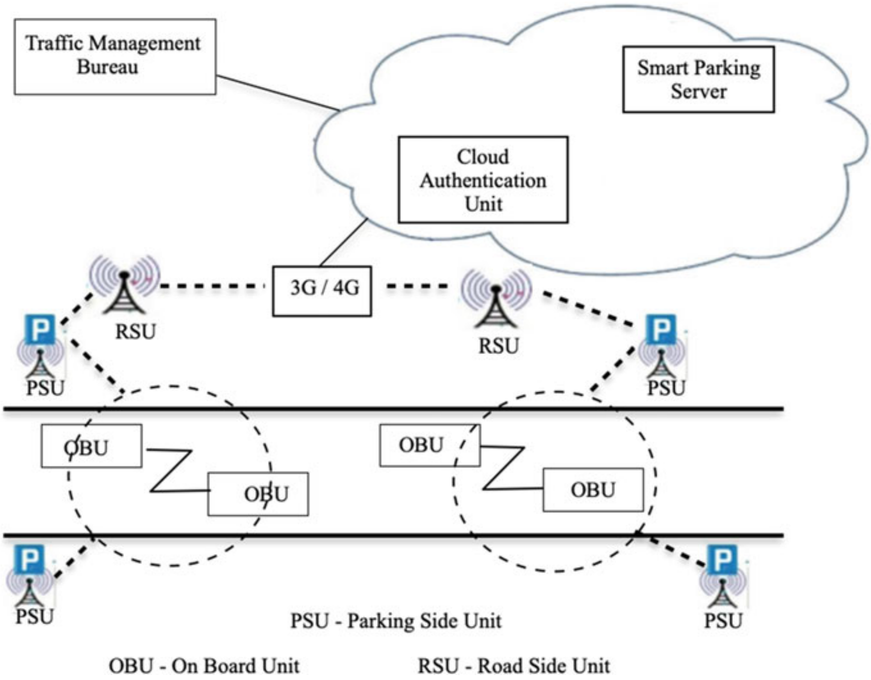


Fig. 1 The architecture of the cloud-based SVPS using VANET

Method: Optical Recognition technique is used to recognise whether a user is registered or not. The number plate of the vehicle is scanned and from this, the driver details are fetched and are sent to the server. The server processes this data and based on available parking slots, the location route is sent to the registered user through the user app. This system works on a specified private campus area.

Result: It allocates spaces that are near to the user. If no such spaces are available nearby, it will allocate space outside the campus area.

Drawback: It does not specify unregistered users. If such users allocate parking space, it creates problems and may degrade performance.

Context: A real-time monitoring system that collects vacant spaces and, based on requests received from drivers, the vacant space is allocated for parking. It makes use of feature extraction and pattern recognition techniques [11].

Objective: To overcome parking issues, a three-fold method is used for parking space allocation. It involves image collection, preprocessing these images using feature extraction and Bayes classifier techniques, and finally utilising the space using an app.

Method: It's a three-step process. Parking space related information is collected from the driving onboard cameras of vehicles, which are preprocessed and used for detection of vacant parking slots. The parking lots are equipped with various IR sensors, the information collected is transferred to Node MCU which will be classified as allocated and unallocated parking space details to the firebase. The firebase also contains the in-time and out-time of each vehicle. This information is later used for allocation of parking space to a requested vehicle. The architecture is shown in Fig. 2. The paper used feature extraction and pattern recognition for preprocessing the collected images.

Result: The study results show this method has a precision of 0.9967and recall of 0.9967.

3 Study on Road Edge Detection in Unfavourable Weather Conditions

A large number of road accidents occur due to vehicle off-road collisions. The drivers are unable to detect road edges especially because of unfavourable environmental conditions like rain, fog, night driving etc. [12]. The speed of vehicles and curvy roads can also result in off-road accidents, but here the study focuses on edge detection in unfavourable weather conditions. Different intelligent methods are used to avoid such problems by clearly identifying road edges.

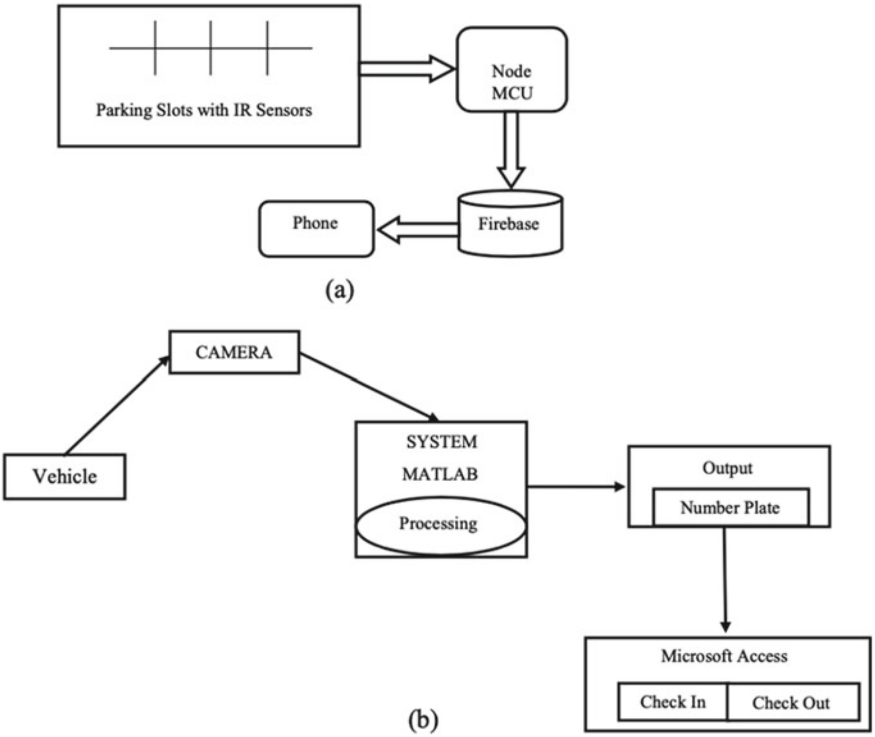


Fig. 2 Architecture for parking space allocation system

Context: The estimation of ambient visibility range in the presence of fog is very difficult. A deep convolutional neural network approach has been used to enhance the awareness of driving weather conditions [13].

Objective: To provide driver awareness about the weather conditions and thereby reduce the safety risk.

Method: A pre-trained deep convolutional neural network is used to extract the features of captured images. For visibility range estimation, $d = -\frac{1}{k}\ln(0.05)$ can be used. Support Vector Machines (SVM) algorithms are used to classify the visibility range. The AlexNet architecture is used as deep CNN, which has 5 convolutional layers, 3 softmax pooling layers and 3 fully connected layers.

Results: The low visibility due to fog or bad weather can be detected using a deep CNNN method with an accuracy of 88.92. The system warns drivers and recommends the proper speed, which reduces the risk of accidents due to bad weather.

Context: Advanced driver assistance systems help to assist drivers by providing safety and comfort. Computer vision also plays a crucial role in providing intelligence to these systems [14].

Objective: To identify the road edge in a foggy climate, using computer vision techniques.

Method: Computer vision helps to reduce the cost and it provides more intelligence in detecting the road edges. If fogs are detected, canny edge detection algorithms are used which enhance the edges in the images. Also, the Hough line detector is used to estimate edge lines of the road images. The vanishing point of the image will be found after this step. Then road and sky segmentation and estimation is carried out. And finally, the visibility distance is estimated.

Results: Visibility distance of roads are found using this computer vision technique.

Context: For safe driving during fog, technological methods have to be applied to detect the edges of roads.

Objective: A fog detection method is used to make driving more safe during foggy times. It selectively uses a de-fogging method if there is a presence of fog [15].

Method: The method is considered only for foggy images that are captured by the onboard camera of the vehicle. It uses de-fog and de-hazing techniques that prevent performance degradation due to decreased visibility. A fog detection algorithm is used to detect the presence of a foggy environment. The method is shown in Fig. 3. Because the image quality is strengthened and the fog from those images has been removed using the fog removal algorithm. The foggy images have low saturation and high value in the HSV color domain, which helps to identify the presence of fog.

Results: The experimental result shows that this algorithm has a fog detection accuracy of more than 97% and it improves the image quality of existing de-fogging algorithms.

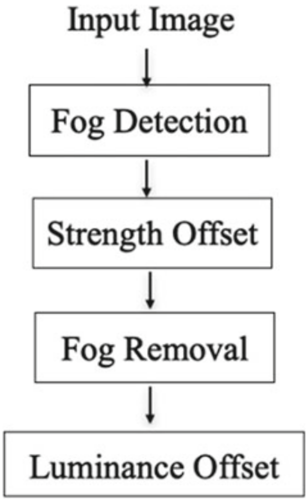


Fig. 3 Overview of de-fogging scheme

Context: Identifying the edges of roads during all weather and illumination conditions are challenging. A driver video data mining technique is used to detect road edges [16].

Objective: Real-time driver video has been used for data mining using all weather and illumination conditions.

Method: Data mining techniques have been applied to driver video data, to find out the qualitative and statistical distribution of road edges. Different weather and illumination conditions are also tested. It also uses the Bayesian method, a probabilistic approach to categories in on-road and off-road regions. Different illumination conditions like sunny and cloudy, rainy and wet roads, snowing and snow-covered roadsides, dark road surfaces, night and invisible roadside etc. can be detected with this approach. So, irrespective of the above-mentioned information, a compact road edge image can be generated from the video data.

Results: The overall boundary accuracy for this method is 87%. By using the Bayesian method, prior knowledge helps to reduce the error of road edges.

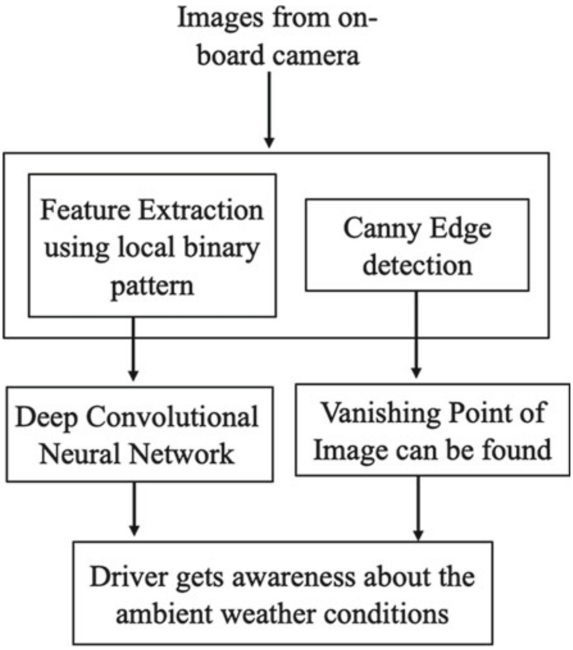
4 Proposed Model

The time wasted searching for parking slots and unclear visuals of road edges, vehicles, pedestrians etc. during unfavourable weather conditions etc. may result in traffic congestion. From the literature, a model has been proposed to provide solutions to such problems. The proposed model has two parts (1) for providing driver awareness about the ambient weather conditions and (2) for allocating parking lots. The proposed architecture for both are shown in Figs. 4 and 5 respectively.

The proposed edge detection model identifies road edges in foggy weather conditions and provides assistance to the driver for accelerating or de-accelerating the vehicle's speed. The on-board camera that captures real images has to be preprocessed using local binary patterns. These preprocessed images need to be fed into a deep convolutional neural network. Parallel to this, another method using canny edge detection has to be carried out to find the vanishing point of the image. Using both these methods will help to converge at a clear image, and thereby it can help to assist the driver in foggy and other unfavourable weather conditions.

The proposed model for parking space allocation will collect images from on-board cameras. The model will identify current vacant slots around the roadside while the vehicle is moving and the details are updated on the cloud. This updated vacant slot information will be very useful for other drivers who might be searching for a slot around that area. The driver needs to request a parking slot along with the location where it is required. The request is processed in the cloud where updated information about the allocated and vacant slot is kept. Based on the request received, the cloud will provide a nearby vacant slot based on auction-based methods. The user or driver then opts for a V2G method, where the electric power of vehicles is stored

Fig. 4 Proposed model for road edge and fog detection


Images from on-board camera
Feature Extraction using local binary pattern
Canny Edge detection
Deep Convolutional Neural Network
Vanishing Point of Image can be found
Driver gets awareness about the ambient weather conditions

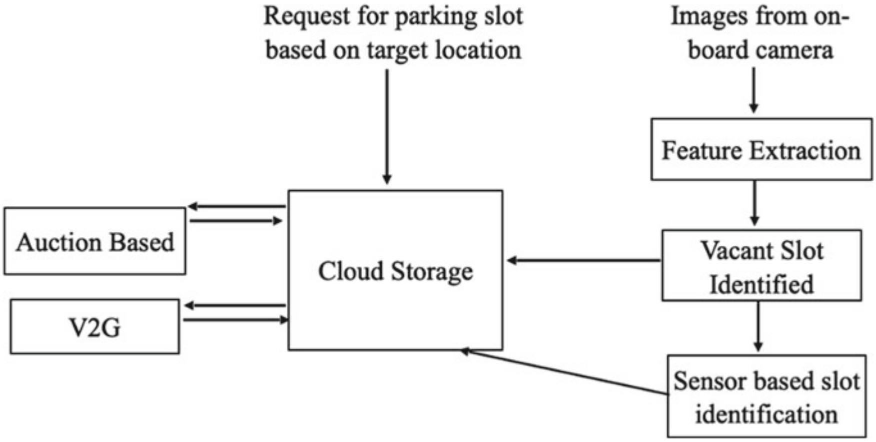

Request for parking slot based on target location
Images from on-board camera
Feature Extraction
Auction Based
Cloud Storage
V2G
Vacant Slot Identified
Sensor based slot identification

Fig. 5 Proposed model for parking space allocation

on to the charging unit near the parking slot. Nowadays, vehicles are moving towards pollution-free methods. So such future options are also added to this model.

5 Conclusion

A model has been proposed to provide solutions to two major problems in VANET. With the advancement of new technologies, parking space allocation and road edge detection can be done with so much ease. Off-road accidents that occur due to poor judgment of road edges, especially during unfavourable environmental conditions, can be avoided with the help of technologies. The study focused on optimization of traffic by providing new cutting edge technology, to solve two major problems i.e., parking space identification and road edge detection. Road edge detections help to reduce off road accidents and thereby ensure smooth traffic on the road. The road edge detection indirectly influences traffic congestion. From this literature it's clear that the new technologies focusing on parking space allocation and road edge detection provide a solution to traffic optimization during peak hours.

References

1. Afrin T, Yodo N (2020) A survey of road traffic congestion measures towards a sustainable and resilient transportation system. Sustainability 12(11):4660
2. Bosch E, Jhme U, Drewitz Jipp M, Oehl M (2020) Why drivers are frustrated: results from a diary study and focus groups. Eur Trans Res Rev 12:52
3. Parmar J, Das P, Dave SM (2020) Study on demand and characteristics of parking system in urban areas: a review. J Traffic Transp Eng (Engl Ed) 7(1):111–124
4. Ruan J, Liu B, Wei H, Qu Y, Zhu N, Zhou X (2016) How many and where to locate parking lots? A space–time accessibility-maximization modeling framework for special event traffic management. Urban Rail Transit 2:59–70
5. Ma Y, Liu Y, Zhang L, Cao Y, Guo S, Li H (2021) Research review on parking space detection method. Symmetry 13(128):1–18
6. Kuehn M, Hummel T, Bende J (2015) Analysis of car accidents caused by unintentional run off road. In: Proceedings of the 24th international technical conference on the enhanced safety of vehicles (ESV), Sweden
7. Hashimoto S, Kanamori R, Ito T (2013) Auction-based parking reservation system with electricity trading. In: 2013 IEee 15th conference on business informatics, Vienna, pp 33–40
8. Ji Z, Ganchey I, O'Droma M, Zhao L, Zhang X (2014) A cloud-based car parking middleware for IoT-based smart cities: design and implementation. Sensors 14(12):22372–22393
9. Safi QGK, Luo S, Pan L, Liu W, Hussain R, Bok SH (2018) SVPS: cloud-based smart vehicle parking system over ubiquitous VANETs. Comput Netw 138:18–30
10. Rane S, Dubey A, Parida T (2017) Design of IoT based intelligent parking system using image processing algorithms. In: 2017 international conference on computing methodologies and communication (ICCMC), Erode, pp 1049–1053
11. Chand V, Karthikeyan J (2020) Design and implementation of parking system using feature extraction and pattern recognition technique, In: Peter J, Fernandes S, Alavi A (eds) Intelligence in big data technologies—beyond the hype. advances in intelligent systems and computing, vol 1167, Singapore, pp 389–400
12. Jagerbrand AK, Sjobergh J (2016) Effects of weather conditions, light conditions, and road lighting on vehicle speed, vol 5. Springer Plus, p 505
13. Outay F, Taha B, Chaabani H, Kamoun F, Werghi N, Yasar A (2021) Estimating ambient visibility in the presence of fog: a deep convolutional neural network approach. Pers Ubiquit Comput 25:51–62

14. Bronte S, Bergasa LM, Alcantarilla PF (2009) Fog detection system based on computer vision techniques. In: 2009 12th international IEEE conference on intelligent transportation systems, St. Louis, USA, pp 1–6
15. Choi KY, Jeong KM, Song BC (2017) Fog detection for de-fogging of road driving images. In: 2017 IEEE 20th international conference on intelligent transportation systems (ITSC), Yokohama, Japan, pp 1–6
16. Wang Z, Cheng G, Zheng JY (2019) Road edge detection in all weather and illumination via driving video mining. IEEE Trans Intell Veh 4(2):232–243

Human Physical Activities Based Calorie Burn Calculator Using LSTM

Jadhav Kalpesh, Jadhav Rushikesh, Kalbande Swaraj, Katta Rohan, and Rakhi Bharadwaj

Abstract Sensors can now recognize human physical activity with the recent technological advances. Accelerometers, gyroscopes, and magnetometers are some of the sensors embedded in smartphones. In today's data-driven world, human activity recognition is important in a variety of fields, including medical applications, fitness tracking, human survey systems, and so on. This research study analyzes the data obtained from mobile sensors such as gyroscopes, accelerometers, linear accelerometers, and magnetometers. Further, the proposed model predicts the human activity by using the data collected from a mobile sensor. Sitting, standing, jogging, and other such activities can also be tracked. Calorie-Meter is an Android application that calculates the calories burned while engaging in such activities. Using the application's predicted activity, the user's calories burned and calorie deficiency can be calculated. This research study proposes the utilization of Long Short-Term Memory (LSTM) and a Neural Network (NN) technique for predicting the human activity based on sensor data.

Keywords Human activity recognition · Sensors · Calorie-meter · Calorie deficit · Long short-term memory

J. Kalpesh · J. Rushikesh · K. Swaraj · K. Rohan · R. Bharadwaj (✉)
Department of Computer Engineering, Vishwakarma Institute of Technology, Pune, India
e-mail: rakhi.bharadwaj@vit.edu

J. Kalpesh
e-mail: kalpesh.jadhav19@vit.edu

J. Rushikesh
e-mail: rushikesh.jadhav191@vit.edu

K. Swaraj
e-mail: swaraj.kalbande19@vit.edu

K. Rohan
e-mail: rohan.katta19@vit.edu

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_31

1 Introduction

Human Activity Recognition (HAR) systems are becoming increasingly popular in the fields of fitness and medical applications. HAR is used in medical applications to care the elderly population by tracking their health records. A magnetometer is used to define the pole position in this project; an accelerometer is used to measure linear acceleration; a gyroscope is used to detect deviation of an object from its desired orientation. Henceforth, the human activities can be predicted by using these sensors. In this project, the dataset used for the training and testing a model is created by using the mobile application named "Sensor Logger" and different actions like running, standing, and climbing stairs were recorded for creating an input dataset. After generating the data, it gets divided into training and testing datasets and the model is then trained by using the Long Short-Term Memory (LSTM) algorithm. This model can now accurately predict the user's current activity. A specific amount of calorie burning rate (Cal/hr) is assigned to every specific activity. Using this, the number of calories burned by that activity in a given amount of time is calculated by providing an exact number of calorie deficit count (Calorie burnt—Calorie intake). The above application allows Android users to easily plan the weight-loss diet plans and exercises.

2 Related Work

Researchers [1] have recently considered the approach of data mining for human activity recognition. Here, big data has been used and further a dataset is created by using 30 test subjects. For the dataset readings, the test subjects used waist-mounted mobile phones for recording and analyzing the user's day-to-day activities. The dataset considered here includes a variety of age groups ranging from 19 to 48 based on many attributes and instances. The dataset contains various physical activities, which are being recognized. This research study has attained an accuracy of 89%.

Both the approaches of human activity recognition i.e., using sensors and human activity recognition using images are considered [2]. Demrozi et al. [2] have carried out a brief existing literature survey by considering 46 papers from various recent conferences. The papers include more information about the different CML methods of ML. It has enlisted a brief information about sensor fusion techniques. The sensor fusion techniques are used to overcome the errors of environmental noise. In normal HAR, the accelerometer can measure the speed accurately but it can be deceived by different angular movements. On the other hand, this research study includes a gyroscope to measure the angular movements, which can be deceived by some linear movements and further to avoid this problem, sensor fusion is used.

In [3], authors have used all 6 basic human activity recognition movements in the dataset. The 6 different signals obtained are then stored as time series and recorded with a sampling rate of 50 Hz. Dataset used here is generated from a gyroscope and accelerometer.

The authors in [4] have proposed a novel method by using physical human activity and energy expenditure estimation in humans with the smartphone application. Also, the author has included an optional heart rate monitor. This method detects the movement of a device by using sensors, which are integrated into smartphones. This method will recognize the movements like linear acceleration and direction of magnetic field by using magnetometer etc.

In [5], the researchers have presented an application, where the innovative method is used for performing activity recognition and analyzing the expenditure of human energy. The proposed application uses data from Smartphone and additionally an optimal heart rate monitor is used to predict the energy expenditure.

In [6], a novel method has been developed for performing human activity classification by using a Support Vector Machine (SVM). They used accelerometers and gyroscope sensors for data collection. Usage of SVM ensemble techniques can leverage an accuracy of 99.1%, sensitivity of 99.6% and specificity of 98.7%.

In [7], a new sensor device has been proposed by using accelerometer vibrations. User can wear sensor device to monitor the physical activities. This paper explains the prevention or detection of human activity by using artificial intelligence, conduction, assessment of rehabilitation exercises, and monitoring of neurological disease progression.

The author of [8], gave a brief idea of support-vector networks. With the explanation of two-group classification problems and the idea of the machine, which implements conceptually. Also, they constructed the feature space in a linear decision surface. Some special properties of decision surface ensure the high generalization ability of the machine learning model.

In [9] the author has conducted a survey on smartphone apps which are used for calorie counting and compared its effect on nutritional awareness and lifestyle modifications. The proposed survey has compared the top 20 apps available on google play store in android applications. The quality of these applications was analyzed using a 55-point scoring system. The characteristics such as accuracy of content, the usability of the interface, standards used, and database source were compared and analyzed in this survey.

In [10], the paper explains the work on suspicious human activity detection using two models, which are trained with a supervised CNN and unsupervised CNN, which doesn't need any records. It uses 250–300 frames of videos in mp4 format as a dataset for performing different activities and successfully predicts the suspicious outdoor human activities which are tracked by an ordinary CCTV camera.

In [11], the authors proposed a novel method to recognize human activity using object information for the detection of different activities. They used activity theory to propose a new technique for activity recognition. For activity theory, authors needed calculation and statistical models i.e., Hidden Markov Model is used. In this paper physiological sensors, cameras, and RFID sensors are used for inputs. Penalized Naive Bayes Classifier is used for different approaches.

In [12], the authors analyzed the difference between the conversation of humans and chatbots. There is various open chat networks where bots are found abusing human beings. This paper studies these bots. Bots are spreading spam and malware through these networks. The chats are distinguished based on entropy i.e., the amount of time between two text messages.

In the [13] paper, the authors use sensors on the body to provide data input. The daily human activities are classified into four categories, they are stationary i.e., not moving (in resting position), light ambulatory i.e., slow walking, intense ambulatory i.e., fast walking or jogging, and abnormal activities like sudden falling of subjects due to a sudden change in environment. The support vector machine based learning and decision tree techniques are used. The accuracy score found is 82.76%, 69.56%, 70.56%, and 60.15% for activities such as Stationary, light ambulatory, intense ambulatory, and abnormal respectively.

In [14], the authors have analyzed different classification techniques like CNN, random forest, and support vector machine. Datasets used here are publicly available. This dataset includes 6 regular human activities calculated with the help of smartphone accelerometers. Activities include walking, jogging, running, sitting, standing, climbing upstairs, and climbing downstairs. The accuracy of different algorithms are calculated, and it was proven that CNN produces the highest accuracy, better than SVM and Random Forest.

In [15] the author used various benchmark datasets to recognize human activity. They selected machine learning and deep learning-based solutions. Also explain the data collection, data pre-processing, and database structure. Importantly they explain that for sensory signals of imaging, a convolutional neural network has been used and for raw sensory, Convolution Neural Network (CNN), Long Short-Term Memory (LSTM), and hybrid models are used.

In [16], authors Harris and Benedict have briefly stated their equation for calculating Basal Metabolic Rate (BMR) after a series of experiments and observations. Normal human individuals of both sex and different ages were used as subjects in 1919.

In [17], Mifflin has proposed a new equation to predict Basal Metabolic Rate (BMR) in healthy individuals. Harris and Benedict's equation for BMR was made more accurate after studies on 498 healthy human subjects. It has then corrected to perform the overestimation of REE by Harris and Benedict.

3 Methodology

3.1 Process Flowchart

Data collection is the first step to start any machine learning model. Data was collected by using the 'Sensors Logger App'. Using this app, the data has been collected from mobile sensors for performing various activities and in various orientations. Data

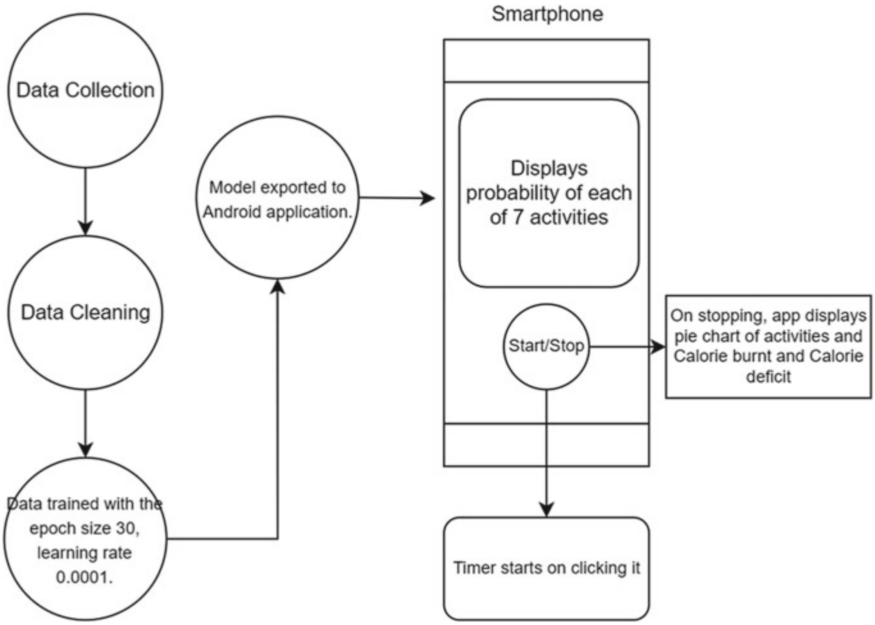

Data Collection
Data Cleaning
Data trained with the epoch size 30, learning rate 0.0001.
Model exported to Android application.
Smartphone
Displays probability of each of 7 activities
Start/Stop
On stopping, app displays pie chart of activities and Calorie burnt and Calorie deficit
Timer starts on clicking it

Fig. 1 Process flowchart

collected from the app is stored in CSV format. Further, it is divided into 80% as training dataset and 20% as testing dataset. Training data has trained the proposed model by using the Long Short-Term Memory (LSTM) algorithm and a model was generated. The pb Model generated here was then exported into an Android application, where it is used to predict activity. Using this we can calculate calories burnt, the calorie deficiency of the user, and show physical activity distribution on the pie chart at a particular time, which will be discussed in Fig. 1.

3.2 Data Collection

Data collection is the first phase of any machine learning project. In this project, the data was collected using the help of the sensors used in the smartphone itself. The Smartphone used was common in all trials. The smartphone model was Xiaomi's Redmi Note 7 Pro. Ten different volunteers participated in this experiment.

The smartphone was attached in five different orientations as follows:

1. left pocket
2. right pocket
3. wrist
4. arm

5. belt.

During each orientation, all seven different activities were recorded for seven minutes.

The activities were as follows:

1. standing
2. sitting
3. running
4. walking
5. jogging
6. stair climbing upstairs
7. stair climbing downstairs.

Sensors of smartphones used to collect data:

1. Accelerometer: An accelerometer uses gravity as a reference to detect the orientation of the phone as well as it can detect motion. When we view the phone in portrait mode or landscape mode, it is detected by the accelerometer, as it detects orientation change. The accelerometer is an electromechanical device that measures acceleration caused due to various natural forces like gravitational force, magnetic force, etc. Accelerometer uses piezoelectric crystals like quartz whose small crystal is attached to a small mass. When the accelerometer moves, that mass squeezes quartz crystal which generates a small voltage that is used to calculate acceleration.
2. Linear accelerometer: This sensor is used in smartphones for getting the acceleration of the phone without considering gravity. This provides data along all three axes in the x, y, and z.
3. Gyroscope: To measure the device's rate of rotation in radian per second we use a Gyroscope. It measures them in all 3 axes of the device. Gyroscopes used in smartphones are Vibration Gyro Sensors that measure angular velocity using Epson's double-T structure crystal element which is a vibrating object. Coriolis force applied on this vibrating object help measure angular velocity in radian per second.
4. Magnetometer: Smartphones come equipped with a magnetometer so that your phone can sense its orientation in space to determine your location for Magnetic North.

All four sensors were used for data gathering. An application from Google Play Store 'Sensors Logger' was used for data collection. This application records the data gathered by the different sensors and converts it into CSV format. 50 Hz is the frequency of data points that are collected using this application. A total of 630,000 samples were collected. 63,000 samples per person. For each activity, 9000 samples were recorded in 5 different orientations per person.

3.3 Data Cleaning and Data Labeling

Removing unusable data from all the datasets was an important factor before labeling the data i.e., removing the starting and ending rows of columns in the CSV files due to unnecessary observations. The observations that were recorded during the transitions from one activity to another were also removed. 1800 sample points per activity per orientation per person were recorded at last. The data labeling was done manually in the CSV files themselves. This data was labeled into the seven activities mentioned above. The final part was to merge all the different CSV files of a single volunteer into one single CSV file for easier access and model generation.

Figures 2, 3, 4 and 5 were generated by using MS Excel from the final Dataset.

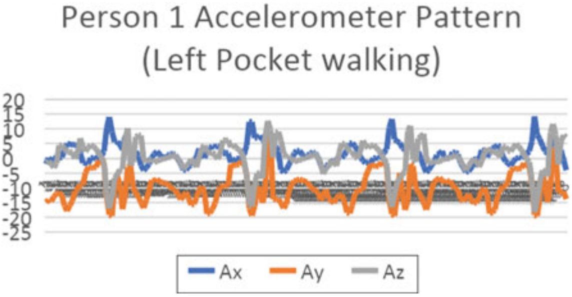


Fig. 2 Accelerometer sensor pattern of person 1. x-axis: a point in time, y-axis: m/s^2

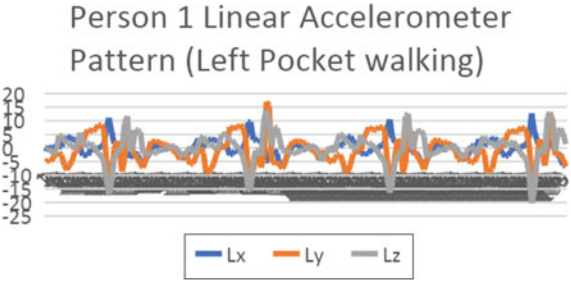


Fig. 3 Linear Accelerometer sensor pattern of person 1. x-axis: a point in time, y-axis: m/s^2

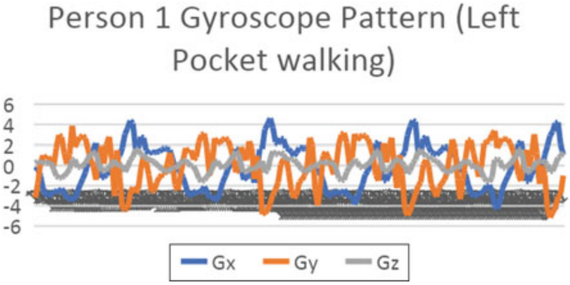


Fig. 4 Gyroscope sensor pattern of person 1. x-axis: a point in time, y-axis: rad/s

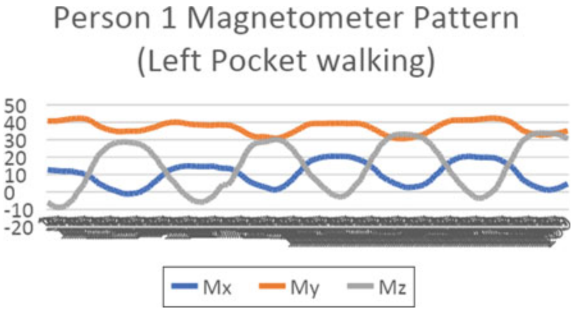


Fig. 5 Magnetometer sensor pattern of person 1. x-axis: a point in time, y-axis: μT (micro-Tesla)

3.4 Patterns in sensor's Data Recorded in the Activity

Many algorithms are used for Time Series Forecasting, they are ARIMA, Prophet, and LSTM. While dealing with a large dataset where lakhs of samples are needed to be trained, LSTM suits best among all. Also, LSTM is easy to use and implement using Python language. It is known that LSTM works on patterns and recognizes them. The challenge of a basic feed-forward neural network is that it can't remember patterns, in short, it has no memory. A neural network that remembers patterns over time series given durational time lags is called Long Short-Term Memory (LSTM). Data of sensors which was recorded had some patterns which we can see in graphs. Each data point in the dataset recorded, corresponds to a point in time.

From the above graphs generated by using MS Excel, we can observe that every reading done by four sensors has a pattern that is a function of time intervals. LSTM's work is to identify this pattern and predict the result of human activity. LSTM is not sensitive to the length of the time gap, which is not the case in RNNs, sequence learning models, and other Markov models. This is one of the significant advantages of using LSTM. For example, if we want to predict something after 10,000 intervals of time, RNNs, and Markov models will forget but LSTM will remember it. But in the case of short intervals like 100, RNNs may act as normal LSTMs.

3.5 Proposed Methodology for Activity Recognition Using LSTM

In theory, RNNs are capable of predicting using long-term dependencies, but in practice, it doesn't seem to be able to learn them. Hochreiter and Bengio discovered why RNNs cannot learn long-term dependencies in practice. In 1997 Hochreiter and Schmidhuber introduced LSTM and it was refined and popularized later [18]. LSTM is a type of recurrent neural network. We can also define it as a Recurrent Neural Network that can remember things over the gap of a certain length or intervals. LSTM is used when we must predict something which is the function of time series or in other words time series forecasting (Fig. 6).

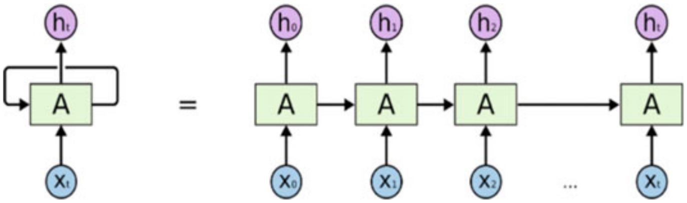


Fig. 6 RNN structure

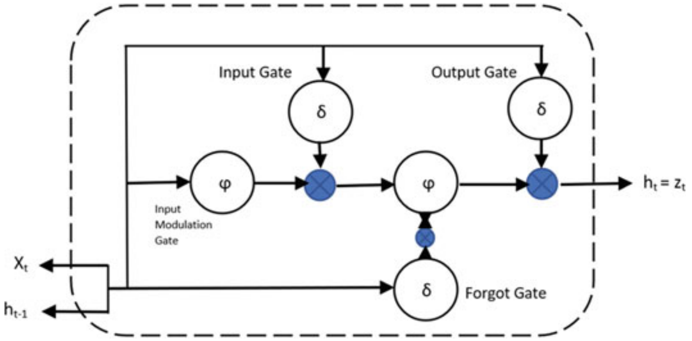


Fig. 7 LSTM cell

RNN cell has two inputs, which help in prediction, that is output of the last state and input given to cell at time t. No information of the past can be retained in a hidden state in the case of RNNs (Fig. 7).

To remember long-term memory, we need another state in the RNN cell, self-state i.e., long term memory besides hidden state i.e. short-term memory. LSTM cell as we observe is similar to RNN cell and has a recurring structure. In RNN, there is only one Neural Network layer of tanh. But this is not true in the case of LSTM a special way of layers are present, where it can have up to four neural networks in simultaneous interaction (Figs. 8 and 9).

The sigmoid activation layer and tanh activation layer are used. Sigmoid function varies between 0 and 1 while tanh function varies between -1 and $+1$. In Sigmoid function is an S-shaped curve that takes real-valued input. tanh function is mostly used in RNN for NLP and Speech Recognition. It varies from -1 and $+1$ which helps in the backpropagation process (Figs. 10 and 11).

Gates have a sigmoid function layer. Gate also has a pointwise multiplication operation. Three gates are present in Long Short-Term Memory. LSTM remembers using a long-term state. If that long-term state needs to be changed, it is done using Forget Gate. The role of Forget Gate is when it encounters certain values during prediction, it must know, it must discard previous long-term memory, and store the new word in this state. To store this new state, we have an input gate. Using Sigmoid and tanh function on Hidden state and new word, cell replaces a new word in long

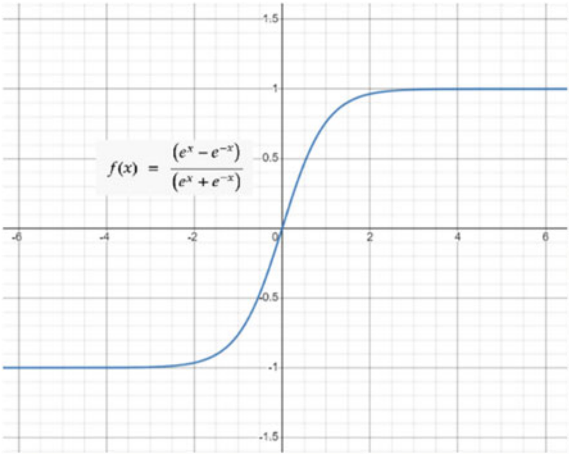


Fig. 8 tanh function activation layer

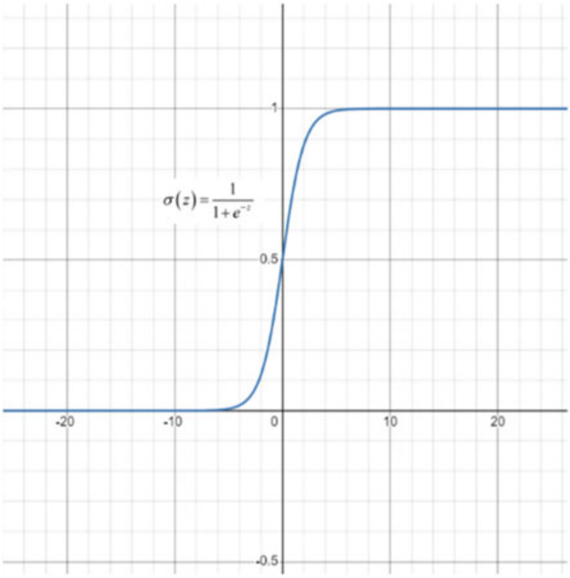


Fig. 9 Sigmoid function activation layer

term state which means we are adding the memory of new value cell read. The third one is the Output gate, where we do the weighted sum of hidden state and value read and then use the sigmoid function on it. Later we multiply it with tanh of new long-term memory to get a new hidden state (Figs. 12, 13 and 14).

Step-by-step LSTM journey.

Step 1: Forget Gate will decide which information that entered cell is important and discard which is not needed. It takes h_{t-1} and *text* as input and outputs number $\in$ [0,1]. See Fig. 12.

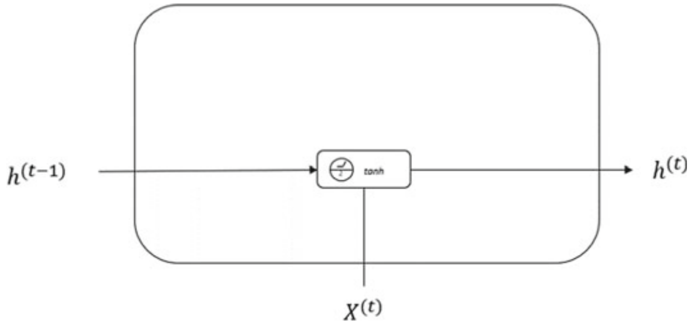


Fig. 10 Short term memory cell in traditional RNN

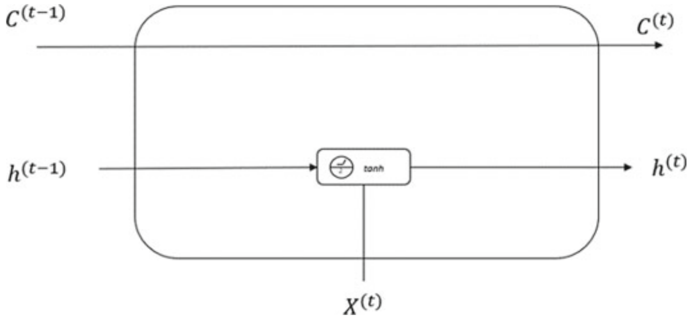


Fig. 11 Short-term and long-term memory

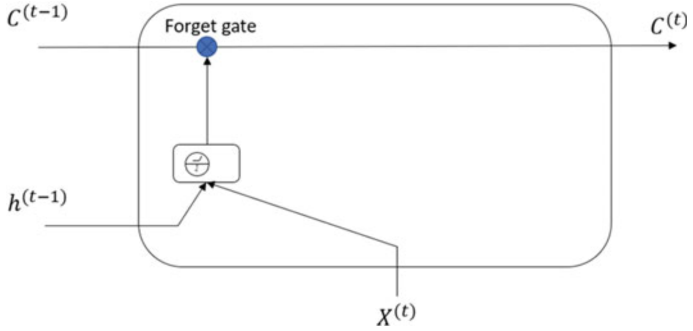


Fig. 12 Forget gate

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{1}$$

here $\mathrm{f_t}$ is an output of forget gate, W_f is the weight of forget gate neuron cell, h_{t-1} is previously hidden state input and x_t is current input, b_f is biased. The sigmoid

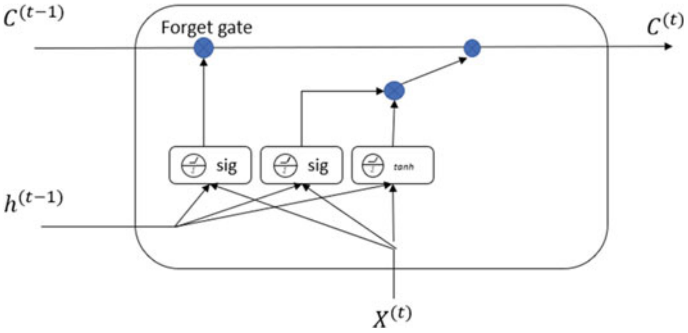


Fig. 13 Input gate

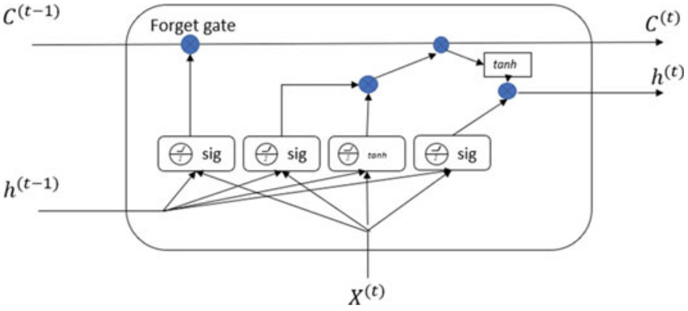


Fig. 14 Output gate

function gives a value between [0, 1] and if its 0 means we must forget the previous memory.

Step 2: Information that needed to be stored in the cell state is decided by LSTM, which has 2 parts. The first one is the Sigmoid function which is also called as Input Gate Layer. The job of the Input Gate layer is to decide which value to update. The second is the tanh layer. It calculates a new candidate value's vector C_t' which could be added to the state.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{2}$$

here i_t is an output of the sigmoid input gate, W_i is the weight of the sigmoid input gate neuron cell, h_{t-1} is previously hidden state input and x_{it} is the current input and $\mathrm{b_i}$ is bias for input gate.

$$C_t' = \tanh(W_c.[h_{t-1}, x_t] + b_c) \tag{3}$$

here C_t' is a temporary new state which is output of tanh input gate, W_c is the weight of tanh input gate neuron cell, h_{t-1} is the previously hidden state input *it* is the current input and b_{cthe} is bias for the input gate.

Now we update the old cell state C_{t-1}. We multiply old state with *f(t)*, forgetting things we decided to forget earlier, then we add $i(t) * C_t'$.

$$C_t = f_t * C_{t-1} + i_t * C_t' \tag{4}$$

here C_t is new Cell State, f_t is output of forget gate, C_{t-1} is the previous Cell State, i_t is the output of the sigmoid input gate and C_t' is the temporary new state which is the output of the tanh input gate.

Step 3: Finally, LSTM needs to decide what the output is going to be. It will be a filtered version and based on cell state. We use sigmoid layer that determines the part of the cell we will return as output. Later apply tanh on cell state to multiply it with output of sigmoid gate earlier, so we output the parts we decided to.

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{5}$$

here *it* is output of sigmoid gate in Output Gate, W_o is weight of output gate neuron cell, h_{t-1} is previous hidden state input and x_t is current input, b_o is bias.

$$h_t = o_t * \tanh(C_t) \tag{6}$$

here h_t is new hidden state which is output of Output Gate, o_t is an output of Output Gate and C_t is the new Cell State which is output of Input Gate.

Using LSTMs, we know patterns in datasets captured by sensors, using which we can do time series forecasting.

3.6 Training of Model and Testing of Model

For building the LSTM model, we considered 7 classes and 32 hidden units, and for activation function we used the ReLU function and kernel initializer as orthogonal. The model is trained for 30 epochs and the batch size is 1024. The training and testing dataset is divided into 80:20 ratio, 504,000 samples are used for training of model while 126,000 are used for testing the model. The time needed for training and evaluation takes 38 s and the average time taken for each epoch is around 1 s 33 ms. The accuracy of the model is 98.17. For checking the performance of model, the confusion matrix is created which is shown in Fig. 15.

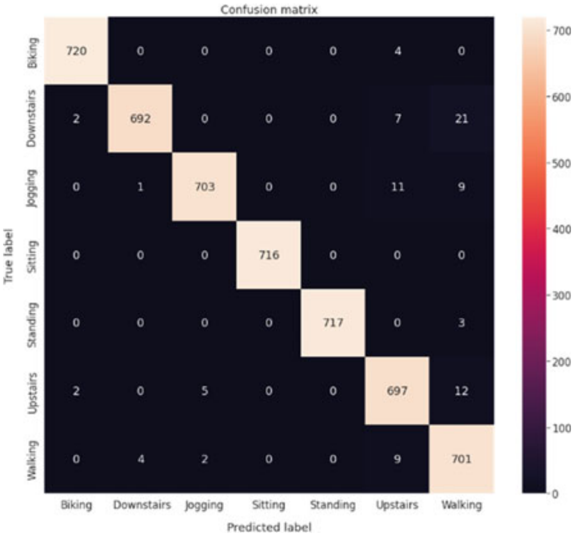


Fig. 15 Confusion matrix

3.7 *Calculating Calorie Deficit Using the Predicted Graph*

Using activity prediction done by LSTM model, it can now approximate the calorie burndown a person is having. Calories are units used to measure energy expenditure. The amount of calories energy our food or drink is measured in calories. A common unit used while dealing with energy dynamics of the human body is kcal. But as kcal is so commonly used in this field, it's commonly called Calorie. The "calorie" we refer to in food is kilocalories. One kilocalorie = (uppercase C) one Calorie. In scientific terms, a kilocalorie is basically amount of heat we require to raise the temperature of 1 kg of water by 1 °C.

Every person, depending on his age, weight, height, and gender has some default energy expenditure in a day, which happens even if you do no physical activity. This is called Basal Metabolic Rate. The basic metabolic rate is equal to minimum number of calories a body need in non-working state. This can be estimated using equations. One of these equations is the Harris-Benedict equation [16, 17], which is a formula that takes one's weight, age, height, gender as input to find BMR or Total Energy Expenditure. BMR calculators are available online.

Men:

$$BMR_m = 88.362 + (13.397 \times W) + (4.799 \times H) - (5.677 \times A) \tag{7}$$

where W is weight in kg, H is height in cm, A is age in years.

Women:

$$\mathrm{BMR_w} = 447.593 + (9.247 \times \mathrm{W}) + (3.098 \times \mathrm{H}) - (4.330 \times \mathrm{A}) \tag{8}$$

where W is weight in kg, H is height in cm, A is age in years.

Calorie burning rate (Cal per hour) for person with weight of 130 lb in activities we are recognizing using model, as per [10] are:

1. standing: 165 Cal per h
2. sitting: 148 Cal per h
3. running: 472 Cal per h
4. walking: 224 Cal per h
5. jogging: 372 Cal per h
6. stair climbing upstairs: 472 Cal per h
7. stair climbing downstairs: 177 Cal per h.

Model will predict activities done by person and then using Calorie burning rate will calculate calories burned.

To calculate Calorie deficit, one must consider calories taken as input. It includes daily items we eat and drink. There are various mobile apps which keep track of these calories we intake.

$$Calorie\,burnt = BMR_u + \sum(t_i * CalA_i) \tag{9}$$

here t_i is time for ith activity, $CalA_i$ is Calorie burning rate in Cal per hour for ith activity, BMR_u is Basal Metabolic Rate of a particular user.

$$Calorie\,deficit = Calorie\,burnt - Calorie\,intake \tag{10}$$

Calorie burnt is energy expenditure due to various physical activities and Calorie intake in energy a person takes after consumption of food. Units are Calories for every term in equation.

Users must input Calories intake into the application developed so that we can calculate Calorie deficit. Using the above technique, we can run an application for a particular amount of time or a day and calculate Calorie deficit for that time or complete day.

4 Result and Conclusion

This paper is proposing new Calorie-Meter which is name of android application using Smartphone sensors that are used to gather data and long short-term memory is used for model generation. There are 7 activities that are classified using this model are: standing, running, walking, upstairs, downstairs and jogging. ‘Sensors Logger” was application used to collect data from smartphone sensor. The signals were received in form of x, y and z coordination system in the 3-dimensional space form every sensor. Data is logged from smartphone sensors and signals given by them while performing the above 7 activities.

This application takes input from the users such as age, height, and weight. model predicts the user's activities by the person and calorie burnt is calculated. The application tells the user how much calories they have burnt in a day and showcases various activities in pie chart representation.

4.1 Accuracy of Model

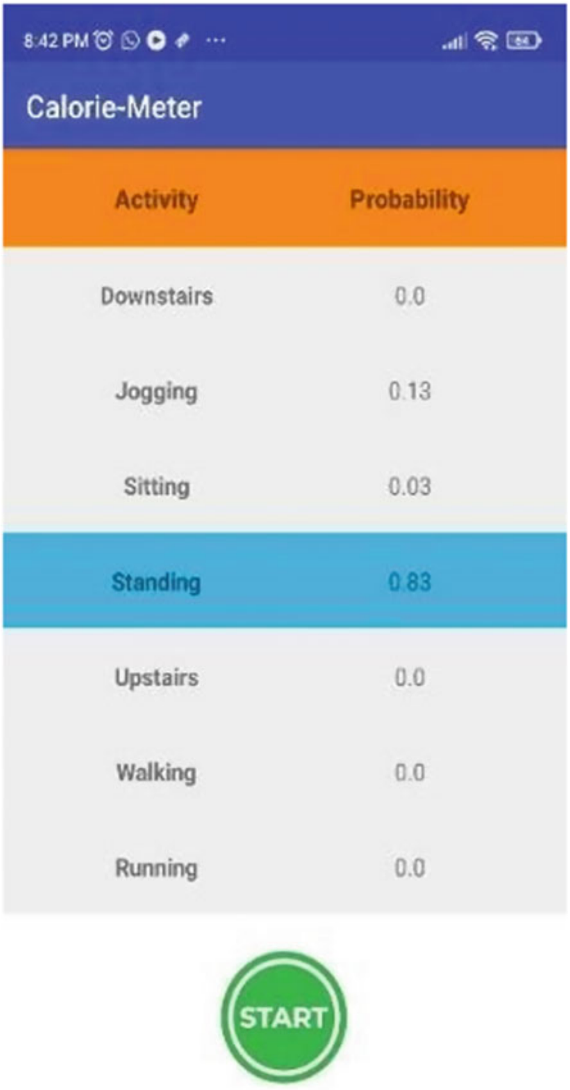
Accuracy of model built is 98.13% and loss is 0.0707.

4.2 Calorie-Meter on Andorid App

Key part of this project is the Android application implementation which carries out live activity detection of a person carrying a smartphone. Design is such that it's easy to understand and user-friendly. Basic user input like weight, height, age, and gender are taken and stored locally in the application. The Calorie-Meter app basically provides users a report of calories they burned and physical activity levels using a Pie chart (screenshots in Figs. 16 and 17).

Using information taken from users, app estimates calories burnt in during activities carried out. User have to start timer of the app. From that point onwards every activity of the user will be recorded along with the time of each activity performed. Once the user stops the timer, a Pie chart is displayed (as shown in Fig. 17) which shows activity distribution in the ratio of time it is performed.

The proposed study has successfully recorded the development of a smartphone Android application that performs live detection of user's physical activities. This app is different from previous works on activity recognition in the following: (1) User interaction needed after setup is close to zero; (2) It relies solely on sensors of smartphones which are available with low-end smartphones too and don't require additional sensing hardware; (3) It recognizes seven different human activities, including walking, running, climbing stairs, descending stairs, sitting, standing, jogging; (4) App is user friendly and easy to understand, which include Graphical representation of data gathered; (5) Accuracy of Calorie-Meter app built is 98.13% which is reliable; (6) App can also calculate Calorie deficit. This functionality is not present in many apps in the marketplace. Application lets user monitor their daily physical activity and enable them to make healthier choices and make good and healthier habits. The application is targeted to encourage people toward a healthy lifestyle.



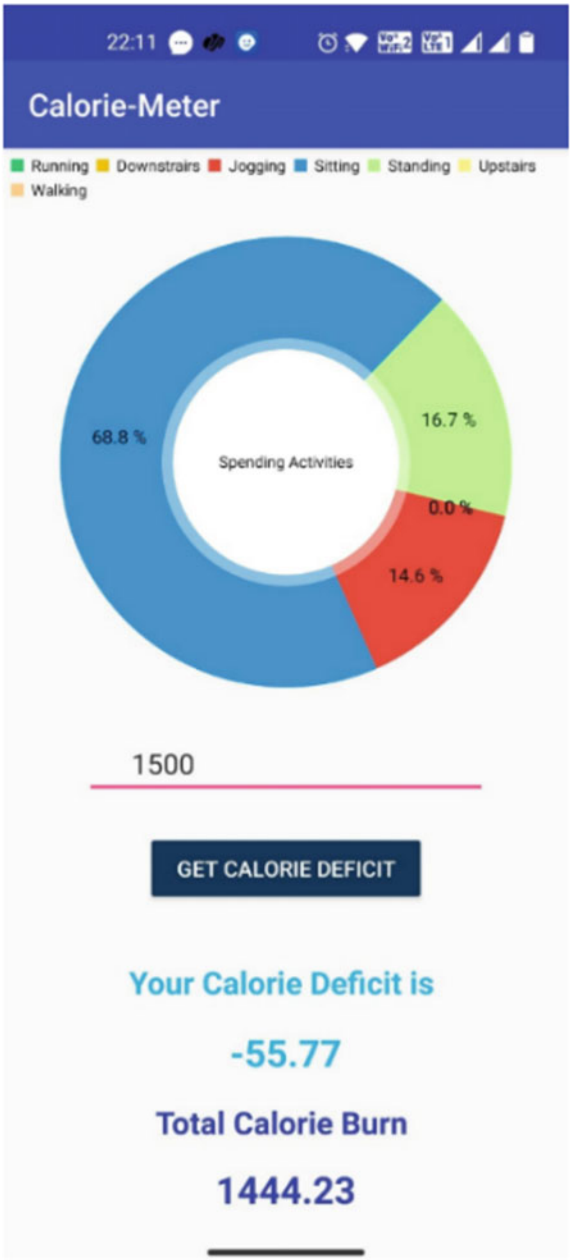
Activity	Probability
Downstairs	0.0
Jogging	0.13
Sitting	0.03
Standing	0.83
Upstairs	0.0
Walking	0.0
Running	0.0

Fig. 16 Human activity recognition dashboard and start-stop timer

5 Future Directions

This project has tremendous scope for future research, wherein the larger dataset can result in increasing the accuracy of the model proposed. Moreover, this will help to maintain the user health after analyzing their daily activities and maintaining calorie goals. The proposed application can be further extended to (1) recognize more physical activities, (2) extra features can be added to provide users with the information about their heart rate, obesity levels, and carbon footprints, and (3)

Fig. 17 Pie chart of human activities in particular time, Calorie deficit calculated, and Calorie burnt



1500

Your Calorie Deficit is

-55.77

Total Calorie Burn

1444.23

develop additional functionality, which uses stored historical information of user's daily activities and help user to make appropriate lifestyle choices, which are healthier and smart.

Acknowledgements We would especially like to thank Vishwakarma Institute of Technology (VIT), Pune for providing us with a platform for producing such a wonderful project. We would like to acknowledge our honorable Director Prof. Dr. Rajesh M. Jalnekar for his inspiration and guidance. Also, the Head of the Department of Computer Engineering Prof. (Dr.) Sandeep Shinde for his valuable support, help, and time. Valuable support and inspiration were given to us by our project guide Prof. Rakhi Bharadwaj for writing this paper.

References

1. Ann OC, Theng LB (2014) Human activity recognition: a review. In: 2014 IEEE international conference on control system, computing and engineering (ICCSCE 2014), pp 389–393. https://doi.org/10.1109/ICCSCE.2014.7072750
2. Demrozi F, Pravadelli G, Bihorac A, Rashidi P (2020) Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey. IEEE Access 8:210816–210836. https://doi.org/10.1109/ACCESS.2020.3037715
3. Bulbul E, Cetin A, Dogru IA (2018) Human activity recognition using smartphones. In: 2018 2nd international symposium on multidisciplinary studies and innovative technologies (ISMSIT), pp 1–6. https://doi.org/10.1109/ISMSIT.2018.8567275
4. Cvetković B, Janko V, Luštrek M (2015) Activity recognition and human energy expenditure estimation with a smartphone
5. Anjum A, Ilyas MU (2013) Activity recognition using smartphone sensors. In: 2013 IEEE 10th consumer communications and networking conference (CCNC), pp 914–919. https://doi.org/10.1109/CCNC.2013.6488584
6. Hardiyanti N, Lawi A, Diaraya, Aziz F (2018) Classification of human activity based on sensor accelerometer and gyroscope using ensemble SVM method. In: 2018 2nd East Indonesia conference on computer and information technology (EIConCIT), pp 304–307. https://doi.org/10.1109/EIConCIT.2018.8878627
7. Caya MVC, Yumang AN, Arai JV, Niñofranco JDA, Yap KAS (2019) Human activity recognition based on accelerometer vibrations using artificial neural network. In: 2019 IEEE 11th international conference on humanoid, nanotechnology, information technology, communication and control, environment, and management (HNICEM), pp 1–5. https://doi.org/10.1109/HNICEM48295.2019.9072850
8. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297
9. Banerjee P, Mendu VVR, Korrapati D, Gavaravarapu SM (2020) Calorie counting smart phone apps: effectiveness in nutritional awareness, lifestyle modification and weight management among young Indian adults. Health Inf J
10. Bhambri P, Bagga S, Priya D, Singh H, Dhiman HK (2020) Suspicious human activity detection system. J IoT Soc Mob Anal Cloud 2(4):216–221
11. Yang J, Lee J, Choi J (2011) Activity recognition based on RFID object usage for smart mobile devices. J Comput Sci Technol 26:239–246
12. Smys S, Wang H (2021) Naïve Bayes and entropy based analysis and classification of humans and chat bots. J ISMAC 3:40–49. https://doi.org/10.36548/jismac.2021.1.004
13. Ali SE, Khan AN, Zia S, Mukhtar M (2020) Human activity recognition system using smart phone based accelerometer and machine learning. In: 2020 IEEE international conference on industry 4.0, artificial intelligence, and communications technology (IAICT), pp 69–74. https://doi.org/10.1109/IAICT50021.2020.9172037

14. Khatun MA, Yousuf MA (2020) Human activity recognition using smartphone sensor based on selective classifiers. In: 2020 2nd international conference on sustainable technologies for industry 4.0 (STI), pp 1–6. https://doi.org/10.1109/STI50764.2020.9350486
15. Ramanujam E, Perumal T, Padmavathi S (2021) Human activity recognition with smartphone and wearable sensors using deep learning techniques: a review. IEEE Sens J 21(12):13029–13040. https://doi.org/10.1109/JSEN.2021.3069927
16. Harris JA, Benedict FG (1918) A biometric study of human basal metabolism. Proc Natl Acad Sci U S A 4(12):370–373. https://doi.org/10.1073/pnas.4.12.370. PMID: 16576330; PMCID: PMC1091498
17. Mifflin MD, St Jeor ST, Hill LA, Scott BJ, Daugherty SA, Koh YO (1990) A new predictive equation for resting energy expenditure in healthy individuals. Am J Clin Nutr 51(2):241–247. https://doi.org/10.1093/ajcn/51.2.241. PMID: 2305711
18. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1

Alternate Tiny Encryption Algorithm: A Modified Tiny Encryption Algorithm for Improved Data Security

Mehak Gupta, Nimit Agrawal, and Manas Ranjan Prusty

Abstract In this era of Industry 4.0, securing file data is very crucial in today's environment with respect to data transfer of Internet of Things (IoT) devices. Over the years with the evolution of technology and file storage systems, many algorithms have been used for encryption and decryption processes for securing the file data, each with its methodologies, advantages, and limitations. An efficient algorithm has very few limitations thus making it a top choice for usage. In this paper, we have proposed a symmetric key cryptographic algorithm called the Alternate Tiny Encryption Algorithm (ATEA) focusing on a strong approach for safekeeping of the file data and minimizing the weak points of the existing Tiny Encryption Algorithm (TEA) and Extended Tiny Encryption Algorithm (XTEA). The Alternate Tiny Encryption Algorithm (ATEA) is a Feistel cipher that utilizes mixed algebraic group operations. The algorithm is simple enough to incorporate into practically any computer program and can be quickly translated into a variety of languages. It uses a unique key generation technique making the encryption of file data more secure.

Keywords Cryptography · Tiny encryption algorithm · Encryption · Decryption

M. Gupta (✉) · N. Agrawal · M. R. Prusty
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu 600127, India
e-mail: mehak.ga25@gmail.com

M. R. Prusty
Centre for Cyber Physical Systems, Vellore Institute of Technology, Chennai, Tamil Nadu 600127, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_32

1 Introduction

With the increase in digitalization of data, though managing and assembling data has become quite a lot easier, it has however increased the risks of security breach. Some vital information transferred within a workspace, across businesses, between divisions of an organization and other external bodies and institutions occasionally falls into the hands of illegitimate parties who could alter the information. If no security measures are adopted, such data and other confidential information will likely be subject to threats such as manipulation, theft, corruption, dispute, and break-in, resulting in substantial harm to the individual or organization. Alongside digitalization, the inception of Industry 4.0 has made a significant role in automation which involves various IoT devices. Data transfer in such devices needs to be upgraded with better and faster secure policies. Hence, this motivates researchers in improving the encryption and decryption algorithms and processes too.

The idea of encryption/decryption algorithms is used where encryption is scrambling the plain text into cipher text and vice versa for decryption. For this process, we can use any of the two types of cryptographic keys present: symmetric or asymmetric. Symmetric or secret-key encryption algorithms encrypt and decrypt using the same symmetric key. Whereas asymmetric or public-key encryption algorithms encrypt and decrypt data two different but related keys. Symmetric algorithms are substantially faster than asymmetric algorithms and can accommodate thousands of keys with very little computing overhead [1]. Therefore, for our algorithm, encoding/decoding is done with the help of a secret key.

A block cipher's strength can be evaluated using characteristics like avalanche, completeness, and statistical independence [2]. The Butterfly Effect, commonly referred to as the Avalanche Effect, is a significant characteristic of cryptographic algorithms. It implies that altering just one bit in the plain text or the key should cause a significant change in the result. We verified the effectiveness of our proposed algorithm by calculating the avalanche effect. Completeness is a required feature of encryption, which means that each bit of the cipher text/output block must depend on each bit in the plaintext [3].

ATEA (Alternate Tiny Encryption Algorithm) is designed as a Feistel cipher. A Feistel network employs a round function that takes two inputs, a data block and a key, and gives one data block-sized output. The round function is applied to the half of the data to be encrypted in each round, and the output is exclusive ORed with the other half. This process is repeated a specific number of times, with the encrypted data as the final outcome. The process of decryption in the Feistel cipher is almost similar but not exactly the same. Instead of starting with the plaintext, the cipher text is sent into the Feistel structure and then the process thereafter is the same as that of encryption. The only difference in the case of decryption is that the sub keys used in encryption are utilized in the reverse order [4].

2 Related Work

There are many algorithms developed for cryptography methods. Mentioned below are the previously existing methods which have been taken into account for developing our algorithm [5].

David Wheeler and Roger Needham developed the Tiny Encryption Algorithm (TEA). It is a 64-round Feistel block cipher with 64-bit block size and 128-bit key that consist of four 32-bit words K[0], K[1], K[2] and K[3]. It uses XOR (Exclusive OR), ADD, SUB and SHIFT operations.

Although TEA is simple and cryptographically strong, it still has some weaknesses. Each key is equivalent to three others resulting in an effective key size of only 126 bits [6] which led to an attack on the gaming console of Microsoft's Xbox where TEA was used as a hash function [7]. It lacks a key schedule algorithm, making it vulnerable to a related-key attack which necessitates 2^{23} chosen plaintexts under a related-key pair and has a 2^{23} time complexity [8]. Because of these flaws, the XTEA (Extended Tiny Encryption Algorithm) cipher was designed.

The first version of XTEA was presented in 1997 by Wheeler and Needham, three years after TEA was first introduced [9]. XTEA is a block cipher that employs a 64 round Feistel structure with a 128-bit key as in TEA. Changes to the key schedule algorithm were introduced in XTEA. Instead of having the subkeys placed in a predetermined order, they were introduced as subkey A and subkey B, which are chosen using two bits of the variable "sum". In addition, a shift of 11 was also added to the key schedule to give the subkeys a more erratic pattern. Rearranged addition, shift and XOR operations are among the other modifications that were made in XTEA.

3 Proposed System

Encryption/Decryption algorithms deal with the data set to be encrypted, a set of keys to encrypt the data set into ciphertext and further, the keys are to be used to decrypt the ciphertext into the original data set.

The proposed algorithm—ATEA is a cryptographically strong algorithm to protect file data. It is an improved version of TEA [5]. ATEA is a Feistel structure having 32 rounds (16 cycles) for each encryption and decryption. The proposed algorithm consists mainly of four modules—Key generation, Subkey generation, Encryption and Decryption.

3.1 Key Generation

The encryption key will be produced automatically using the user's random mouse pointer coordinates on the screen, making each key unique. Then the key is further converted into a hash key of 128 bits using a cryptographic hash function called salted MD5 (Message-Digest algorithm 5) [10]. MD5 is a simple and fast hashing algorithm that is added with salt to force its uniqueness, improve complexity without increasing user requirements, and protect against password attacks [11]. Salting a hash means adding a random string of letters and numbers, called a salt, before hashing it [12]. Hashing is a one-way process that forms a fixed-length binary sequence that is difficult to invert computationally. Hence, the attacker cannot retrieve the master key from the hash key, which makes the system more secure.

3.2 Subkey Generation

The hash key generated using salted MD5 can now be used to generate subkeys that are to be inputted for encryption and decryption processes. We have proposed a new technique for this purpose that is easy to understand and implement. The proposed technique generates 64 unique subkeys which ensure that the subkeys are uniquely chosen for each cycle.

First, the 128-bit hash key is divided and split into four 32-bit pieces- K[0], K[1], K[2] and K[3]. Then an array P of size 16 (P0, P1, …, P14, P15) is initialized with elements having 32-bit value each. Now the subkeys (S[0], S[1], …, S[62], S[63]) are generated by XOR operation between the elements of the P array and the four hash key pieces, as shown in Fig. 1, which will be used for the encryption and decryption process.

3.3 Encryption

The file text to be encrypted is first split into two halves vv[0] and vv[1]. ATEA contains two rounds for each encryption or decryption cycle. In total, encryption contains 32 rounds that are 16 cycles. Round one and following odd rounds operate on vv[0]. Round two and following even rounds operate on vv[1]. The round function which consists of some basic operations like Bitwise shift, XOR and addition is applied to one half using subkeys. The round function output is added to the other half creating repeated mixing of the data and all the bits of the subkey. The two halves are then exchanged. The same pattern is observed in each round using unique subkeys generated in the previous step. Figure 2 shows an ATEA encryption cycle consisting of two rounds.

$S[0] = P0 \oplus K[0]$	$S[16] = P4 \oplus K[0]$	$S[32] = P8 \oplus K[0]$	$S[48] = P12 \oplus K[0]$
$S[1] = P0 \oplus K[1]$	$S[17] = P4 \oplus K[1]$	$S[33] = P8 \oplus K[1]$	$S[49] = P12 \oplus K[1]$
$S[2] = P0 \oplus K[2]$	$S[18] = P4 \oplus K[2]$	$S[34] = P8 \oplus K[2]$	$S[50] = P12 \oplus K[2]$
$S[3] = P0 \oplus K[3]$	$S[19] = P4 \oplus K[3]$	$S[35] = P8 \oplus K[3]$	$S[51] = P12 \oplus K[3]$
$S[4] = P1 \oplus K[0]$	$S[20] = P5 \oplus K[0]$	$S[36] = P9 \oplus K[0]$	$S[52] = P13 \oplus K[0]$
$S[5] = P1 \oplus K[1]$	$S[21] = P5 \oplus K[1]$	$S[37] = P9 \oplus K[1]$	$S[53] = P13 \oplus K[1]$
$S[6] = P1 \oplus K[2]$	$S[22] = P5 \oplus K[2]$	$S[38] = P9 \oplus K[2]$	$S[54] = P13 \oplus K[2]$
$S[7] = P1 \oplus K[3]$	$S[23] = P5 \oplus K[3]$	$S[39] = P9 \oplus K[3]$	$S[55] = P13 \oplus K[3]$
$S[8] = P2 \oplus K[0]$	$S[24] = P6 \oplus K[0]$	$S[40] = P10 \oplus K[0]$	$S[56] = P14 \oplus K[0]$
$S[9] = P2 \oplus K[1]$	$S[25] = P6 \oplus K[1]$	$S[41] = P10 \oplus K[1]$	$S[57] = P14 \oplus K[1]$
$S[10] = P2 \oplus K[2]$	$S[26] = P6 \oplus K[2]$	$S[42] = P10 \oplus K[2]$	$S[58] = P14 \oplus K[2]$
$S[11] = P2 \oplus K[3]$	$S[27] = P6 \oplus K[3]$	$S[43] = P10 \oplus K[3]$	$S[59] = P14 \oplus K[3]$
$S[12] = P3 \oplus K[0]$	$S[28] = P7 \oplus K[0]$	$S[44] = P11 \oplus K[0]$	$S[60] = P15 \oplus K[0]$
$S[13] = P3 \oplus K[1]$	$S[29] = P7 \oplus K[1]$	$S[45] = P11 \oplus K[1]$	$S[61] = P15 \oplus K[1]$
$S[14] = P3 \oplus K[2]$	$S[30] = P7 \oplus K[2]$	$S[46] = P11 \oplus K[2]$	$S[62] = P15 \oplus K[2]$
$S[15] = P3 \oplus K[3]$	$S[31] = P7 \oplus K[3]$	$S[47] = P11 \oplus K[3]$	$S[63] = P15 \oplus K[3]$

Fig. 1 Subkey generation technique

Pseudo Code for Encryption Using ATEA

```
Initialize sum = 0 and delta = 0x9E3779B9
Declare a byte array vv to store the left and right halves of the file data
for j = 0 to 16
Initialize i = 0
sum += delta
vv[0] += ((((vv[1] << 4) + S[i]) ^ ((vv[1] >> 5) + S[i + 1])) ^ (vv[1] + sum))
```

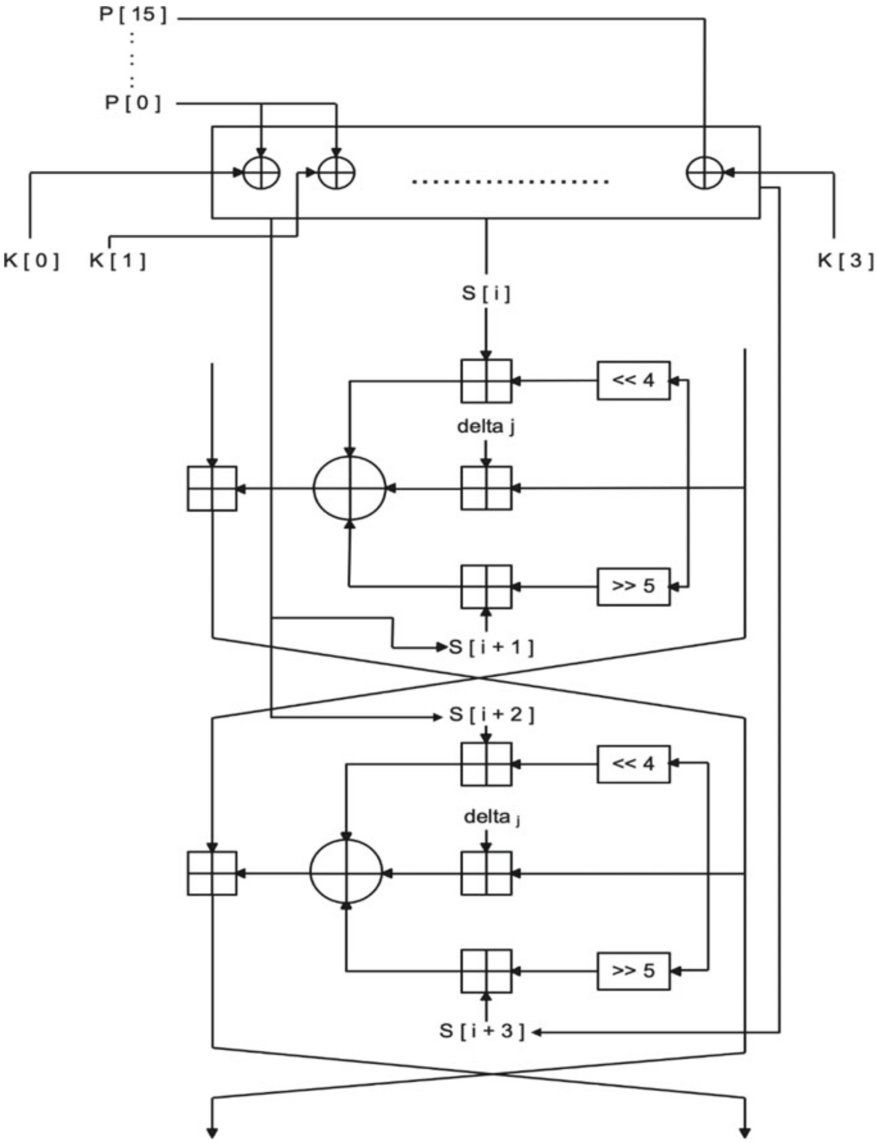


Fig. 2 One cycle of encryption using ATEA

7. vv[1] += ((((vv[0] << 4) + S[i + 2]) ^ ((vv[0] >> 5) + S[i + 3])) ^ (vv[0] + sum))
8. i = i + 4
9. End for loop
10. Swap left and right halves

3.4 Decryption

It simply involves the inverse operations in reverse order, i.e. 16 cycles of addition and use of the subkeys in the opposite order of encryption. Figure 3 shows an ATEA decryption cycle consisting of two rounds.

Pseudo Code for Decryption Using ATEA

```
Initialize delta = 0x9E3779B9 and sum = 16 * delta
Declare a byte array vv to store the left and right halves of the file data
for j = 0 to 16
    Intialize i = 63
    $vv[1] -= ((((vv[0] << 4) + S[i - 2]) \wedge ((vv[0] >> 5) + S[i - 3])) \wedge (vv[0] + sum))$
    $vv[0] -= ((((vv[1] << 4) + S[i]) \wedge ((vv[1] >> 5) + S[i - 1])) \wedge (vv[1] + sum))$
    sum -= delta
    $i = i - 4$
End for loop
Swap left and right halves
```

We implemented the proposed ATEA algorithm to make a web application whose architecture is shown in Fig. 4. The application aims to maintain the secrecy of the data files that the client sends to it [13].

Aside from cryptography, the application also supports the upload of encrypted files. This feature allows a user to easily download and install an encrypted file on the server. The files are stored in separate encrypted partitions for each user. Also, the server sends a message to the recipient when the link is sent. Only encrypted files are allowed to be uploaded on the server and links to the uploaded files are automatically generated and returned to the user. Also, the encrypted files uploaded on the server can be easily shared with any other registered users of the application using the share option. The link of the encrypted file on the server is sent as a message to the recipient user which can be easily downloaded by a single click and decrypted if the recipient has the respective key [14].

4 Results

ATEA is a simple algorithm to implement. The encrypting and decrypting operations are nearly similar, needing only a key schedule reversal. As a result, the size of the code needed to implement ATEA is nearly halved.

The number of rounds in a system depends on a trade-off between efficiency and security. More rounds provide a more security. More rounds, on the other hand, imply inefficiently slow encryption and decryption. Hence, we have reduced the number of rounds to 32 to make these processes fast. And for increasing the security, we

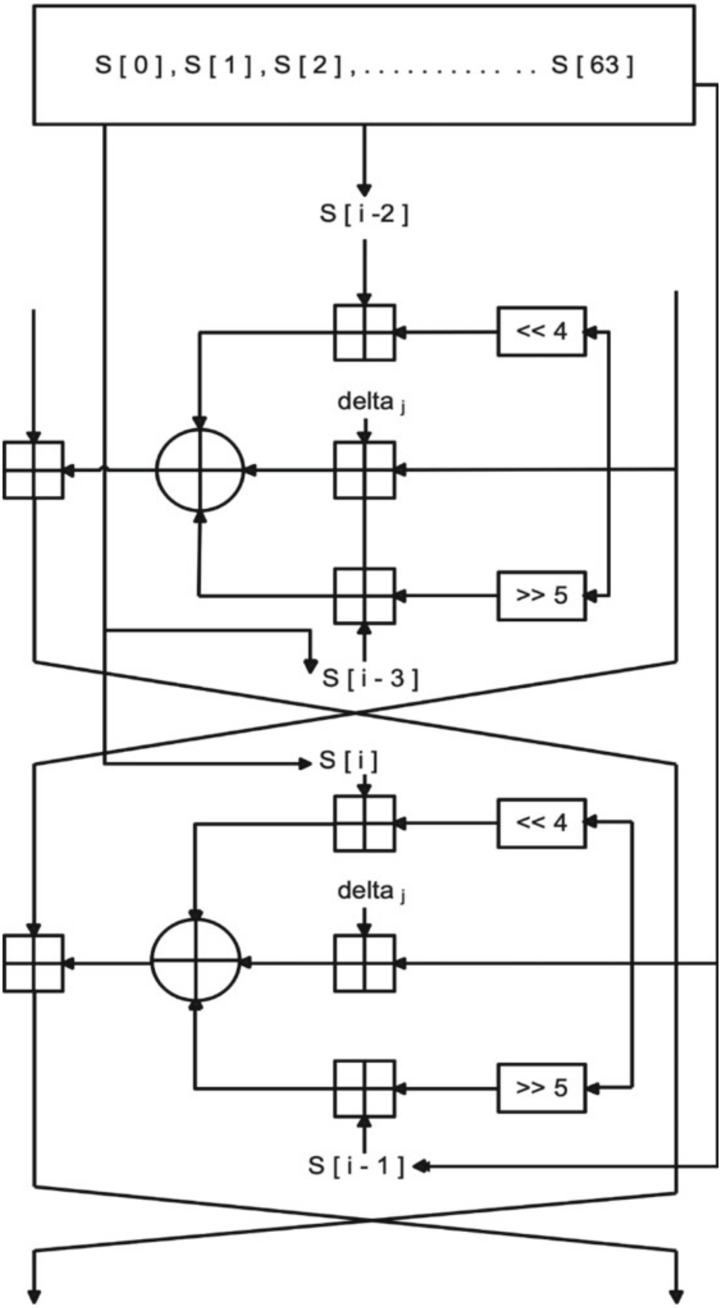


Fig. 3 One cycle of decryption using ATEA

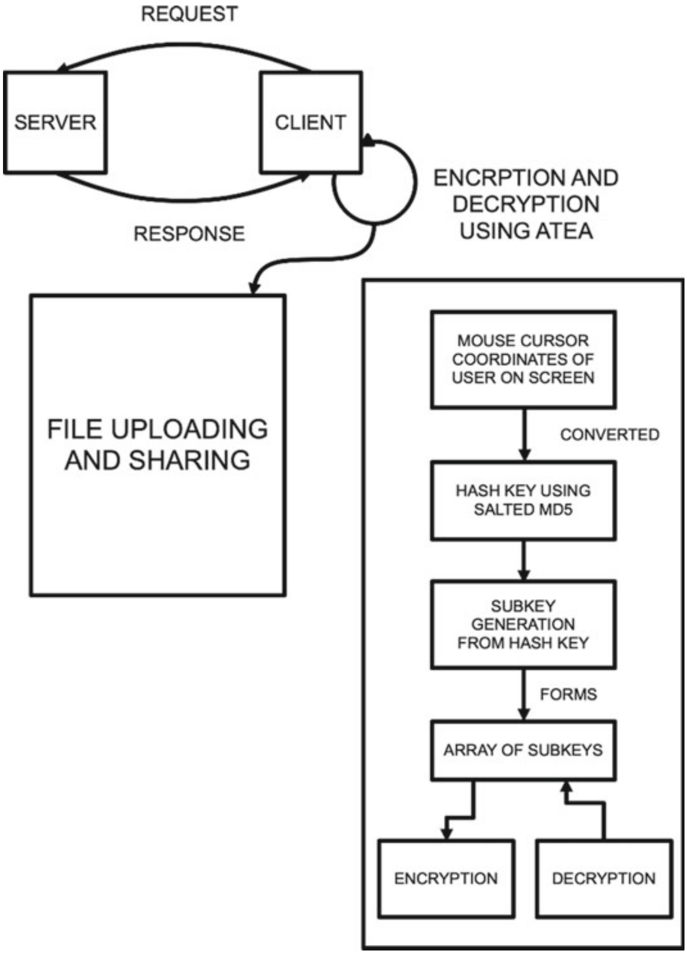

REQUEST
SERVER
CLIENT
RESPONSE
ENCRPTION AND DECRYPTION USING ATEA
FILE UPLOADING AND SHARING
MOUSE CURSOR COORDINATES OF USER ON SCREEN
CONVERTED
HASH KEY USING SALTED MD5
SUBKEY GENERATION FROM HASH KEY
FORMS
ARRAY OF SUBKEYS
ENCRYPTION
DECRYPTION

Fig. 4 Web application architecture

have added the key and subkey generation module which ensures that the subkeys are independently chosen for each round.

TEA employs a basic key schedule that breaks down the 128-bit key into four 32-bit chunks and uses them again and again in consecutive rounds which makes it prone to related-key attacks whereas ATEA has a key generated from mouse cursor coordinates which are further converted to a hash key of 128 bits and further divided into 64 unique keys using the proposed subkey generation technique thus making it more random and acting as an added layer of security.

XTEA uses a subkey generation technique that provides better security as compared to TEA [9]. The subkeys generation technique of XTEA is simple and easy to crack as compared to our proposed key and subkey generation technique.

We have summarized some similarities and differences between our proposed algorithm ATEA and the existing algorithm TEA and XTEA for encryption and decryption as shown in Table 1.

Tables 2 and 3 show the approximate encryption and decryption times of our algorithm and the existing TEA and XTEA security algorithms for a file size of 12 kilobytes with different key sizes.

We verified the effectiveness of our algorithm by calculating the Avalanche Effect that refers to the number of modified bits in a ciphertext divided by the number of bits in the ciphertext. An avalanche greater than 50% must always be satisfied by a good cipher [15]. For our proposed algorithm approximately 50% of ciphertext bits differ after every round as shown in Fig. 5. Thus, using ATEA for encryption, the input bits get thoroughly mixed in each round resulting in many changes in output bits making cryptanalysis very difficult. Also, by calculating the average of the avalanche effect in

Table 1 Comparison between ATEA, TEA and XTEA algorithms

Characteristics	ATEA	TEA	XTEA
Rounds/cycles	32 rounds (16 cycles)	64 rounds (32 cycles)	64 rounds (32 cycles)
Master key size	128 bits	128 bits	128 bits
Key generation	Mouse cursor coordinates	Not unique	Not unique
Key used in each round	Unique	Repeating	Unique
Subkey scheduling algorithm	Yes	No	Yes
Security	Secure and strong	Less secure and weak	Secure and strong
Structure	Feistel cipher	Feistel cipher	Feistel cipher
Performance	Better than TEA	Relatively low	Better than TEA

Table 2 Time taken for encrypting a 12 KB file

Key size (in bits)	Encryption time (in ms)		
	TEA	XTEA	ATEA
32	1.173	2.287	1.329
48	1.178	2.572	1.201
64	1.248	2.089	1.092
96	1.208	2.32	1.301
128	1.067	2.301	1.287
160	1.137	2.608	1.053
192	1.39	2.327	1.129
240	1.439	2.866	1.306

Table 3 Time taken for decrypting a 12 KB file

Key size (in bits)	Decryption time (in ms)		
	TEA	XTEA	ATEA
32	1.127	2.249	1.360
48	1.195	2.572	1.239
64	1.241	2.084	1.145
96	1.226	2.299	1.297
128	1.029	2.265	1.301
160	1.093	2.264	1.049
192	1.363	2.278	1.159
240	1.402	2.827	1.401

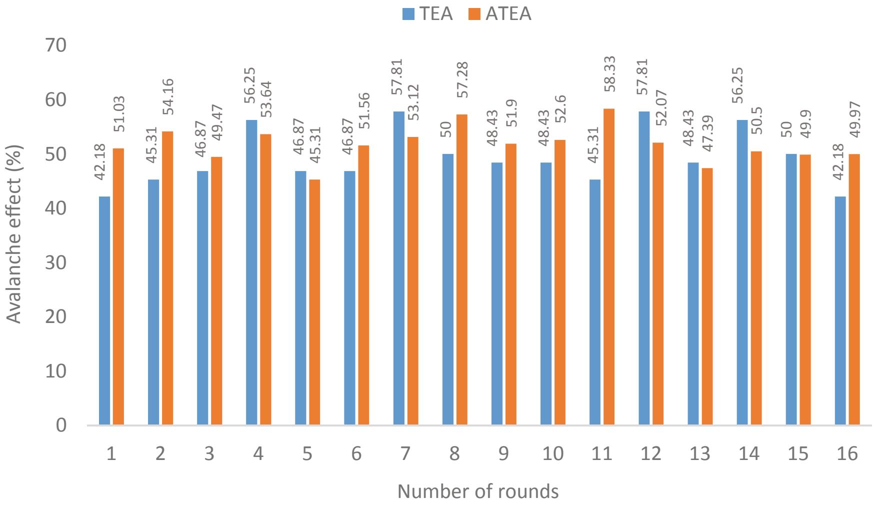


Fig. 5 Avalanche effect comparison in ATEA and TEA

TEA which is approximately 49.32% and ATEA approximately 51.77% from Fig. 5, we can conclude that the ATEA shows a better avalanche effect than TEA.

5 Conclusion

This paper proposes an improved technique for data encryption, working on the limitations of the previous versions. It shows that there is no significant degradation in the considered cryptographic quality by using ATEA over standard TEA and XTEA.

ATEA provides strong and secure encryption/decryption. Rather than using the primary key directly in the encryption/decryption process for the file data one can use it as a master key to derive unique subkeys and use them for the actual cryptographic processing. Hence ATEA can be considered as a good cryptographic algorithm to be implemented.

Although the proposed method offers robust and safe encryption and decryption, it might eventually be improved by increasing performance and decreasing complexity.

References

1. Liu S, Gavrylyako O, Bradford P (2004) Implementing the TEA algorithm on sensors. In: Proceedings of the 42nd annual Southeast regional conference, pp 64–69
2. Sharma M, Arora JB (2017) Cryptography and its desirable properties in terms of different algorithm. IITM J Manag IT 75–81
3. Ramanujam S, Karuppiah M (2011) Designing an algorithm with high avalanche effect. IJCSNS Int J Comput Sci Netw Secur 106–111
4. Narendra K, Pareek VP, Sud K (2010) Block cipher using 1D and 2D chaotic maps. Int J Inf Commun Technol 2(3):244
5. Shoeb M, Gupta VK (2019) A crypt analysis of the tiny encryption algorithm in key generation. Int J Commun Comput Technol 01(01)
6. Wheeler D, Needham R (2004) TEA, a tiny encryption algorithm. In: Proceedings of the fast software encryption: second international workshop. Lecture notes in computer science, vol 1008, pp 363–366
7. Moon MD, Hwang K, Lee W, Lee S, Lim J (2002) Impossible differential cryptanalysis of reduced round XTEA and TEA. In: Daemen J, Rijmen V (eds) FSE 2002. LNCS, vol 2365. Springer, pp 49–60
8. Kelsey J, Schneider B, Wagner D (1997) Related-key cryptanalysis of 3-way, Biham-DES, CAST, DES-X, NewDES, RC2, and TEA. In: Proceedings of the first international conference on information and communication security. Lecture notes in computer science, vol 1334, pp 233–246
9. Lee E, Hong D, Chang D, Hong S, Lim J (2006) A weak key class of XTEA for a related-key rectangle attack. In: Nguyen PQ (ed) VIETCRYPT 2006. LNCS, vol 4341. Springer, pp 286–297
10. Hossain MA, Islam MK, Dasand SK, Nashiry MA (2012) Cryptanalyzing of message digest algorithms MD4 and MD5. Int J Crypt Inf Secur (IJCIS) 2(1)
11. Sriramya P, Karthika RA (2015) Providing password security by salted password hashing using Bcrypt algorithm. ARPN J Eng Appl Sci 10(13)
12. Boonkrong S (2012) Security of passwords. Inf Technol J 112–117
13. Benson E, Marcus A, Karger D, Madden S (2010) Sync kit: a persistent client-side database caching toolkit for data intensive websites. In: Proceedings of the 19th International conference on world wide web, USA, Apr 2010
14. Zaman RU, Utkarsh P, Tanwani L, Kumar A (2017) Data encryption and file sharing. Int Res J Eng Technol (IRJET) 04(05)
15. Ramanujam S, Karuppiah M (2011) Designing an algorithm with high Avalanche effect. Int J Comput Sci Netw Secur (IJCSNS) 11(1)

Crystal Clear Analysis of Open–Source Automation Platforms

Kiran Jadhav, Mangesh Nikose, and Sagar Shinde

Abstract The advantages of free software are innumerable, and it is highly encouraging to witness many domestic computerization sources that provide free and efficient software program to Internet of Things researchers across the globe. The developers behind such computerization sources have worked hard to create a solid system foundation that anyone can use. Moreover, these sources are freely used by the people to obtain individual responses. Similar to most of the other software systems, for developing an advanced solution, a strong community that is willing to return it to its initial position is required. These characteristics have motivated to compile and compare a greater number of open-source automation software, which is considered as the most exciting domain of IoT. The user can utilize the open source software based on the suitability of the system available and preferred, programming language support for the system, number of users, desired security level, service protocol, and so on.

Keywords OpenHAB · Home assistant · Homebridge.io · Jeedom · Eventghost · Pidome · IFTTT · FHEM · Calaos

K. Jadhav (✉)
Sandip University, Nashik, Maharashtra, India
e-mail: jadhavkiran32@gmail.com

M. Nikose
Department of Electrical and Electronics Engineering, Sandip University, Nashik, Maharashtra, India
e-mail: mangeshnikose@sandipuniversity.edu.in

S. Shinde
PCET's-NMVP's Nutan College of Engineering and Research, Pune, Maharashtra, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_33

1 Introduction

Home automation software is a software that allows to control and display basic home and office appliances using a PC. Home automation used to be limited to switching the lights and other appliances. However, the possibilities are extensive, allowing customers to create a wireless community, automate Television and Hi-Fi to record both audio and video while the user is away, set up an answering device, and create a climate station by integrating a plethora of different domestic automation technologies into a single advanced technology. According to the UK Department of Trade and Industry [1], a smart home is a place of residence incorporating information and communication technology to connect the key electrical appliances and services that allows them to be remotely controlled, monitored, or accessed. Shareholder network protocols are currently used by many home automation systems. The protocols employed may be exclusive to the business that advanced the machine. The software industry may also prefer this approach because it effectively connects the buyer to their products. However, these can be extensively harmful to the consumer of the home automation device. As a result, it is critical to evaluate a home automation device to ensure that it is built on open protocols.

2 Literature Summary of Open Source Software

The proposed study has collected and summarised the available open source software resources in terms of the analysis of most important open-sources of home automation software, including their main features such as system requirements, development language, messaging service, security, speed, and web interface availability, as individual resources and links [2]. The following Table 1 compares some of the most popular open-source home automation software.

2.1 OpenHAB

OpenHAB is an open source domestic automation software program available in Java Script. It is set up on premises and integrated to device, which can offer special companies. From 2019, approximately three hundred ties are considered as OSGi modules. Actions, inclusive of latching of lighting are caused through policies, speech instructions or controls at the openHAB client edge. The openHAB assignment started out in 2010 and about 2013, the middle functionality has become a legitimate mission of the Eclipse Foundation under the name eclipse smart home. OpenHAB is primarily based on Eclipse Smart Home and remains as the undertaking for the improvement of bindings. As per Black Duck Open Hub, it remains miles advanced by certainly using one from largest open-source squads present across the globe.

Table 1 Comparison of open source software

Open source software	System supports	Language	Nos. of user	Security	Speed	Service protocol	License
Open-HAB [3]	Windows, MAC OS, Linux, Raspberry pi PINE 64	Java with OSGI, Apache, Karaf	1500	Secure	Very fast	MQTT 2.0	Copyleft
Home assistant [4]	Raspberry pi 3	Python 3 (Polymer and YAML)	1000	Secure	Fast	MQTT	Apache 2.0
Jeedom [5]	Hardware hubs and paid support	PHP (Android and IOS)	1000+	Secure	Intermediate	MQTT	GPL V2
Io-Broker [6]	ARM with GNU/Linux	JavaScript with node.js and Redis	285+	Secure (Https)	Fast	MQTT and UPnP	MIT
Google home assistant [7]	Android 5.0+, Google app	Node.js, java, C++, GO	10,000	Highly secure (AES)	Very fast	SMS	Google Inc
AGO Control	Raspberry pi wire	YAML	10+	Secure but lot of contradictions	Satisfactory	AMQP enterprise message	GNU public license
Domoticz [8]	Raspberry pi, UNIX, Apple, Windows	C++	255	Safe	Quite fast	MQTT	GPL V3
FHEM	UNIX, Windows, Fritz Box, Raspberry pi 3	Perl	430	Secure with AES algorithm	Fast	MQTT	GPL V2
Calaos	Wago PLC, Raspberry pi, Cubie Board, Squeezebox	C++	255	Secure	Fast	N.A	GPL V3
Pimatic [9]	Android, IOS	Node.js	N.A	Secure automate HTTPS	Fast enough	MQTT	GNU GPL
Homebridge [10]	Raspberry pi, Linux, IOS, Windows, Docker	Node.js	support class 2 of devices	Highly secure	Fast	MQTT	Apache 2.0
IFTTT	IOS and Android	Node.js, Java	10,000+	Secure	Fast	SMS	Apache 2.0

(continued)

Table 1 (continued)

Open source software	System supports	Language	Nos. of user	Security	Speed	Service protocol	License
Node-RED	Mac, Raspberry pi	Java Script and Node.js	10,000+	Not secure	Fast	MQTT with TLS	Apache 2.0

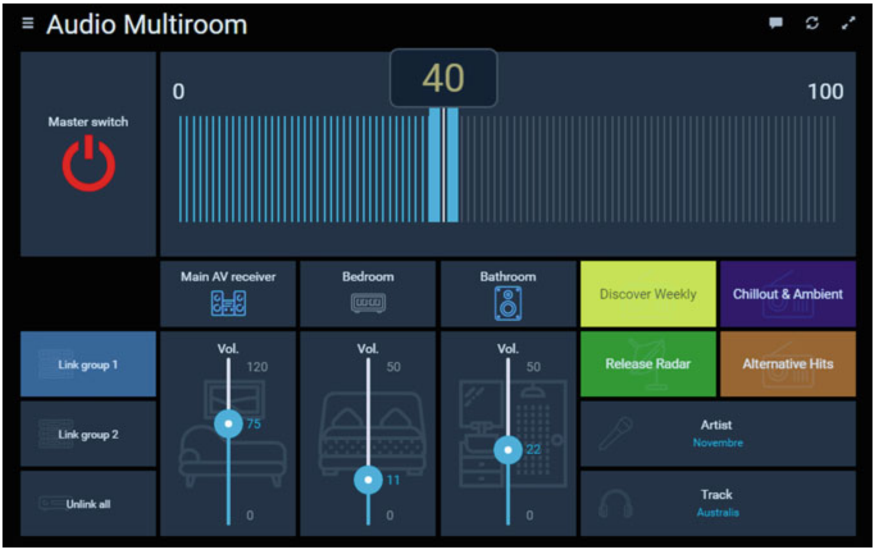


Fig. 1 A glimpse of OpenHAB

It additionally has a lively consumer era [4]. A glimpse of OpenHAB is shown in Fig. 1.

- Merits of OpenHAB [11]:
  - It can be integrated with more than 200 different technologies, systems and thousands of devices.
  - Ease of automation.
  - No cloud required but cloud friendly.
  - It runs on almost all types of system hardwares.

2.2 *Home Assistant*

Home Assistant is an open source robotization programming, which is intended to be the centralized control framework for smart home gadgets with a focus on nearby

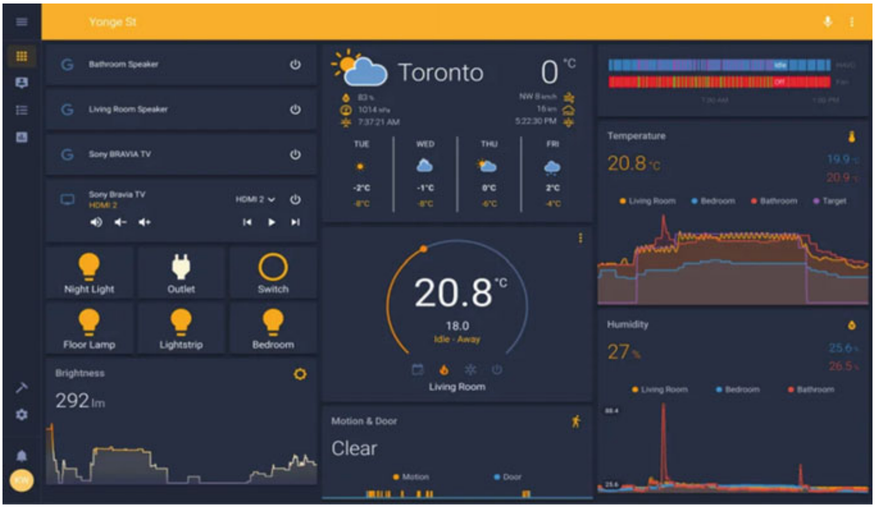


Fig. 2 A glimpse of home assistant

control and security. It can be accessed through an electronic UI, peer applications for Android and iOS, or voice commands for smart devices like Google Assistant and Amazon Alexa. Advanced IoT gadgets, programming applications, and administrations are supported by the measured incorporation segments, which not just incorporate local combinations for nearby network conventions like Bluetooth, MQTT, Zigbee, and Z-Wave, but additionally support for controlling private environments if they give access through a public API for outsider interfaces. When the home assistant programming application is installed on a PC, it serves as a central control framework for home automation. Information from all substances can be used and controlled from scripts trigger mechanizations by utilizing the planning and outline subroutines to control lighting, environment, theatre setups, and apparatus [4]. Figure 2 shows an overview of Home Assistant.

- Merits of Home Assistant [12]:
  - Works with over 1900 devices
  - Powerful automation
  - All data stays local
  - Home energy management.

2.3 *Jeedom*

Jeedom is ideal for the users searching for a reliable and high-end home computerization package at a reasonable price. Due to its small size and smooth brushed metal design, it will fit in any space. It comes with a high-gain Z-Wave+ Antenna, a four-center 1.5 GHZ processor, and an eMMC hard drive. Jeedom is delivered with the Jeedom Service Pack Power programming, which includes administrations that are not available in the open source version. Jeedom's focal point is only locally viable with the Z-Wave+ (module preinstalled and arranged). If it needs to work with different conventions, the necessary modules and equipment should be downloaded and purchased. Adding the required module implies that each unique home robotization convention or potentially network with different ventures can be coordinated to Jeedom. Jeedom is a hub that empowers a centralization of the entire associated gadgets. Jeedom has a web interface to empower the design of a home mechanization framework and a portable application can also be developed for use through a cell phone or tablet. It is also completely adaptable to whatever requirements of using virtual devices, situations, or modules. Whether remotely, locally, by voice, instant message, or touch screen, you can manage your home whenever and from wherever you want and be alerted whenever an event occurs [5]. Figure 3 depicts a glimpse of Jeedom.

- Merits of Jeedom [13]:
  - Compatible with various protocols
  - No need to access external servers to work
  - Flexible and lots of customization settings available.

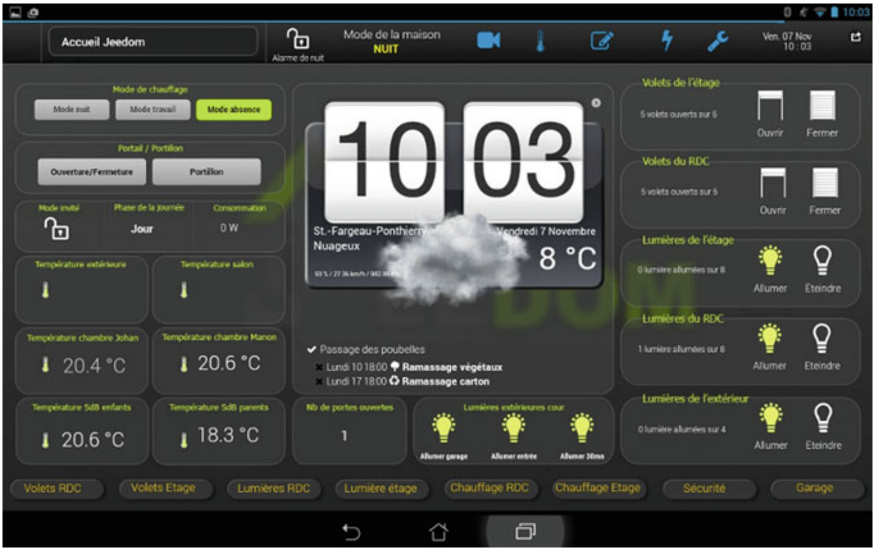
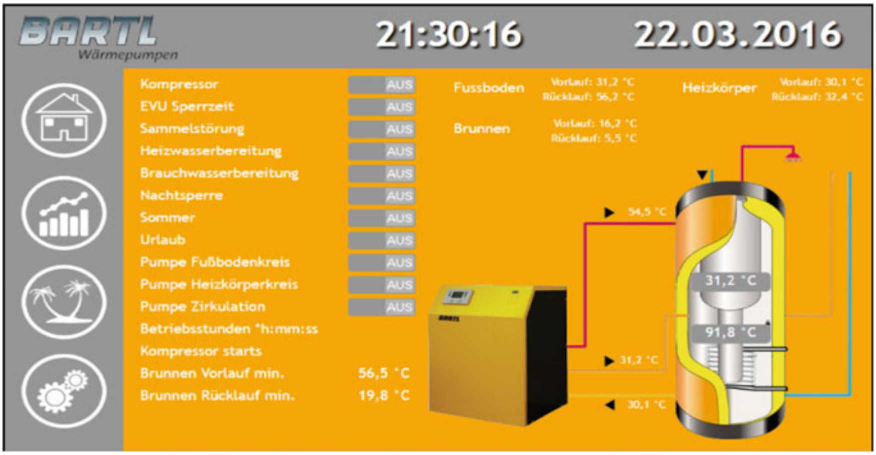


Fig. 3 A glimpse of Jeedom

2.4 IoBroker

IoBroker is open-source software with unique and efficient household components (IoTs), which co-operatively combine into an entire machine (Smart Home System). This gadgets work on their very own but this gives us a complicated data management panel with a graphical user interface, which may be accessed on the neighborhood community with an internet browser. IoBroker is composed of different modules called adapters. Smart devices join IoBroker with adapters. The required adapters can be triggered with one click and further it will be used after a brief setup. IoBroker may be set up on a SoC unmarried-card pc (Raspberry pi, Orange pi, and so on.) or on a computing device PC strolling Windows, Linux. Installing the running device on a Raspberry Pi four USB HDD or SSD is considered as a proper solution. On an old PC, it runs flawlessly under Debian-10, and with java script, one can create automation approaches, connect adapters, and so on. It enables the creation of customized and manipulated strategies. For this reason, there are also Blocky, Node-red, and Scenes; one of the most important adapters is MQTT. Since MQTT allows for establishing a low-cost two-way communication, it is an ideal choice for performing data exchange between a server and a microcontroller [6]. Figure 4 depicts an overview of IoBroker.

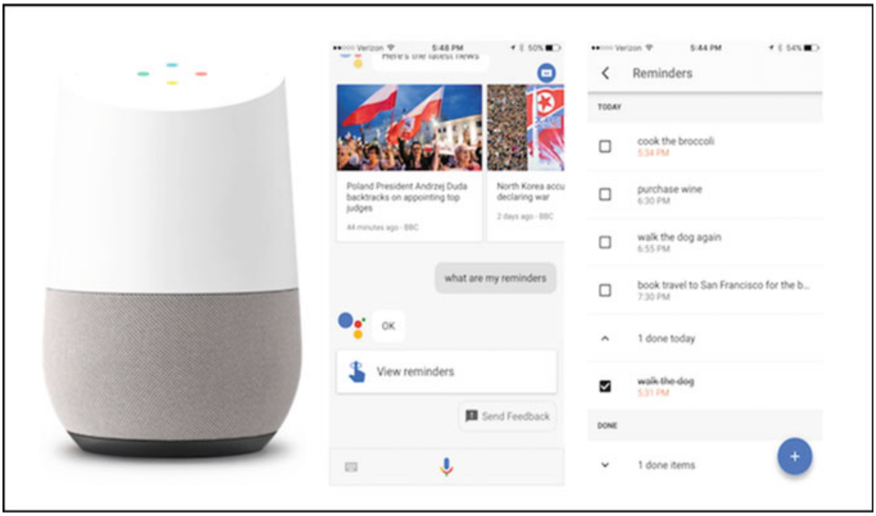
- Merits of IoBroker [14]:
  - Unique graphics and attractive interfaces
  - RUNS ON: Windows, Linux, OSX, Raspberry Pi, ARM or PC.


21:30:16 22.03.2016

Kompressor AUS
EVU Sperrzeit AUS
Sammelstörung AUS
Heizwasserbereitung AUS
Brauchwasserbereitung AUS
Nachtsperre AUS
Sommer AUS
Urlaub AUS
Pumpe Fußbodenkreis AUS
Pumpe Heizkörperkreis AUS
Pumpe Zirkulation AUS
Betriebsstunden *h:mm:ss
Kompressor starts
Brunnen Vorlauf min. 56,5 °C
Brunnen Rücklauf min. 19,8 °C

Fussboden Vorlauf: 31,2 °C Rücklauf: 56,2 °C
Heizkörper Vorlauf: 30,1 °C Rücklauf: 32,4 °C
Brunnen Vorlauf: 16,2 °C Rücklauf: 5,5 °C

Fig. 4 A glimpse of IoBroker



Poland President Andrzej Duda backtracks on appointing top judges

44 minutes ago · BBC

North Korea accu declaring war

2 days ago · BBC

what are my reminders

OK

TODAY

□ cook the broccoli
5:34 PM

□ purchase wine
6:30 PM

□ walk the dog again
6:55 PM

□ book travel to San Francisco for the b...
7:30 PM

1 done today

☑ ~~walk the dog~~
5:31 PM

DONE

1 done items

Fig. 5 A glimpse of google home assistant

2.5 Google Home Assistant

Google Assistant uses speech instructions, speech based search, and speech based control devices, and removes some of the obligations after user state “OK Google” or “Hey Google” phrases. It is intended to respond user in dialogue exchanges. Google Assistant confers command to smart devices and smart home [7]. Google Home Assistant glimpse is shown in Fig. 5.

2.6 AGO Control

AGO control is a framework for tool management. The primary goal is to provide a complete automation solution. It is also used in other applications including agriculture. Previously, manipulate makes use of an AMQP enterprise message bus as a backend and a low footprint protocol, which is readable by users and auto readers, a modern and modular structure, cloud features, and much more. AGO manipulate boasts excellent performance and even works on embedded devices such as Raspberry Pi and multi-plug computers such as Guruplug, Pogoplug, and Sheevaplug. AGO control provides assistance for many peripherals with different protocols, they are 433 MHz trancievers, KNX, Z-Wave, X10, EnOcean, 1wire, Dreambox/Enigma2, Asterisk PBX, Chromoflex USP3 RGB LED dimmer, Onkyoe ISCP AVR, APC Power Distribution Unit, DMX Interfaces by the OLA, Phillips TV units, IRTrans Ethernet purchaser infrared blaster, Webcam guide, Rain8net irrigation controller,

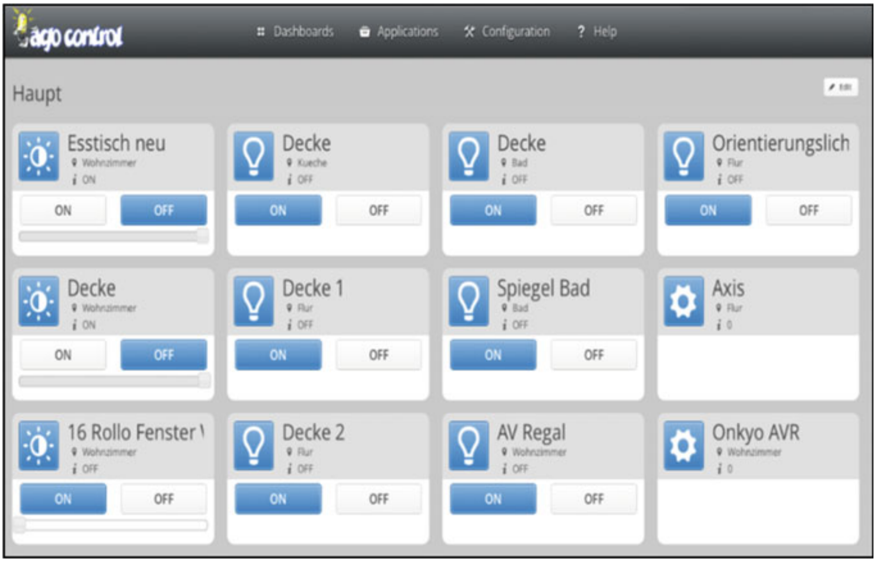


Fig. 6 A glimpse of AGO control

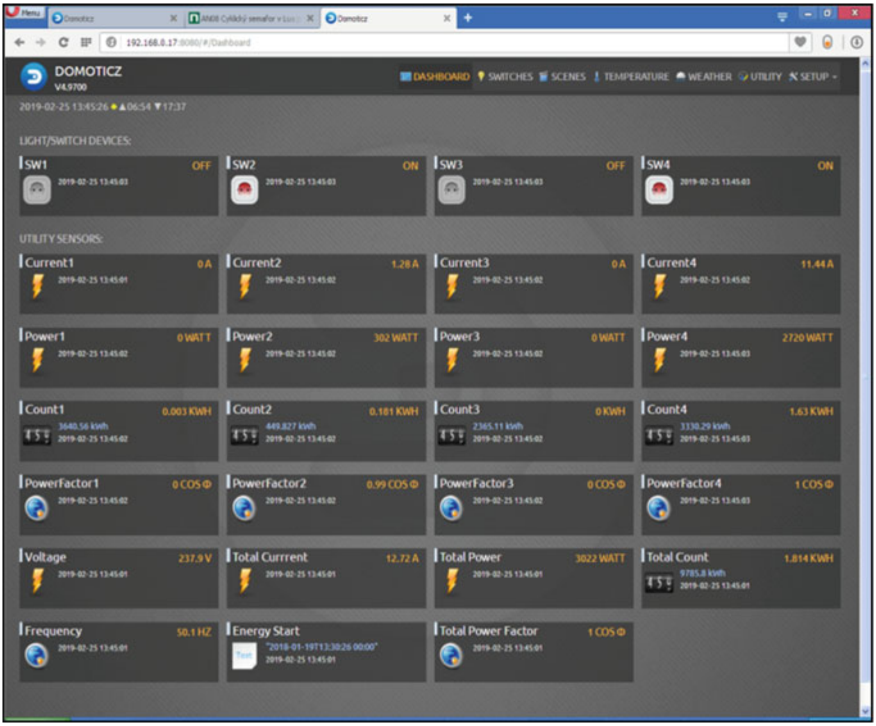
Webcam guide, and BlinkM LED suit. In the past, it was simple to add a list of user suitable tool drivers. Figure 6 shows a glimpse of AGO control.

2.7 Domoticz

Domoticz is an absolutely compact home automation device that assist user to supervise and set various devices, like lighting, switches, different devices which sense or measure different parameters like warmth, precipitation, air pressure, radiation, radiation, energy usage/manufacturing, fuel intake, water consumption, etc. Notifications/indicators can also be sent to any mobile tool [8]. A glimpse of Domoticz is shown in Fig. 7.

2.8 FHEM

FHEM (TM) is a perl-based GNU-GPL server for developing smart homes. FHEM performs common tasks within the home such as controlling shutters/lamps/heating and recording events such as temperature/humidity/energy consumption. The application runs as a server, and it can be directly controlled via net or phone frontends, telnet, or TCP/IP. To use FHEM, the user must have a server (such as a NAS, RPi, PC, or MacMini) with a perl translator and hardware such as the CUL-, EnOcean-,



2019-02-25 13:45:26 ▲06:54 ▼17:37

LIGHT/SWITCH DEVICES:

SW1 OFF 2019-02-25 13:45:03

SW2 ON 2019-02-25 13:45:03

SW3 OFF 2019-02-25 13:45:03

SW4 ON 2019-02-25 13:45:03

UTILITY SENSORS:

Current1 0 A 2019-02-25 13:45:01

Current2 1.28 A 2019-02-25 13:45:02

Current3 0 A 2019-02-25 13:45:02

Current4 11.44 A 2019-02-25 13:45:02

Power1 0 WATT 2019-02-25 13:45:02

Power2 302 WATT 2019-02-25 13:45:02

Power3 0 WATT 2019-02-25 13:45:02

Power4 2720 WATT 2019-02-25 13:45:03

Count1 0.003 KWH 3640.56 kWh 2019-02-25 13:45:02

Count2 0.181 KWH 449.827 kWh 2019-02-25 13:45:02

Count3 0 KWH 2365.11 kWh 2019-02-25 13:45:02

Count4 1.63 KWH 3330.29 kWh 2019-02-25 13:45:03

PowerFactor1 0 COS Φ 2019-02-25 13:45:02

PowerFactor2 0.99 COS Φ 2019-02-25 13:45:02

PowerFactor3 0 COS Φ 2019-02-25 13:45:02

PowerFactor4 1 COS Φ 2019-02-25 13:45:03

Voltage 237.9 V 2019-02-25 13:45:01

Total Currrent 12.72 A 2019-02-25 13:45:01

Total Power 3022 WATT 2019-02-25 13:45:01

Total Count 1.814 KWH 9785.8 kWh 2019-02-25 13:45:01

Frequency 50.1 HZ 2019-02-25 13:45:01

Energy Start "2018-01-19T13:30:26 00:00" 2019-02-25 13:45:01

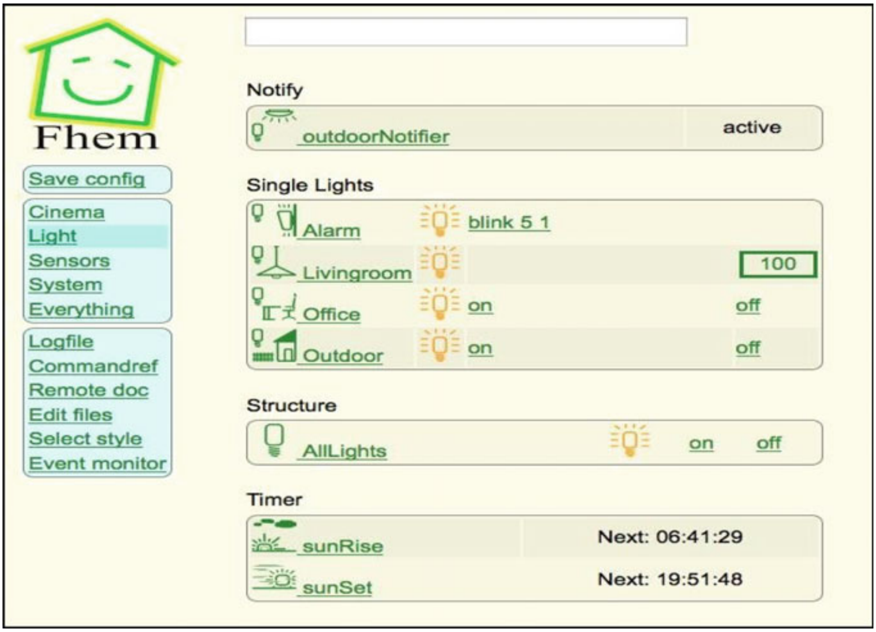
Total Power Factor 1 COS Φ 2019-02-25 13:45:01

Fig. 7 A glimpse of domoticz

and Z-Wave-USB-Stick to gain access to the sensors. Figure 8 depicts a glimpse of FHEM.

2.9 Calaos

Calaos is a powerful open-source domestic automation software program. It functions as a whole set of components to automate house from lighting fixtures to shutter or cameras. It incorporates a complete solution for home automation. The enlightenment foundation libraries are used to strengthen its touch-screen interface. Calaos OS is a whole linux distribution used on its personal. It is primarily based on open-embedded development and provides binary images for a set of machines. Calaos client and server with web-app are mandated for automation. A glimpse of Calaos is shown in Fig. 9.



Notify

outdoorNotifier active

Single Lights

Alarm blink 5 1

Livingroom 100

Office on off

Outdoor on off

Structure

AllLights on off

Timer

sunRise Next: 06:41:29

sunSet Next: 19:51:48

Fig. 8 A glimpse of FHEM

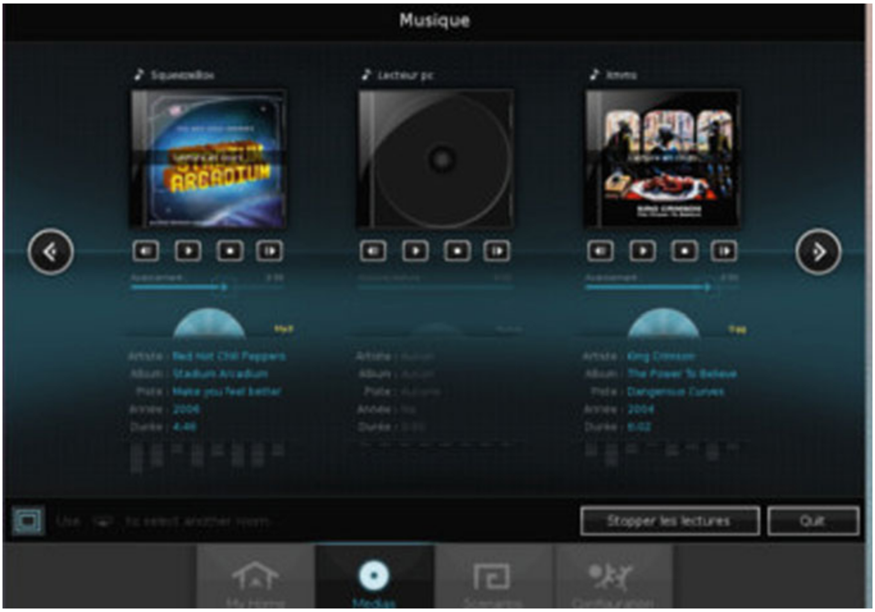


Fig. 9 A glimpse of calaos

2.10 Pimatic

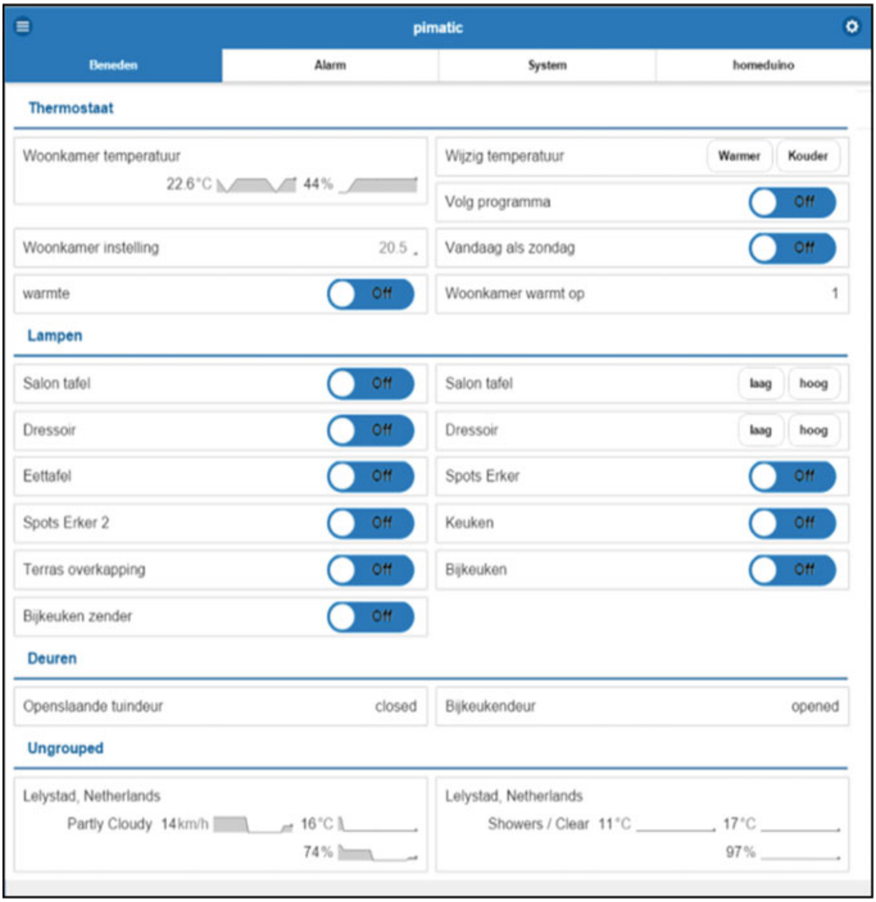
Pitmatic is a domestic automation framework that runs on Node.Js. It presents a common extensible platform to manage home automation tasks. It defines several schemata for developing exceptional home gadgets and sensors, so that any device can be controlled uniformly and are provided with a commonplace interface. Automation duties can be described with the help of policies in the shape of “if this then that”, where the “this” and “that” element may be completely custom designed via plug-in. More information can be found on the rule of thumb page. The mobile frontend plug-in provides a user-friendly internet frontend with sensor evaluation, device manipulation, and rule definition. Express and jQuery Mobile are being used to build the internet interface. Pitmatic can be extended to various plug-ins, which include functions and integrate existing hardware and software [9]. Figure 10 depicts a glimpse of Pimatic.

2.11 Homebridge.io

Homebridge.io is a Node.js based server that acquires less memory and user runs it on a private home community that emulates the iOS HomeKit API. It helps plug-in, which are considered as the network-contributed modules that provide a simple bridge from HomeKit to diverse APIs supplied by using producers of smart domestic devices. A glimpse of Homebridge.io is shown in Fig. 11.

2.12 Node Red

To connect various edge boards, API, and over the internet facilities in a novel and exciting thought, Node-Red is considered as the ideal solution. It provides a browser to simply utilize a large number of nodes, which can be deployed to its runtime with a single click on. Node-Red includes a browser-based glide editor to make it simple to connect flows by using the palette’s large number of nodes. After the software is triggered, signalling can be configured. JavaScript features may be created within the editor by utilizing a wealthy text editor. A built-in library permits to shop beneficial feature, templates or flows for re-use [10]. A glimpse of Node-Red is shown in Fig. 12.



Thermostaat

Woonkamer temperatuur 22.6 °C 44 %

Wijzig temperatuur Warmer Kouder

Volg programma Off

Woonkamer instelling 20.5

Vandaag als zondag Off

warmte Off

Woonkamer warmt op 1

Lampen

Salon tafel Off

Salon tafel laag hoog

Dressoir Off

Dressoir laag hoog

Eettafel Off

Spots Erker Off

Spots Erker 2 Off

Keuken Off

Terras overkapping Off

Bijkeuken Off

Bijkeuken zender Off

Deuren

Openslaande tuindeur closed

Bijkeukendeur opened

Ungrouped

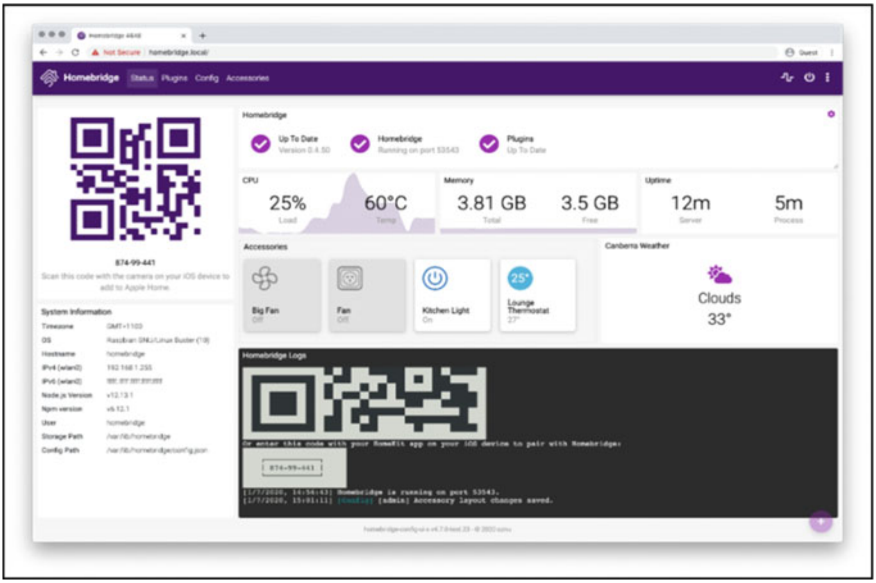
Lelystad, Netherlands
Partly Cloudy 14 km/h 16 °C 74 %

Lelystad, Netherlands
Showers / Clear 11 °C 17 °C 97 %

Fig. 10 A glimpse of pimatic

2.13 IFTTT

If This Then That (IFTTT) offers services to control a reaction to different activities inside the smart home framework. IFTTT has integrated with exclusive carrier carriers to deliver timely announcement to IFTTT and follows instructions that execute the replies. In some instances, command interfaces are considered as public APIs. The applications, referred to as applets are simply created graphically. User can create applications and in any other case manipulate IFTTT with an iOS or internet interface or Android application [15]. A glimpse of IFTTT is shown in Fig. 13.



System Information

Timezone	GMT+1100
OS	Raspbian GNU/Linux Buster (10)
Hostname	homebridge
IPv4 (wlan0)	192.168.1.255
IPv6 (wlan0)	[illegible]
Node.js Version	v12.13.1
Npm version	v6.12.1
User	homebridge
Storage Path	/var/lib/homebridge
Config Path	/var/lib/homebridge/config.json

```
Or enter this code with your HomeKit app on your iOS device to pair with Homebridge:

    | 874-99-441 |

[1/7/2020, 14:54:43] Homebridge is running on port 53543.
[1/7/2020, 15:01:11] [Config] [admin] Accessory layout changes saved.
```

Fig. 11 A glimpse of Homebridge.io

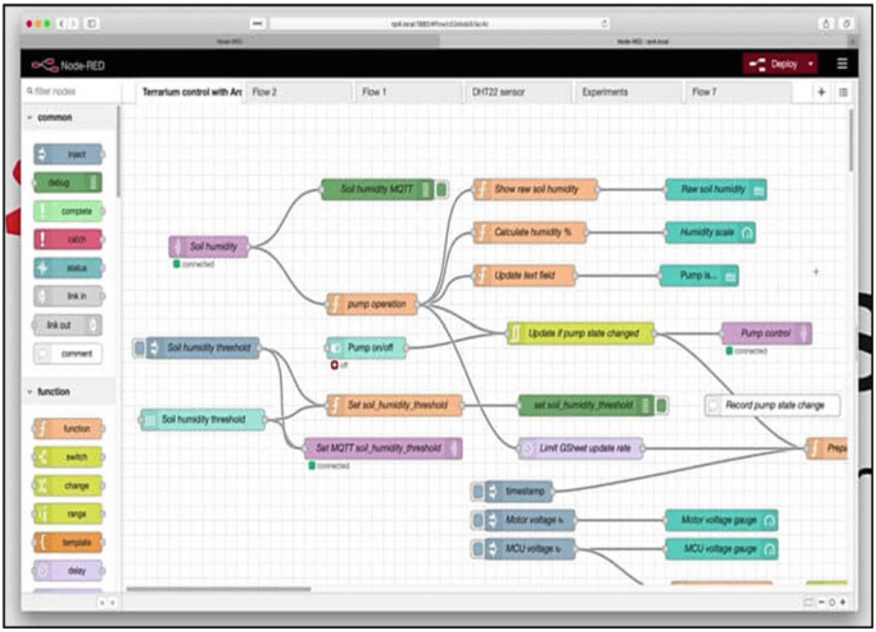


Fig. 12 A glimpse of NodeRed

Fig. 13 A glimpse of IFTTT

Do more with the services you love

Applets bring your favorite services together

Discover Applets for every part of your life

3 Conclusion

These are the important open-source home automation software, which can be used by the developers in the field of IoT. The current review comprises a brief and detailed comparative analysis of the most important open-source home automation software by including their main features like system requirements, development language, messaging service, security, speed and web interface availability.

The proposed has also analyzed the different licensing techniques in a detailed manner. Some of the open source applications dealt in the report are Home Assistant, openHAB, Jeedom, ioBroker, Google home assistant, Calaos, Domoticz, FHEM, Pimatic, AGO control etc. The user can utilize the perfomred research study to select different entities essential for IoT based automation depending on features and availability of system requirements.

4 Future Work

With the proposed comparison on open source IoT based automation and their selections for users, the users can select desired open source software depending upon the system available or vice versa with service protocol for set up communication.

In further work, there is a scope to set up a few cases with various open source software installed on differ support systems mentioned in Table 1 in order to check results and improve the performance.

References

1. What is a “smart home”?: Smart home energy [Online]. Available: http://smarthomeenergy.co.uk/what-smart-home. Accessed on 10 July 2022
2. Pelaez A (2021) 16 open source home automation platforms to use in 2020. Last accessed 19 Mar 2022. https://ubidots.com/blog/open-source-home-automation/
3. Baker J (2017) 6 open source home automation tools, build a smarter home with these open source software solutions. Last accessed 14 Dec 2017. https://www.opensource.com/tools/home-automation/
4. Saxena S, Jain S, Arora D, Sharma P (2019) Implications of MQTT connectivity protocol for IoT based device automation using home assistant and OpenHab. In: 6th International conference on computing for sustainable global development (INDIACom), pp 475–480
5. Stolojescu-Crisan C, Crisan C, Butunoi BP (2021) An IoT-based smart home automation system. Sensors 21:3784. https://doi.org/10.3390/s21113784
6. Korachentsov A, Vasilevna KA (2019) Automatic control method of indoor lighting using motion sensors in smart home systems. https://doi.org/10.13140/RG.2.2.16213.17122
7. Park MJ, Joshua I (2020) Preliminary study of a google home mini. J Digit For 12(1)
8. Zinca D (2018) experiments with protocols and frameworks used in IoT. In: IEEE 2018 12th international conference on communications (COMM)-Bucharest, 14 June 2018–16 June 2018, pp 277–280. https://doi.org/10.1109/ICComm.2018.8484269
9. Wilhelm S (2021) Activity-monitoring in private households for emergency detection: a survey of common methods and existing disagreeable data sources. https://doi.org/10.5220/0010180002630272
10. Macheso P, Manda TD, Chisale S, Dzupire N, Mlatho J, Mukanyiligira D (2021) Design of ESP8266 smart home using MQTT and node-RED. In: International conference on artificial intelligence and smart systems (ICAIS), pp 502–505. https://doi.org/10.1109/ICAIS50930.2021.9396027
11. Why openHAB? [Online]. Available: https://www.openhab.org. Accessed on 10 July 2022
12. Features of home assistant? [Online]. Available: https://www.home-assistant.io/. Accessed on 10 July 2022
13. Jeedom Official Web Page. Available online: https://www.jeedom.com. Accessed on 10 July 2022
14. IoBroker Official Web Page. Available online: https://www.iobroker.net/. Accessed on 10 July 2022
15. Sălăgean M, Zinca D (2020) IoT applications based on MQTT protocol. In: 2020 international symposium on electronics and telecommunications (ISETC), pp 1–4. https://doi.org/10.1109/ISETC50328.2020.9301055

A Review Paper on Network Intrusion Detection System

Nongmeikapam Thoiba Singh and Raman Chadha

Abstract Computer networks are prone to cyber as a consequence of global internet use; as a consequence, academics have developed several Intrusion Detection Systems (IDSs). Identifying intrusions is one of the main significant study topics in data security. It aids in the detection of misuse and attacks as a safeguard for the network's integrity. Machine Learning, Bayesian-based method, nature-inspired meta-heuristic methods, swarm intelligent approach, and Markov neural network is some ways to find the most effective characteristics and thus improve the effectiveness of Intrusion Detection System (IDS). Over the years, numerous databases have been used to evaluate various projects. This publication provides a comprehensive assessment of IDS with machine learning approaches.

Keywords Machine learning · Single classifiers · Hybrid · Ensemble · Misuse identification · IDS

1 Introduction

An IDS is used to identify illegal or aberrant conduct. An attack is initiated in a system that is exhibiting unusual activity. Attackers use system flaws, including poor security measures and practices, and program defects like buffer overflow, to cause network breaches. The intruders might be less privileged equipment owners seeking more network access or black hat hackers accused of stealing data from ordinary internet users [1]. Approaches for identifying intruders can focus on identifying exploits or detecting anomalies. Misuse-based IDS monitors network traffic and compares it to a database of signs of known malicious behaviour. Attacks are discovered in the method of anomaly detection when they are matched to activities that depart from typical user activities [2].

N. T. Singh (✉) · R. Chadha
Department of Computer Science and Engineering, Chandigarh University, Mohali, Punjab, India
e-mail: nthoiba12@gmail.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_34

NIDS or HIDS are two kinds of IDS. To monitor and assess activity on a given machine, computer network (CN) managers use the HIDS method [3]. When traveling over a system, HIDS offers the advantage of allowing encrypted information to be collected. Since each host must configure and manage data, HIDS has the drawback of being hard to maintain. Additionally, certain Denial-of-Service (DOS) assaults may be able to wipe out HIDS. A network-based intelligently deployed software or HIDS that tracks packets as they transit through the point is called a NIDS. The NIDS consists of two components: one for general monitoring and the other for tracking network conversations. The advantage of NIDS is that it only takes a few people to administer a wide network and is frequently camouflaged from various intruders; products are safe against attack. The disadvantage of NIDS is that it makes it hard to identify an attack during high traffic periods.

Many IDS are built on obsolete datasets, such KDD Cup 99, NSL-KDD, and many more, that are devoid of the most recent attack tags. The current initiatives have a poor rate of recognition. This happens as a result of the option to eliminate all unnecessary and irrelevant columns. This occurs when lawful traffic is mistakenly classified as an attack. The FPR expands the IDS's capabilities while decreasing its effectiveness.

The remaining portion of the document is organized as follows. The IDS architecture is contained in Sect. 2. A description of ML algorithms may be found in Sect. 3. Data reduction techniques used in IDS are covered in Sect. 4. Explaining the literature review is Sect. 5. The benchmark datasets for network intrusion detection systems (NIDS) are described in Sect. 6. In Sect. 7, the conclusion is provided.

2 Architecture of IDS

The first IDS was created by Dorothy E. Denning in 1986 in collaboration with the SRI International research team. The signature detection element of the double conceptual framework was characterized as containing an attack rule base and an anomaly identification stage using a statistically based analysis to identify novel attacks. IDS became an exciting topic of inquiry in the research community after then. Signature and anomaly IDS are the two most popular forms of IDS. Anomaly-based IDS employs a statistical method to detect actions that differ from typical resource utilization and behavior characteristics. Anomaly-based IDS employed machine learning techniques to create and train IDS models. These models dealt with any incoming suspicious activities. On the other hand, simple statistical methods like mean, median, and quantiles can detect univariate anomalies in datasets. Data visualization and data analysis techniques are also used to detect anomalies. In anomaly-based detection, the proportion of False Positive and False Negative remains high. Figure 1 shows how these IDS work.

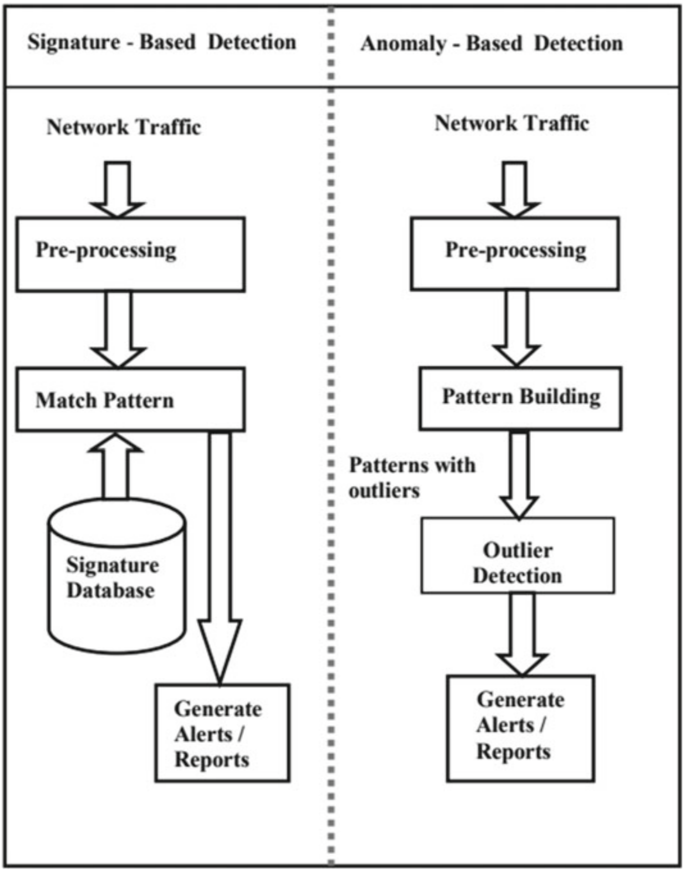


Fig. 1 Signature and anomaly intrusion detection system

Following are the various performance metrics used in IDS:

True Positive Rate: It is computed as the ratio between the number of accurately predicted attacks and the total number of attacks. We can express TPR as:

$$\text{True Positive Rate} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

where

TP = True Positive
FN = False Negative.

False Positive Rate: It is computed as the ratio between the number of normal instances incorrectly predicted as an attack and the total number of normal instances. We can express FPR as:

$$\text{False Positive Rate} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{2}$$

Where

FP = False Positive
TN = True Negative.

False Negative Rate: It indicates when a detector fails to identify an anomaly and classifies it as normal. We can express FNR as:

$$\text{False Negative Rate} = \frac{\text{FN}}{\text{FN} + \text{TP}} \tag{3}$$

where

FN = False Negative
TP = True Positive.

Accuracy: It may be defined as the proportion of examples that are properly categorized to all occurrences.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

where

TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative.

Anomaly-based detection procedures must be devised to identify irregular behaviors and achieve optimal accuracy. ML approaches use complex equations. If enough training data is provided and well-trained methods are built, IDS may even be able to detect zero-day assaults. Cyber security managers could use these strategies must make sure security.

3 Overview of ML Approaches

ML can be defined as a method of learning algorithms to automatically learn and develop quality characteristics so that they do not need to be specified [2, 4] by utilizing prior experience or example information. The ML model is centered on training sets to forecast distinct class labels per attribute. Machine learning models use existing datasets as model inputs to predict outputs. The machine learning models are fed with sufficient training sets to learn from. The dimensions of the datasets are

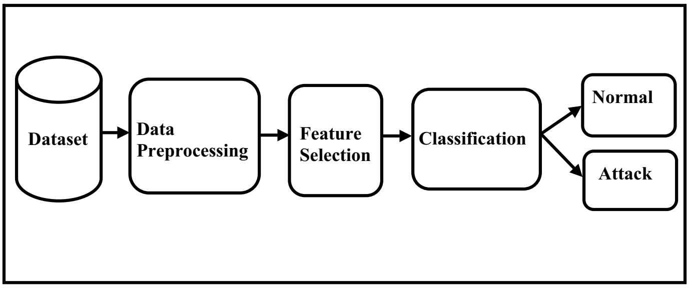


Fig. 2 Block diagram of IDS

reduced. At the same time, the features are extracted. In machine learning, classification is a supervised learning approach where labeled datasets are used in training. The entries in the datasets are labeled with distinct class labels. Supervised Machine Learning models are trained in order to forecast the class labels [5]. Block Diagram of IDS is shown in Fig. 2. ML is often divided into three categories.

3.1 Supervised Learning

The samples of the input vector and their corresponding anticipated output vectors make up the database needed for supervised learning training. NB, KNN, ANN, Decision Tree, SVM, Ensemble methods, and Logistic Regression (LR) are a few of the approaches used in this kind of learning [3].

3.2 Unsupervised Learning

Unsupervised Learning does not have groups to train from, so it must discover an organization's information on its own [1]. This is often referred to as self-teaching. Unsupervised methods include the SOM, Apriori method, Éclat approach and outlier detection, Hierarchical Clustering, and Cluster Analysis.

3.3 Reinforcement Learning

The system is trained to create a series of decisions using reinforcement learning. The aim is accomplished in a hazy and frequently challenging manner [1]. The system uses trial and error to find a solution to this issue. Several reinforcement learning (RL) techniques include DQN, Q-Learning, SARSA, and DDPG.

4 Data Reduction Approaches Used in IDS

Due to the enormous size and complexity of information, the weight of ML and data mining (DM) algorithms did not function well with malware detection [4, 6]. These approaches take a long time to categorize assaults, making deployment in real-time situations more problematic. This is due to the increasing amount of features in-network data that Malware Detection must analyze. The quantity and quality of characteristics are important for categorization, as they help us grasp their significance or association [7]. If the number of characteristics chosen is insufficient, classification accuracy will suffer, and if they are in order to be able, generalization will suffer. Studies showed that using feature extraction (FE) methods in IDS improves accuracy and reduces processing costs [5]. FE methods like Principal Component Analysis (PCA), Linear Discriminant Analysis, and Exhaustive Feature Selection help in reducing the number of features in datasets. Maximum effort is made to retain as much information as possible while reducing the number of features. These feature extraction methods also help in handling overfitting, training speed up, and improving data visualization. Lastly, they help in extracting maximum information from the models. As a preprocessing stage, dimensionality and feature reduction algorithms are utilized to enhance quality and reduce attack identification time.

4.1 Feature Selection

Feature Selection (FS) strategies are being utilized to find a collection of the great attributes that could enhance the procedure's ultimate result while generating minimal errors. One objective is to reduce simulation response and storage usage. FS approaches are used in IDS to increase attack detection performance. Principle Component Analysis (PCA), Information Gain and Gain Ratio are among the most used feature selection approaches [8]. Filter and Wrapper are two types of feature selection algorithms that combine various FS methods.

- In the Wrapper technique, a classifier serves as a "black box" for assessing optimum properties. Such approaches yield great speculation but may suffer from excessive complexity because to the computational cost of creating the classifier.
- Filter approaches don't use classification to evaluate features and are reasonably resistant to overfitting, but they do so using autonomous estimating methods such as distance measurements, consistency measures, and correlation measures.

4.2 Feature Extraction

Showing various samples and columns represent characteristics in a database, with features resulting from both quantitative and subjective discoveries. Feature extraction is a method for reducing the dimensionality of information collecting by choosing a subset so that the accuracy of attack detection accuracy remains unchanged and the amount of time spent discovering is reduced. In the environment, there are a variety of feature extraction approaches. Self-Organizing Maps (SOM), Principle Component Analysis (PCA) and other techniques are instances.

4.3 Clustering

Data samples are sorted into data sets in clustering, with information samples in each set being comparable in some manner.

5 Literature Survey

An IDS is a system traffic monitoring network that tracks suspicious behavior and transmits out notifications when it is found. It's a part of the software that investigates a computer or network for malicious activities or policy violations. We've compiled a list of IDS that have been explored in this study.

Yedukondalu et al. [9] executed ML algorithms to the data set, comparing and evaluating their results. In order to identify intrusion rates, the suggested app uses the SVM and ANN methods. Every technique is utilized to calculate whether the information being sought is allowed or includes any irregularities. While the IDS scans the information requested, if it detects anything malicious material, the request is dropped. These techniques used a feature selection method based on correlation and Chi-Squared to minimize the collection by removing unnecessary data. The preprocessed database is instructed and evaluated with the algorithms to produce notable findings, which improves predictive performance. The testing was performed using the NSL-KDD database. Eventually, an SVM technique achieved 48% accuracy, while the ANN method achieved 97% accuracy. In this sample, the ANN performance was better than the SVM.

Gupta et al. [10] a real-time solution for hybrid IDS that employed anomaly detection to find new threats and signature-based identification to identify well-known incursions was proposed. The anomaly identification approach may be used to identify intrusions that escaped the signature-based method, leading to a high proportion of detection value in this investigation. On the last day of the test, the system's precision achieved a substantial value of 92.65%, and as the approach improved and the machine was taught daily, the percentage of false negatives significantly fell. However, when the model was applied to a big data collection, the issue of a sluggish detection rate surfaced.

Lin et al. [11] in both the FFANN and PRANN were used to develop the ANN-based IDS, which used scaled conjugate gradient and Bayesian regularization approaches. Particular result measures were utilized to analyze the work's quality and competence. The two models have been proved to outperform each other in several output tests on distinct attack observations from the proven effectiveness. Consequently, the FFANN increased precision by 98.0742%. The production's dependability should be enhanced through testing the design on multiple databases.

Hajisalem et al. [12] a single ML classifier was used to create an IDS. They used the NSL-KDD database to test RF and DT methods. The random classifier outperforms the decision tree in accuracy and produces superior outcomes. Both the detection rate and the FPR were not examined in the study.

Atefi et al. [13] single classifiers were employed to more accurately identify the assault. The techniques used include SVM, LR, RF, and DT. The NSL-KDD database was used to evaluate the work. The IDS performs best using the RF classifier, according to the study. They also discovered that the RF method had the quickest completion time. The study has the disadvantage of only being able to examine one database efficiently.

Ankome et al. [14] proposed for MANET a HCIDM. NS 2.35 was used to create HCIDM. 50 normal MN and five black hole nodes traveled freely in a 4000 m square grid in the experiment. The authors place the black hole on the shortest path between the source and DN. The HCIDM's productivity is assessed using measures such as PDR, packet loss, and throughput. The system with HCIDM had average network congestion of 26.92, but CCIDM had an average of 19.85, which could be due to HCIDM's maximum bandwidth. Throughout a BHA, the average total performance of the system with HCIDM is 23.23 Mbps, a 94% increase.

Nie et al. [15] GAN was used to create a new intrusion detection tool (Generative Adversarial Network). For feature learning, the GAN-based technique may extract less-dimensional characteristics from actual system flows. MATLAB was used to develop the suggested GAN-based system. The proposed approach was evaluated using the CIC-DDoS2019 and CSE-CIC-IDS2018 databases. Their strategy can significantly enhance intrusion detection accuracy based on the simulation results. The accuracy of both databases was 95.32% for the CSE-CIC-IDS2018 dataset and 98.53% for the CIC-DDOS2019 dataset.

Sankaranarayanan et al. [16] the intruders are effectively identified using the suggested Secure Intrusion Detection approach, which uses the RSA approach. The simulation is run on Ubuntu and the Network Simulator NS2.34 environment. The findings showed that in the existence of intruders, the suggested secure IDS approach provides a higher PDR. Even when routing overhead is on the higher side, it is permissible when authors need to detect attackers.

Huang et al. [17] based on traffic pattern learning, created an effective IDS for VSNs. In order to characterise the dynamic properties of network traffic in Visual Sensor Networks, the suggested approach builds a traffic model. By utilising Omnet++ to create a VSN, it is possible to compare the outcomes of our IDS with the traffic data that has been produced in order to find the best feature set for traffic pattern learning. An HSOM is then used to comprehend traffic circumstances and identify intrusions. In order to hasten the HSOM's training and improve its understanding of attack patterns, an active learning approach is also developed. The suggested approach provides excellent real-time performance and great detection precision.

6 NIDS Datasets

KDD CUP'99: It is a popular NIDS dataset with 5 million training and 2 million testing records. Each record includes 41 features. The attacks are labeled as DoS, R2L, U2R, and Probe [18].

UNSW-NB15: The Australian Center for Cyber Security releases this NIDS dataset. It contains around 2 million records with 49 features. This dataset includes nine attack types: Shellcode, Worms, Generic, Reconnaissance, Port Scans, DoS, Fuzzers, and Backdoors [19].

NSL-KDD: This is the updated version of the KDD Cup'99 dataset. It contains four different attack types, including DoS, R2L, and Probe, and 41 characteristics [20].

CIC-IDS2017: This NIDS dataset is made available by the Canadian Institute of Cyber Security. It contains normal flows as well as updated real-world attacks. Brute force, botnet, HeartBleed, DoS, web, DDoS, and infiltration are some of the attack techniques it may use [21].

Kyoto 2006+: Kyoto University produced this dataset from network traffic logs. It has exactly 24 statistical characteristics [22].

CSE-CIC-IDS2018: This dataset contains seven attack types: Web, Infiltration, Brute Force, DDoS, Botnet, DoS, and Heart Bleed [23].

7 Conclusion

The development of ML introduces new techniques to IDS, with many researchers and academics using various types in the construction of IDS designs. In today's environment, intrusion detection is crucial for information security, so ML-based apps have aided in the discovery of new assaults. In recent years, several classifiers, such as hybrid systems and ensemble learning approaches, have dramatically enhanced the reliability of attack detection strategies. However, there is still a problem with the proportion of FP and FN. Authors encourage scientists and researchers to consider using more approaches with a higher accuracy rate.

References

1. Sangkatsanee P, Wattanapongsakorn N, Charnsripinyo C (2011) Practical real-time intrusion detection using machine learning approaches. Comput Commun 34(18):2227–2235
2. Nader P, Honeine P, Beauseroy P (2016) Detection of cyber attacks in a water distribution system using machine learning techniques. In: 6th international conference on digital information processing and communications, ICDIPC 2016
3. Jabbar MA, Aluvalu R, Reddy SSS (2017) Cluster-based ensemble classification for intrusion detection system. In: ACM international conference proceeding series, vol Part F1283, pp 253–257
4. Verma P, Anwar S, Khan S, Mane SB (2018) Network intrusion detection using clustering and gradient boosting. In: 9th international conference on computing, communication and networking technologies (ICCCNT), pp 1–7
5. Shakya S (2021) Modified gray wolf feature selection and machine learning classification for wireless sensor network intrusion detection. IRO J Sustain Wirel Syst 3(2):118–127
6. Denning DE (1986) An intrusion-detection model. In: IEEE symposium on security and privacy, pp 118–131
7. Saha S, Sairam AS, Yadav A (2012) Genetic algorithm combined with support vector machine for building an intrusion detection system. In: Proceedings of the international conference on advances in computing, communications and informatics. ACM, pp 566–572
8. Vergara JR, Estevez PA (2014) A review of feature selection methods based on mutual information. Neural Comput Appl 24(1):175–186
9. Yedukondalu G, Hima Bindu G, Pavan J, Venkatesh G, Sai Teja A (2021) Intrusion detection system framework using machine learning. In: Third international conference on inventive research in computing applications (ICIRCA)
10. Gupta S (2016) Analyzing the machine learning algorithms—Naïve Bayes, random tree, and support vector machines SVM using the kdd99 data set to predict and classify the, vol 2, pp 452–459
11. Lin WC, Ke SW, Tsai CF (2015) CANN: an intrusion detection system based on combining cluster centers and nearest neighbors. Knowl Based Syst 78(1):13–21
12. Hajisalem V, Babaie S (2018) A hybrid intrusion detection system based on ABC-AFS algorithm for misuse and anomaly detection. Comput Netw 136:37–50
13. Atefi K, Yahya S, Rezaei A, Hashim SHBM (2016) Anomaly detection based on profile signature in network using machine learning technique. In: Proceedings—2016 IEEE region 10 symposium, TENSYMP 2016, pp 71–76
14. Ankome T, LusilaoZodi G-A (2021) Hierarchical cooperative intrusion detection method for MANETs (HCIDM). In: 15th international conference on ubiquitous information management and communication (IMCOM)
15. Nie L, Wu Y, Wang X, Guo L, Wang G, Gao X, Li S (2021) Intrusion detection for secure social internet of things based on collaborative edge computing: a generative adversarial network-based approach. IEEE Trans Comput Soc Syst 1–12
16. Sankaranarayanan S, Murugaboopathi G (2017) Secure intrusion detection system in mobile ad hoc networks using RSA algorithm. In: Second international conference on recent trends and challenges in computational models (ICRTCCM)
17. Huang K, Zhang Q, Zhou C, Xiong N, Qin Y (2017) An efficient intrusion detection approach for visual sensor networks based on traffic pattern learning. IEEE Trans Syst Man Cybernet Syst 47(10):2704–2713

18. Bay S (1999) The UCI KDD archive. University of California, Department of Computer Science, Irvine, CA. http://kdd.ics.uci.edu
19. Moustafa N, Slay J (2016) The evaluation of network anomaly detection systems: statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. Inf Sec J A Global Perspect 25(1–3):18–31. https://doi.org/10.1080/19393555.2015.1125974
20. Tavallaee M, Bagheri E, Lu W, Ghorbani AA (2009) A detailed analysis of the KDD CUP 99 data set. Paper presented at: proceedings of the IEEE symposium on computational intelligence for security and defense applications. IEEE, Ottawa, ON, Canada, pp 1–6
21. Abdulhammed R, Musafer H, Alessa A, Faezipour M, Abuzneid A (2019) Features dimensionality reduction approaches for machine learning based network intrusion detection. Electronics 8(3):322. https://doi.org/10.3390/electronics8030322
22. Song J, Takakura H, Okabe Y, Eto M, Inoue D, Nakao K (2011) Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation. Paper presented at: proceedings of the 1st workshop on building analysis datasets and gathering experience returns for security. Salzburg, Austria, pp 29–36
23. Sharafaldin I, Lashkari AH, Ghorbani AA (2018) Toward generating a new intrusion detection dataset and intrusion traffic characterization. Paper presented at: proceedings of the 4th international conference on information systems security and privacy (ICISSP). Madeira, Portugal, pp 108–116

ESP32 Based Irrigation System

M. Koteswara Rao, M. Satish Kumar, M. Jaijaivenkataramana, and Ch. Sai Sowjanya

Abstract In India, Agriculture is remaining as the major occupation among people. Farmers cultivate various crops depending on the type of land and season. Additionally, farmers also irrigate agricultural land while cultivating. However, farmers cannot predict how much water they will use for irrigation. Currently, farmers are supplying water to the field without knowing the moisture content of the soil. If this continues, there will be a scarcity of water, and underground water will also be depleted and may not be available for future generations. Furthermore, the crops are destroyed if more water is supplied than the required amount. This research study has utilized an advanced technology called Internet of Things (IoT) to design and develop agricultural monitoring systems for evaluating soil moisture content and other agricultural parameters. This system includes soil moisture sensors for measuring the soil moisture content. Rain drop sensor, water level sensor, and DHT11 sensor are used for measuring rain, field water level content, temperature and humidity. The outputs from the sensors are sent to the ESP32 module, which then sends it to the motor (used for irrigation and ejection). These values can also be visualized in the ThingSpeak cloud platform.

Keywords NodeMCUESP-32 · Soil moisture · Raindrop · Water level · DHT11 sensors · Cloud

1 Introduction

Agriculture is important in our country because it highly contributes to India's GDP. Agriculture is also important to our country's population. However, with the increased use of synthetic fertilizers, soil quality is deteriorating. Furthermore, by not knowing

M. Koteswara Rao (✉) · M. Satish Kumar · M. Jaijaivenkataramana · Ch. Sai Sowjanya
Department of Electrical Communication Engineering, Sri Vasavi Engineering College and Pharmacy (Autonomous), JNTUK, Tadepalligudem, India
e-mail: koteswararao.maram@srivasaviengg.ac.in

M. Satish Kumar
e-mail: satish.margani@srivasaviengg.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_35

how much water must be used for irrigation, the water supply is depleted resulting in long-term consequences for future generations. Each agricultural field has a different irrigation value. Furthermore, the moisture content of soil varies from place to place. Since irrigation is dependent on soil moisture content, farmers can easily determine how much water to use by gaining an adequate knowledge on the moisture content of the soil [1, 2]. In addition, water is saved. In this case, a raindrop sensor is used to detect the raindrops, DHT11 sensor measures the temperature and humidity on that particular day, and the water level sensor detects the level of water. These inputs are then sent to the ESP32 board. Finally, the output is sent to the motor, and then the data gets saved in the ThingSpeak cloud [3, 4].

2 Literature Survey

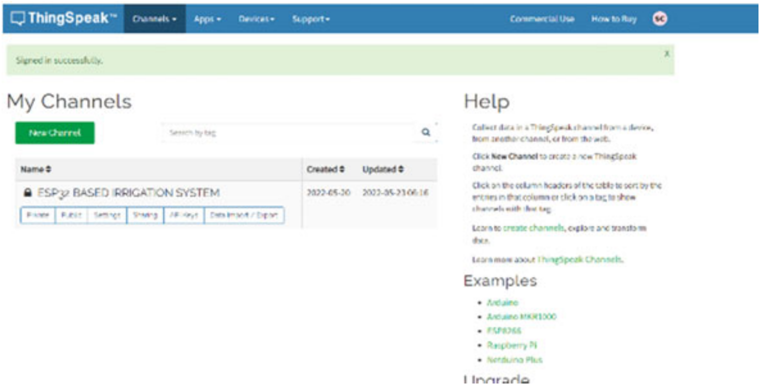
Even though agriculture plays a major role in contributing to the country's economy, farmers are still using the traditional farming methods. Irrigation challenges result in the wastage of underground water and water bodies. To eliminate the problems associated with water wastage, Das et al. [1] proposed an innovative model to avoid water wastage in the process of irrigation by automating the water control process. Vadapalli et al. [2] have utilized the cloud platform known as Thingspeak to collect data on the field parameters of an automatic motor pumping system. As water level systems has become necessary for farmers at any climatic conditions, Suresh et al. [5] have briefly described about the incorporation of new technologies in handling the ever-changing weather conditions. Further, Babiuch et al. [4] have described about the working principle and working environment of ESP32.

3 Proposed System

The proposed system has used ESP32 and some sensors like soil moisture sensor, raindrop sensor, DHT11 sensor and water level sensor. The sensor will then be sent to the cloud by using an inbuilt Wi-Fi module for ESP32-WROOM-32.

3.1 ThingSpeak

ThingSpeak is an IoT platform as in Fig. 1; it is a modern cloud platform with a user-friendly interface, which allows to store the sensor data, visualize the data, and then analyze it by using MATLAB or any other software. This cloud application works with HTTP and MQTT IoT protocols.



Signed in successfully.

My Channels

Name ⇕	Created ⇕	Updated ⇕
🔒 ESP32 BASED IRRIGATION SYSTEM	2022-05-20	2022-05-23 06:16

Help

Collect data in a ThingSpeak channel from a device, from another channel, or from the web.

Click **New Channel** to create a new ThingSpeak channel.

Click on the column headers of the table to sort by the entries in that column or click on a tag to show channels with that tag.

Learn to create channels, explore and transform data.

Learn more about ThingSpeak Channels.

Examples

- Arduino
- Arduino MKR1000
- ESP8266
- Raspberry Pi
- Netduino Plus

Upgrade

Fig. 1 ThingSpeak cloud platform

3.2 Soil Moisture Sensor

The soil moisture sensor is designed specifically to measure the moisture content of the soil. The soil moisture sensor is highly used in a variety of applications, including root zone measurements and regulating the existing conventional irrigation timer. Soil senor values are shown in Fig. 2.

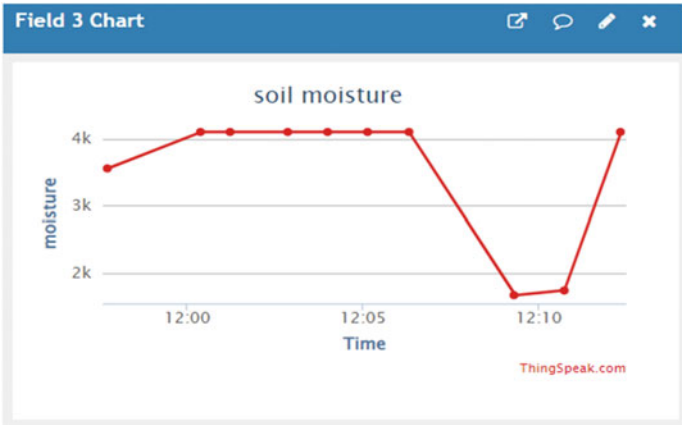


Fig. 2 Soil sensor values

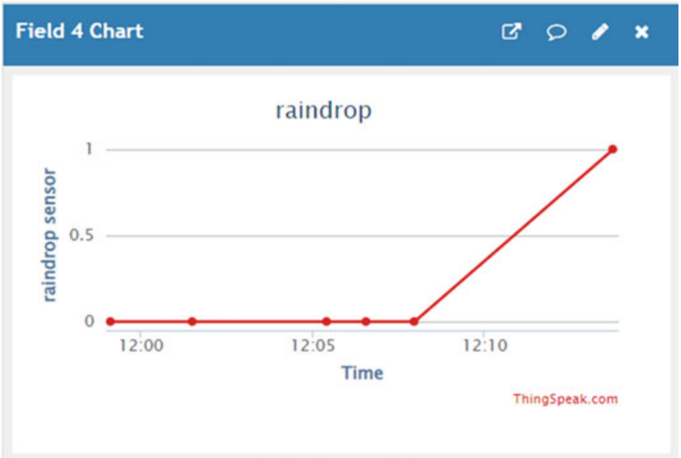


Fig. 3 Raindrop sensor values

3.3 Raindrop Sensor

The raindrop sensor is a device that detects raindrops. It consists of two modules: a rain board that determines the rain droplets and a control module that differentiates analog values and converts them to digital values. Raindrop sensors are used in a variety of applications, including home automation and protecting the internal parts of an automobile from rain [5]. Raindrop sensor values are shown in Fig. 3.

3.4 DHT11 Sensor

The DHT11 is a temperature and humidity sensor with a dedicated NTC for monitoring the temperature and an 8-bit microcontroller for monitoring the temperature and humidity output values. This sensor is used to measure the humidity in medical equipment, home automation systems, automotive, and other weather control applications as in Fig. 4.

3.5 Water Level Sensor

The water level sensor module has a series of parallely exposed lines, which measure droplets or water volume to determine the water level as I Fig. 5 [6]. Water level sensor used in different applications to measure water levels in tanks help to control water

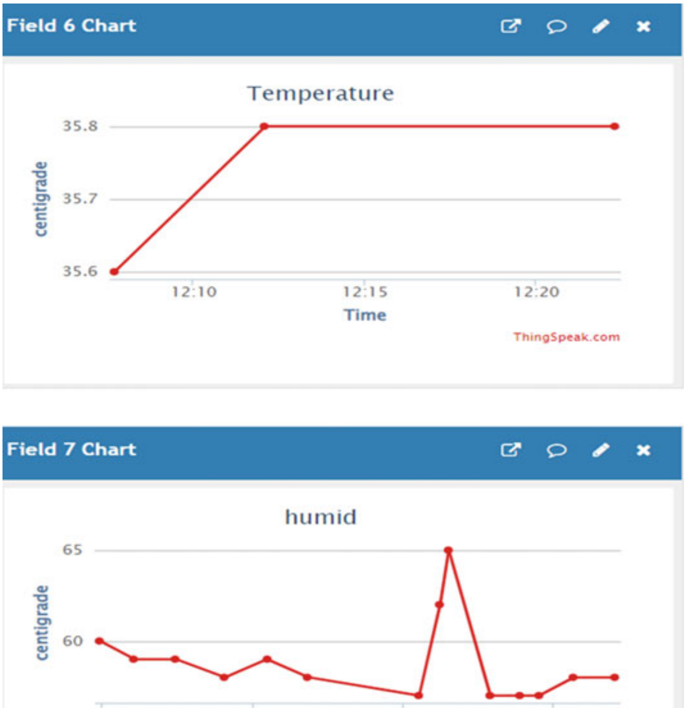


Fig. 4 DHT11 sensor

levels. This can also be used to automatically turn ON/OFF pumps and life station switches.

3.6 Methodology of the Proposed Work

The overall block diagram of the proposed work is depicted in Fig. 7. Each and every block has a detailed explanation of each sensor included in the proposed work. The block diagram consists of various sensors like soil moisture sensor, raindrop sensor, water level sensor, and DHT11 sensor. These sensors are then connected to the ESP32 board module. Here, the input values are given and the output values are fed to the actuators and finally the data is stored in cloud platform i.e., Thingspeak [7].

Soil moisture sensors, rain drop sensors, water level sensors, and DHT11 sensors are some of the sensors used in agriculture. Analog sensors include the soil moisture sensor, rain drop sensor, and water level sensor. The DHT11 sensor is a digital sensor. The ESP32 module includes a Bluetooth and Wi-Fi module. The sensors are connected to the ESP32 module in this case. Both analogue and digital values can

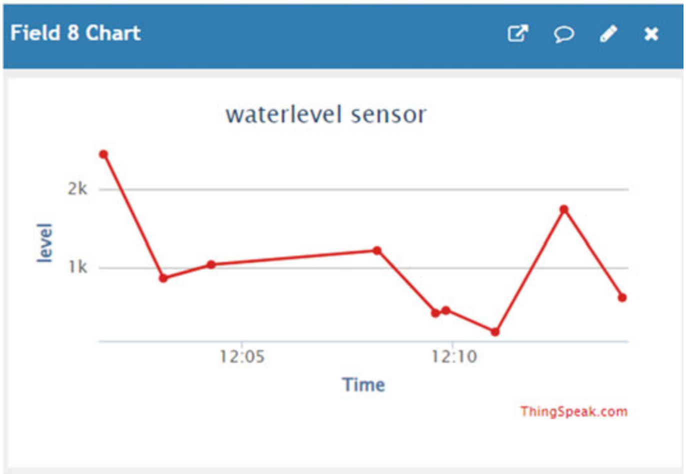


Fig. 5 Water level sensor values

be accepted by ESP32. The output is routed to actuators for additional processing. Here, the rain drop sensor is used to sense the rainfall and the sensed data is then provided to the actuators for performing automatic irrigation in field [8]. Moreover, during rain fall the underground water usage gets decreased.

Here, the water level sensor is used for sensing the water level of the field. Here, the water levels differ based on the different types of agricultural field. If the water level exceeds the threshold value, the water is moved to nearby water bodies for future use.

The DHT11 sensor is used for measuring the humidity and temperature of the agricultural field. DHT11 sensor consists of a capacitive humidity sensing element and a thermistor for sensing the temperature. Here, the cloud platform used here is Thingspeak. Every sensor and actuator value is uploaded to the cloud platform Thingspeak. All sensor values are shown in Fig. 6.

4 Proposed System

See Fig. 7.

4.1 Methodology of the Proposed Work

The process flow diagram is shown in Fig. 8.

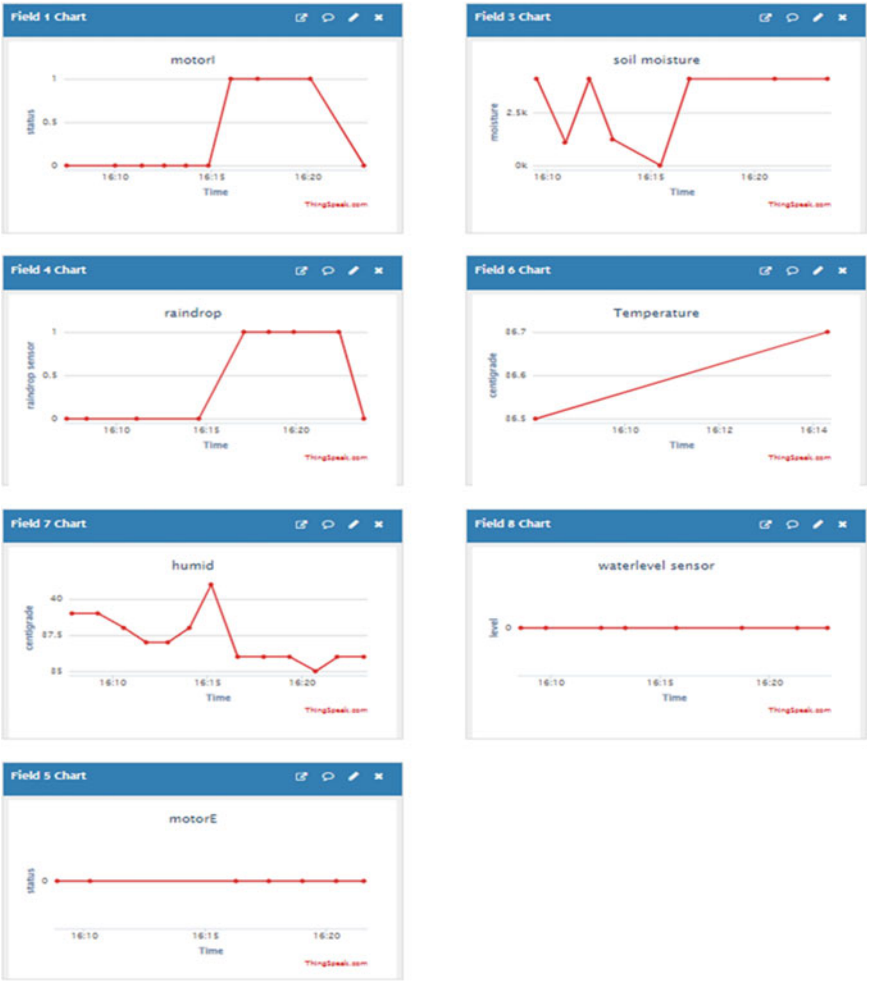


Fig. 6 All sensor values

The proposed research focuses on the interfacing of sensors such as soil moisture sensors and rain drop sensors. The motor will turn ON when the moisture content of the land rises, otherwise it will turn OFF. If there is rain, the motor will be turned OFF; otherwise, it will be turned ON. If rain falls but the required threshold value of moisture content is not present in the soil, the motor will turn ON. The two conditions will be checked here. A water level sensor is also used to monitor the water levels in the field. If the level of water exceeds required level, the water will be removed by using an ejection motor. DHT11 sensor, a digital sensor will be used to measure the temperature and humidity of the agricultural field. All the sensors are linked to the ESP32 microcontroller [9, 10]. The measured parameter values will be displayed

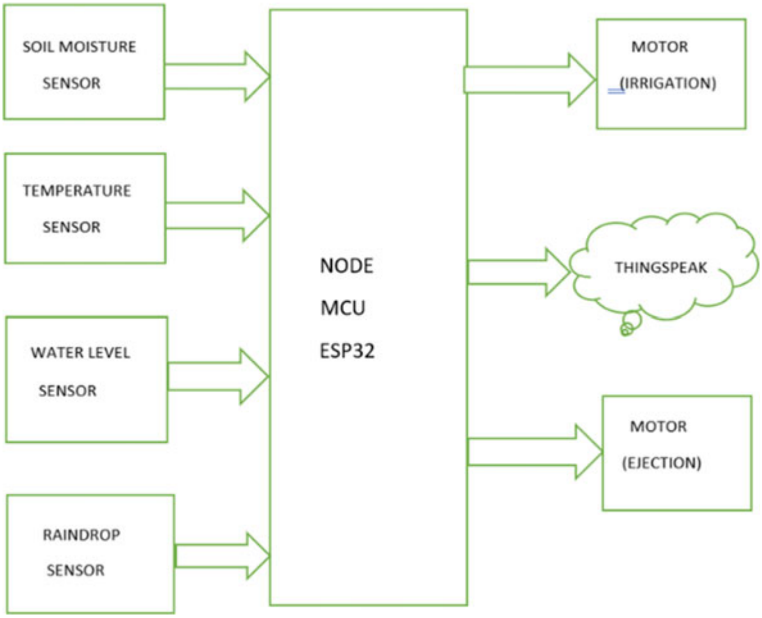


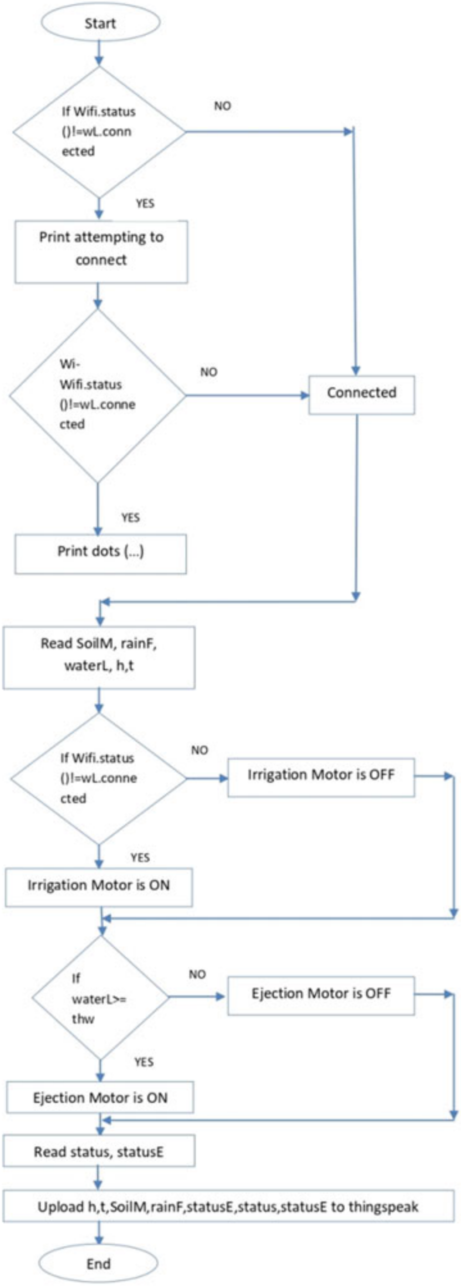
Fig. 7 Block diagram

and saved in the Thingspeak cloud platform. The project's hardware configuration is shown in Fig. 9. Sensor values on serial monitor are shown in Fig. 10.

5 Conclusion

The primary goal of this ESP32-based irrigation system is to be more innovative, user-friendly, resourceful, and efficient than the existing systems. Different agricultural parameters such as soil moisture, temperature, humidity, water level, and raindrop sensing are continuously measured. The data for these parameters is then uploaded to the cloud. By using the proposed system, the farmer can learn about the field's current parameters at any time and from any location by using these updated values.

Fig. 8 Flow chart


Start
If Wifi.status ()!=wL.connected
NO
YES
Print attempting to connect
Wi- Wifi.status ()!=wL.connected
NO
Connected
YES
Print dots (...)
Read SoilM, rainF, waterL, h,t
If Wifi.status ()!=wL.connected
NO
Irrigation Motor is OFF
YES
Irrigation Motor is ON
If waterL>= thw
NO
Ejection Motor is OFF
YES
Ejection Motor is ON
Read status, statusE
Upload h,t,SoilM,rainF,statusE,status,statusE to thingspeak
End

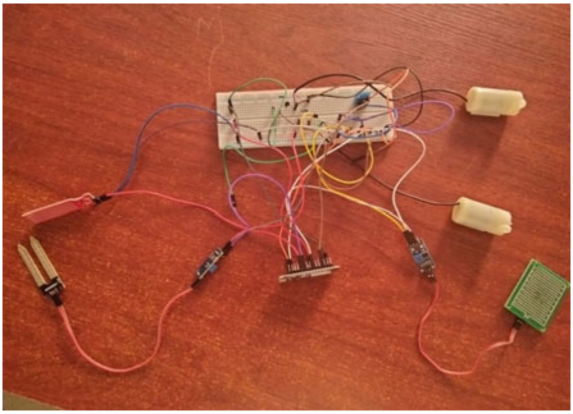
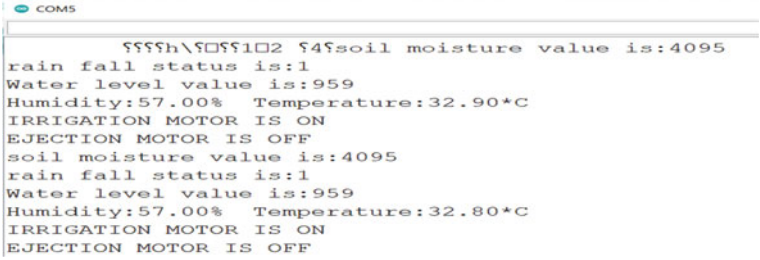


Fig. 9 Hardware setup of ESP32 irrigation based system



```
          ⸮⸮⸮⸮h\⸮□⸮⸮1□2 ⸮4⸮soil moisture value is:4095
rain fall status is:1
Water level value is:959
Humidity:57.00%  Temperature:32.90*C
IRRIGATION MOTOR IS ON
EJECTION MOTOR IS OFF
soil moisture value is:4095
rain fall status is:1
Water level value is:959
Humidity:57.00%  Temperature:32.80*C
IRRIGATION MOTOR IS ON
EJECTION MOTOR IS OFF
```

Fig. 10 Sensor values on serial monitor

References

1. Das K, Sahu S (2020) Smart farming using IoT. Int Organ Sci Res 15(02)
2. Vadapalli A, Peravali S, Dada VR (2020) Smart agriculture system using IoT technology. Int J Adv Res Sci Eng 09(09)
3. Anitha A, Sampath N, Jerlin MA (2021) Smart irrigation system using internet of things. In: International conference on emerging trends in information technology and engineering
4. Babiuch M, Foltynek P, Smutny P (2020) Using the Esp32 microcontroller for data processing. In: International Carpathian control conference
5. Suresh N, Hashiyana V, Kulula VP, Thotappa S (2019) Smart water level monitoring system for farmers. In: International conference on emerging trends in information technology and engineering
6. Sharma A (2020) Review on IoT based water level sensing and controlling. Int J Eng Res Technol 09(07)
7. Alsahi QN, Marhoon AF (2020) Design health care system using Raspberry pi and Esp32. Int J Comput Appl 36
8. Krishna BV, Priyanka K (2014) Soil moisture sensor design for crop management system by using cellular communication. Int J Adv Res Electr Electron Instrum Eng 03(10)
9. Sai Ram KS, Gupta ANPS (2021) IoT based data logger system for weather monitoring using wireless sensor networks. Int J Eng Trends Technol 32
10. Tyagi A, Gupta N, Navani JP, Tiwari R, Gupta A (2017) Smart irrigation system. Int J Innov Res Sci Technol 03(10)

RFID (Radio Frequency Identification) Tag Collision Risk Mitigation Analysis and Avoidance

Aditya Sukhwal, Gourab Kundu, and Chandrani Chakravorty

Abstract RFID (Radio Frequency Identification) transponder is a small beacon tag responsible for transmitting the radio waves in the range of 30 kHz to 3 GHz to an antenna. The antenna is then connected to the reader, which sends and receives signals from the antenna. The use of multiple tags in a perimeter increases the chance of tag collision. Tag collision occurs when multiple tags send signal to the reader at the same time. Such a situation can lead to miscommunication between tags and readers present within the same perimeter; such miscommunication between reader and tags can lead to the failure of the entire RFID system. To avoid such a situation, an algorithm mechanism helps to avoid collision risks. The main goal is to collect data and identify various ranges and proximity in which an RFID tag collision may occur, as well as to conduct preventative analysis to avoid such failures.

Keywords Tag · Transponder · Reader · Antenna · Collision

1 Introduction

Radio Frequency Identification (RFID) tag is the small beacon device, which emits the radio wave signals. The signals emitted by the RFID tags are received by an antenna connected to the main device, which is called as reader. Reader is considered as the main brain of RFID tag [1] as it is responsible for converting the signals into raw data, which then gets stored in a database and from there all the tags database can be maintained and manipulated. The reader communicates to the antenna to obtain the radio waves signal emitted from the tags [2].

A. Sukhwal (✉) · G. Kundu · C. Chakravorty
Department of MCA, RV College of Engineering, Bengaluru, India
e-mail: adityas.mca19@rvce.edu.in

G. Kundu
e-mail: gourabkundu.mca19@rvce.edu.in

C. Chakravorty
e-mail: chandrani@rvce.edu.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*, Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_36

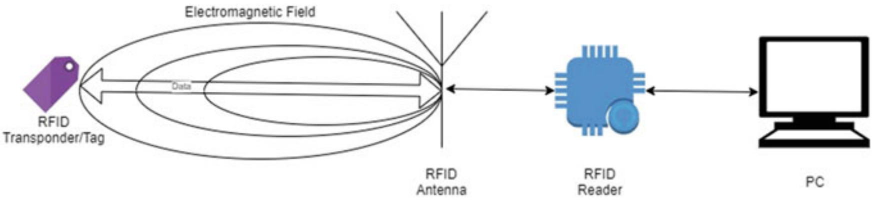


Fig. 1 Radio frequency identification mechanism

In a specific area, the tags have a fixed perimeter. When a tag leaves the perimeter, the antenna stops receiving signals from the tag, and thus a security alarm can be triggered, which is essentially a use case of RFID systems in supply chain management, where it is easy to keep track of inventory and avoid the risk of losing inventory [3].

One common example for RFID use case is the attachment of RFID tags to the clothes at cloth store, these tags are attached to clothes so until the cloth is purchased the tag is not removed, hence if anyone take the cloth out of store without buying the alarm will set ON and burglar will get caught. RFID tags have unique identity to recognize automatically. Additionally, RFID tags are tough and waterproof as they still work after getting dropped into water.

Figure 1 shows the working mechanism of a RFID system. Here, a transponder tag emits the radio wave signal in a perimeter, wherein the signals emitted by the transponder gets received by an antenna, which is further connected to the main brain of RFID system called RFID Reader. The RFID reader gets connected to a computer and also to a local switcher throughout which multiple reader will be connected to a centralized cloud based computer. The central connection of RFID reader ecosystem expands the area of reader especially when a system is required for large warehouses. Such system provides an edge in terms of connecting an ecosystem to the cloud based distributed system, which can be used to trace the inventory from any place in the world via internet. Antennas are responsible for transmitting the electromagnetic waves.

2 Types of RFID Tags

There are mainly 3 types of RFID tags, all 3 types of tags have different applications and use cases, as shown in Fig. 2. The classification of different types of tags have their variation in specifications. All of them provide different radio waves and wavelength ranges and hence the cost of operating them is also directly proportional to their specification (Table 1).

A. *Low Frequency transponder* is highly cost effective and operates at the range of 30.0–300.0 kHz. The absence of battery source makes them long lasting with less maintenance cost. Application of these type of tags are electronic identity card, bus passes etc.

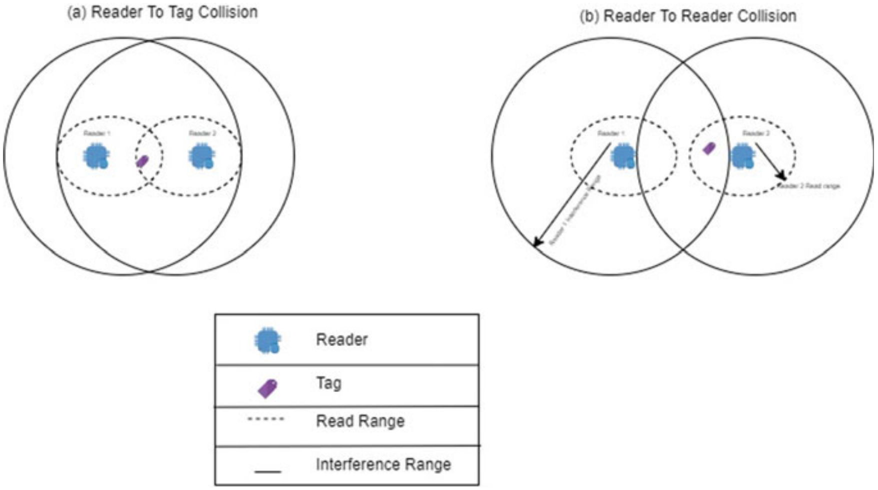


Fig. 2 Two type of collision present in RFID. **a** Reader and tag. **b** Reader and reader [4]

Table 1 Comparison of 3 types of RFID transponders

Types of RFID tags			
	Low-frequency (RFID)	High-frequency (HF)	Ultra-high-frequency (UHF)
Frequency range	30–300 kHz	3–30 MHz	300 MHz to 3 GHz
Common frequency	125 kHz or 134 kHz	13.56 MHz (NFC)	860–960 MHz (UHF Gen2)
Relative cost	$$	$$–$$$	$
Read range	≤10 cm	≤30 cm	≤100 m
Benefits	More resistant to interference by liquids and metals	Higher memory capabilities, NFC tags can function as both reader and tag	Lower cost, with good read range and fast read rates
Common applications	Animal tracking, automobile inventorying	Promotional packaging and labels, contactless payment, library collections	Inventory control, item-level tracking, supply chain visibility and efficiency

Credit Resource label

B. *High Frequency transponders* are the cost-efficient NFC (Near Field Communication) tags. NFC used here is mobile phones, which helps to transfer data at a lightning speed by including more memory. It works as both integrated reader and transponder at the same time and cover a range up to 30 cm.

C. *Ultra-High-Frequency transponders* have a battery, which allows it to emit a wide range of radio signals, such as 860.0–960 MHz and a perimeter range of 100 m. UHF transponders are used in the supply chain management applications used for inventories. These tags can be seen in shopping malls connected to garments and apparels to prevent theft. They have a high operating range of up to 100 m.

D. *Active RFID Tags* includes a small battery, which power them for carrying out active control transmission of radio waves with antenna, meanwhile Passive RFID tags get energized when it comes into proximity of radio waves emitted by the antenna.

3 Background

Types of tag collisions which occur in highly dense environment are usually classified into two types, Reader and Tag collision and Reader and Reader collision. The first type, reader and tag collision occurs when one tag emit signals to two readers at the same time. The second type, reader and reader collision occurs when a tag is supposed to come under reader X enters the proximity of reader [4].

ALOHA algorithm and binary search algorithms can be classified as the two of major algorithms which are specifically used to avoid the RFID tag collisions. At present, the algorithms that are used to avoid the RFID tag collisions are dependent on TDMA (Time Division Multiple Access) [5].

Different types of RFID tags operate at different frequencies and to optimize the tags for high frequency the inclusion of the battery is needed in High Frequency (HF) and Ultra High Frequency (UHF) based RFID systems. By implementing the two way handshake mechanism, RFID systems assist in avoiding different cyber-attacks [1, 6].

Q algorithm has been massively used by RFID systems since 2005 due to its tag sorting and handling capabilities, Q algorithm collects the responses from the tags and group them according to their response time and with help of various iterations it identifies the tags [7].

The compact exclusion validation method avoids limiting tags by testing them one by one in each slot. The inclusion of a key filtering method allows to perform fast tag filtering while sorting the necessary tags [8].

Multiple readers are frequently present in large warehouses and inventory tracking environments; as the number of readers increases, so does the number of tags. For such a complex system, it is critical that all readers remain critically coordinated in real time. Communication latency is considered as a major issue in such sophisticated systems [9].

Dynamic and backtracking based binary search algorithm is used to prevent tag collisions and make the identification process more time efficient. The binary search algorithm outperforms other algorithms in identifying tags [10].

Tag collision can be avoided by implementing and modifying multiple existing protocols that are fundamentally based on ALOHA, Tree, and other hybrid protocols. The complexity and cost of the RFID network are two components that are primarily concerned with the entire RFID system setup, where the efficiency of the entire ecosystem is directly related to the cost and complexity of the RFID network [11].

When an RFID tag does not respond to a reader within a specified time frame and returns a null response, the reader terminates the idle slots. This is where the M optimal algorithm comes into effect; it is a splitting algorithm in which the value of M is carefully selected when the value of M is greater than 2 [12].

The DFSA algorithm has the capability to recover almost 99% of communication channels present in RFID system. With the help of DFSA reader, the algorithm implements a Field Programmable Gate Array (FPGA) and then the entire RFID ecosystem gets verification through the communication tests accompanied by the tags used in commercial spaces [13].

There are many fields where RFID is being used in today's world. The applications of RFID have been significantly increasing since the inception of RFID system. The RFID system admin and any unauthorized user can use RFID card to gain access to confidential places. This limits the application of RFID tags in high intensity defense and security areas [2].

Passive RFID tags emit radio wave signals in a proximity, which further acquired by the antenna. By using the amount of energy emitted, the distance of the tag from RFID system can be measured. After receiving the signals, the reader goes to the idle state. Here, the overall aim is to increase the efficiency of slots by incorporating K-means algorithm [3].

Ultra-High Frequency RFID transponders may have a range up to 50 m, at the same time it results in high power consumption to acquire such high range radio waves. Bayesian algorithm with filter results in the transmission of power in UHF based RFID tags. The main intent of this study is to save the power when RFID transponder remains in idle states by using machine learning algorithm [14].

RFID systems are often vulnerable to relay attacks, which are referred as man-in-the-middle attacks in which the communication between sender and receiver is disrupted by a third-party unauthorized user, and such interception poses a significant risk to RFID systems. Lightweight method provides the best solution for reducing the likelihood of such successful relay attacks [15].

Existing RFID system challenges are the actual reason why it still stands far from the market capital across the globe. According to largest RFID manufacturers such as Zebra, Alien Technology Inc., and Tyco Retail Solutions, RFID's future sustainability is dependent on how efficient the existing research and development proves to be in the future. RFID can also be used to track the emerging needs of humankind such as vaccines and medications in times of pandemic [16, 17].

Table 3 demonstrates the comparison among different anti-collision algorithms with severity of failure. It also mentions the complexity level of different algorithms along with the application of these algorithms in suitable domain along with its cost and efficiency.

4 Types of Tag Collision

Tag collision is mainly classified into two different categories: Reader and Tag collision (Fig. 2a), Reader and Reader Collision (Fig. 2b). Whenever two readers present in same vicinity and send the messaging signal to the tag located in their perimeter range then there are high chances that the tag won't be responding to any of them. In exceptional circumstances, if the frequencies of the readers are dissimilar, the tag will not respond to any of the readers. In such a Reader to Tag Collision, the tag will never listen to the communication requests send by the readers [14].

In some cases, when a reader send a strong signal even when the reader's interrogation perimeter is small, the reader may develop a large interference range. Reader and Reader Collision occurs when multiple readers are placed in the same interference range and use the same frequencies of other readers at the same time, as shown in Fig. 1 (Fig. 2b) [4, 5].

5 Risks of Tag Collision

Tag collision is referred as a major threat to any of the RFID system, especially when it comes to the real-life application, the collision of tag leads to more complexity in supply chain management system. Hence, it is important to detect such situation and handle it at the earliest. Such problem occurs in the presence of a huge number of Tag/Transponder devices in a perimeter of the RFID system. The tags gets energized by the multiple RFID systems and starts reflecting the radio wave signals back to the multiple or a single reader simultaneously [15].

I. Failure of the RFID system
II. Losing the inventory item tracking details
III. Security breach in the vicinity
IV. Theft detection if the valuables are attached to RFID tag.

6 Avoiding RFID Tag Collision

Failure of an RFID system in a supply chain management challenge poses a significant threat to inventory management; such failures can temporarily suspend the supply chain of carriers and goods; therefore, having a sustainable and robust RFID system should be given utmost importance. RFID systems can be made more secured by using cryptographic algorithms and two-way handshakes. Looking at all of the challenges, tag collision remains as one of the most recurring and difficult challenges that RFID systems face in today's world. Since its inception, various algorithms such as DBTSA and Q protocol have proven to be more efficient than previous algorithms. The below-mentioned are the description of two of the most versatile and widely famous anti tag collision algorithms, which has proven to be highly effective against tag collision.

6.1 Dynamic Binary Tree Slot ALOHA (DBTSA)

DBTSA (Dynamic Binary Tree Slot ALOHA) [12] was discovered as a leading constructive anti-collision tag algorithm. DBTSA employs dynamic frame utilization and efficient split frame in the algorithm to improve the effectiveness of high frequency passive tag anti-collision performance. The proposed algorithm presented in Fig. 3 provides a non-estimation based protocol, which maintains the protocol effectiveness even as the number of transponders in the RFID system increase. To achieve such a high efficiency, the algorithm focuses on reducing the idle slots and time required for tag identification [12].

As shown in Table 2 the DBTSA is more focused on adjusting the frame dynamically when the efficiency is increased by reducing the idle slots and identifying the tags in minimal response time. Here, the DBTSA achieves an efficiency of 0.441, which is 0.21 times higher than the traditional DFA-OS. DBTSA uses dynamic frame allocation by having a non-static frame size, which can store a higher or smaller number of tags depending on the Q-value.

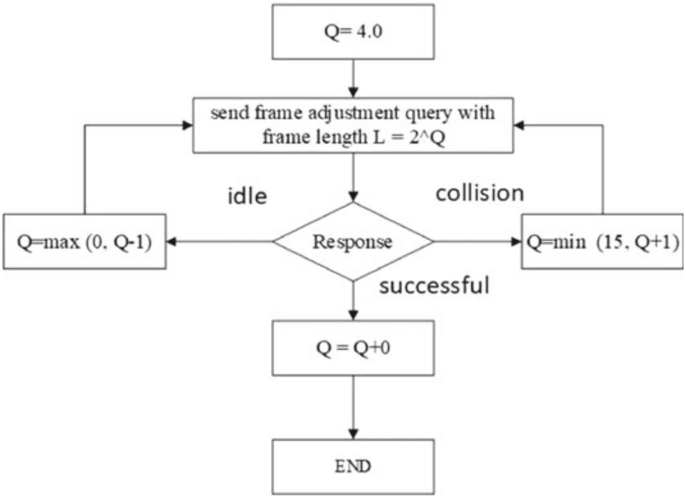


Fig. 3 Dynamic frame adjustment algorithm (DBTSA) [12]

Table 2 DBTSA efficiency matrix [12]

Proposed algorithm—dynamic binary tree slotted ALOHA (DBTSA)					
Number of tags	Number of idle slots	Number of collisions	Identification time (ms)	Number of iterations	System efficiency
0–50	31–38	30–34	140–180	80–90	0.441–0.420
100–500	392–480	344–345	800–970	850–1050	0.411–0.400

Table 3 Comparison of ALOHA based tag anti-collision protocol

Protocol name	PA	SA	FSA	DFSA	Q
Protocol feature	Tags send signals at random time intervals, in case of collision tags resends the signal to reader	Tag sends ID in synchronous form of time slots, if collision then resends in random time interval	All tags respond once per frame only	Tags send signal once every window. Reader set tag analysis function to change frame size	Reader sets value of Q dynamically based on types of reply sent by tags
Disadvantages	When number of tags is high in vicinity the collision increases	High quantity of tags results in more collision so reader requires synchronization with tag	Uses fix size of frame	Can't process to next frame without finishing the current frame process	May face lower throughput while adjusting value of Q when frame size is greater than quantity of tags
Efficiency (%)	18.40	36.80	36.80	42.60	36.80
System cost	Very low	Low	Expensive	Expensive	Expensive
Complexity	Very easy	Easy	Medium	High	Medium

6.2 *Q Protocol*

Q protocol remains as a basic framework for Dynamic Frame Slotted ALOHA (DFSA), which centers on manipulating the size of the frame. Based on the last feedback received, RFID tag sends the value of Q to tag. Depending on number of tags considered per round, Q_{New} is calculated. Q algorithms have two fundamental variables Q and a constant remaining C, where the range of the variable Q can be from 0.0 to 15.0 (in Fig. 4). The Q protocol uses 3 different types of commands:

- *Query Command*: It is transferred through reader to the transponder present in the vicinity and forces the transponder to choose a slot number to initialize the process of identification with providing a new value assigned to the Q.

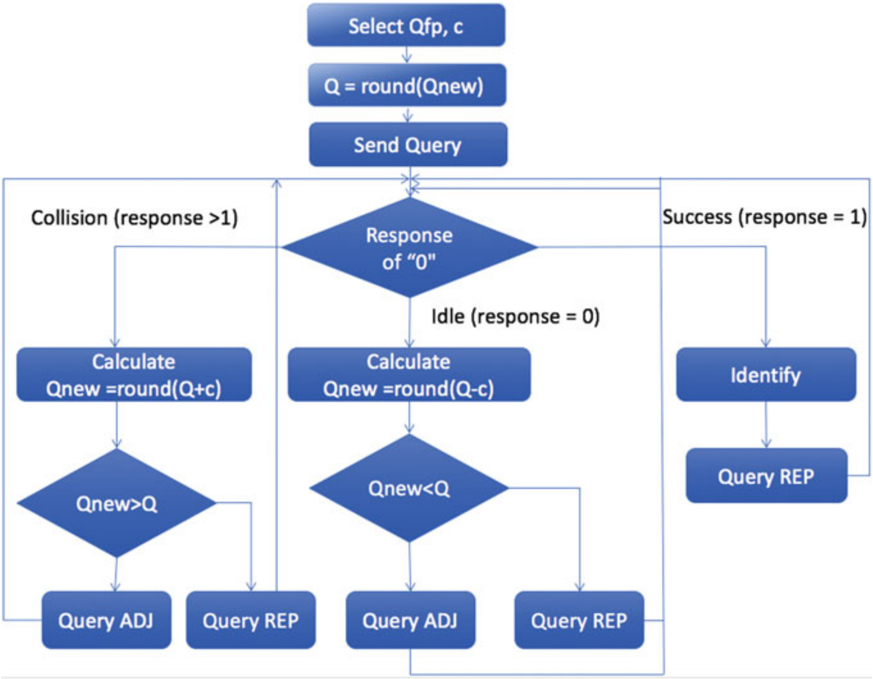


Fig. 4 Q protocol flow chart [11]

- *Query Adjust Command*: It is used to send instructions to every transponder to decrease, grow and maintain the value of Q unchanged. According to the algorithm value of Q, it can be increased or decreased based on the value of C.
- *Query Rep* comes to the picture when the tags are supposed to decrease the value of Q.

DBTSA (Dynamic Binary Tree Slot ALOHA) and Q algorithm both have their own pros and cons, DBTSA is proven to provide 43% efficiency which is 7% higher than Q protocol but Q provides the preprocessing of next frames by processing one frame in one slot, whereas DBTSA does not proceed to next frame without finishing the process of current frame. So, it can be concluded that the DBTSA is idle for warehouse with high flow of RFID tags but when the slots are sitting for long and have less flow activity, Q protocol seems to be better than DBTSA. In terms of avoiding the rate of active tag collision, Q and DBTSA are more efficient to include tag in frames by reducing the collision rate.

7 Comparison of ALOHA Based Anti-collision Protocols

As shown in Table 3. PA (Pure ALOHA), SA (Slot ALOHA), FSA (Frame Slot ALOHA), DFSA (Dynamic Frame Slot ALOHA) and Q algorithm are proven to be the best algorithmic protocols to favor tag collision avoidance in RFID system. The systematic comparison between these algorithm prove the DFSA as the most efficient algorithm to avoid the collision of the tag in any RFID system. On the other hand, DFSA possess the threat of having the size of frame more static, while the Q algorithm provides the dynamic frame allocation to make it more useful in case of handling high intense perimeter with the presence of huge number of tags and readers. By considering the complexity of Q and DFSA, the DFSA appears to be more complex than Q protocol. DFSA manipulates the frame size depending on the number of tags sending signals and reader changes function to change the frame size. This algorithm focuses on the reduction of slots by identifying and sorting the tags without crashing due to tag collision in such a way this algorithm focuses on the reduction of slots [1, 11].

8 Future of RFID (Radio Frequency Identification) Technology

In 2014, the global market revenue of RFID technology was barely 9 billion US dollars. This has multiplied three times in a span of 8 years by making global market revenue to 27 billion US dollars as shown in Fig. 5. This strongly indicates the fact that RFID will continue to grow stronger in the industry and hence more and more research and development works will be initiated in future to improvise the problems existing within RFID systems such as tag and reader collision [16].

9 Conclusion

Tag collision is one of the major concerns present in the field of RFID, Tag and Tag, Reader and Reader collision pose a threat to entire RFID system, which can be used for supply chain management or inventory tracking in warehouses. RFID also present opportunity in the global market due to its cost effectiveness and ability to integrate RFID technology into IoT. Along with that RFID provides different advantages such as high reliability, energy savings and low maintenance cost.

This paper has successfully discussed about the RFID processes and proposes the most efficient and dependable tag anti-collision protocols. The extensive literature review carried out in this study demonstrates that a sufficient research and development has been conducted in this field. Whereas the current efficiency of the

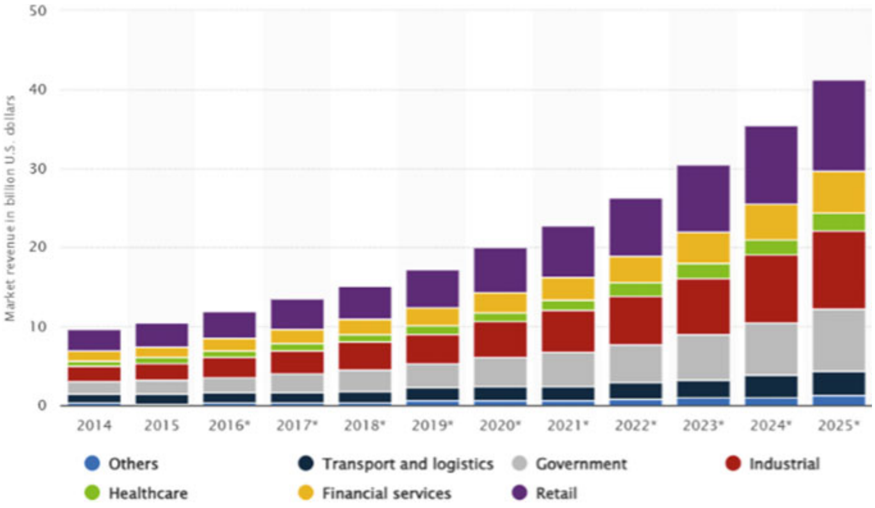


Fig. 5 RFID worldwide market revenue in billion U.S. dollars (credit statista) [18]

protocols indicates that more R&D is required to overcome the emerging RFID challenges. The key findings of the proposed study were that various algorithms and protocols have proven to be more efficient than the traditional methods over time, resulting in increased complexity. RFID must become more robust in the future, and anti-collision protocols play a critical role in this regard.

References

1. Chechi D, Kundu T, Kaur P (2012) The RFID technology and its applications: a review. Int J Electr Commun Instrum Eng Res Dev (IJECIERD). 2:109–120
2. Jung K, Lee S (2015) A systematic review of RFID applications and diffusion: key areas and public policy issues. J Open Innov 1:9. https://doi.org/10.1186/s40852-015-0010-z
3. Ibrahim AAA, Nisar K, Hzou YK, Welch I (2019) Review and analyzing RFID technology tags and applications. In: 2019 IEEE 13th international conference on application of information and communication technologies (AICT), pp 1–4. https://doi.org/10.1109/AICT47866.2019.8981779
4. Assariana A, Khademzadeh A (2018) A beacon analysis-based RFID reader anti-collision protocol for dense reader environments. Comput Commun 128:18–34
5. Guo Z, Yuan J, Gu J, Liu Z (2010) Research of tag anti-collision technology in RFID system. In: Luo Q (eds) Advances in wireless networks and information systems. Lecture notes in electrical engineering, vol 72. Springer, Berlin. https://doi.org/10.1007/978-3-642-14350-2_29
6. Haoxiang W, Smys S (2021) A survey on digital fraud risk control management by automatic case management system. J Electr Eng Autom 3(1):1–14
7. Liu Z, Guan Z, Shang K, Cai W (2013) Anti-collision algorithm for RFID based on continuous collision detection. Telkomnika Indones J Electr Eng 11. https://doi.org/10.11591/telkomnika.v11i12.3661

8. Liu X, Yin J, Liu J, Zhang S, Xiao B (2022) Time efficient tag searching in large-scale RFID systems: a compact exclusive validation method. IEEE Trans Mob Comput 21(4):1476–1491
9. Yang Q, Liu X, Guo S (2021) No wait, no waste: a novel and efficient coordination algorithm for multiple readers in RFID systems. In: 2021 IEEE/ACM 29th international symposium on quality of service (IWQOS), pp 1–10
10. Pal K (2019) RFID tag collision problem in supply chain management. Int J Adv Perv Ubiquit Comput 11:1–12. https://doi.org/10.4018/IJAPUC.2019070101
11. Cmiljanic N, Landaluce H, Perallos A (2018) A comparison of RFID anti-collision protocols for tag identification. Appl Sci 8(8):1282. https://doi.org/10.3390/app8081282
12. Memon MQ et al (2018) Improving efficiency of passive RFID tag anti-collision protocol using dynamic frame adjustment and optimal splitting. Sensors (Basel, Switzerland), vol 18(4):1185. https://doi.org/10.3390/s18041185
13. Tan X, Wang H, Fu L, Wang J, Min H, Engels DW (2018) Collision detection and signal recovery for UHF RFID systems. IEEE Trans Autom Sci Eng 15(1):239–250. https://doi.org/10.1109/TASE.2016.2614134
14. Zhang J, Lyu Y, Patton J, Periaswamy SCG, Roppel T (2018) BFVP: a probabilistic UHF RFID tag localization algorithm using Bayesian filter and a variable power RFID model. IEEE Trans Ind Electron 65(10):8250–8259
15. Zhang D, Huang H, Jo M (2015) Future RFID technology and applications: visions and challenges. Telecommun Syst 58:193–194. https://doi.org/10.1007/s11235-014-9865-8
16. Nayak R (2019) Challenges and future directions of RFID technology. https://doi.org/10.1201/9781351238250-9
17. Sungheetha A (2021) COVID-19 risk minimization decision making strategy using data-driven model. J Inf Technol 3(01):57–66
18. Alsop T (2022) Global RFID technology market revenue 2014–2025. https://www.statista.com/statistics/781338/global-rfid-technology-market-revenue/. Date of Publication: 2 Mar 2022

BizGuru 1.0: Design and Development of a Mobile-Based Digital Marketing Guide for Elderly

Ahmad Sofian Shminan, Nur Zulaikha Mohamed Aziyen, Lee Jun Choi, and Merikan Aren

Abstract BizGuru 1.0 is an online learning platform using mobile devices known as mobile-based learning. It is a modernized alternative to acquiring knowledge which is suitable with the current digitalized environment. BizGuru provides learning materials that promote business-related knowledge, focusing on Digital Marketing. However, in this study, the mobile application design will be focusing on the elder's group to cater for their needs. The target users are people aged 60 years old and above, who use an Android smartphone and are interested in gaining new knowledge. The purpose of the proposed application is to help these retired elderlies find an alternative that enables them to gain income at late age to continue supporting their living expenses. With the current pandemic situation and how they are often related to poverty, both circumstances result in the elders having to struggle to survive financially. Therefore, by using BizGuru, the elderlies do not only get to familiarize themselves with modern devices, but also they could look for other alternatives to gain income and avoid poverty which helps to fulfil the 1st goal of Sustainable Development Goals (SDG) on the eradication of poverty issues. Besides, this proposed application also provides learning opportunities for elderlies who have the desire to gain knowledge at late age which can help fulfil the 4th goal of SDG which is promoting life-long learning opportunities for all.

Keywords Mobile-based learning · Digital marketing · Elderly community

A. S. Shminan (✉) · N. Z. M. Aziyen · L. J. Choi · M. Aren
Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, 94300
Kota Samarahan, Sarawak, Malaysia
e-mail: sasofian@unimas.my

N. Z. M. Aziyen
e-mail: zulaikhaaziyen@gmail.com

L. J. Choi
e-mail: cljun@unimas.my

M. Aren
e-mail: amerikan@unimas.my

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_37

1 Introduction

The world is now becoming more digitalized with the aid of Information and Communication Technology (ICT) in the forms of mobile phones, computers, laptops, tablets, and TVs until it becomes a necessity in everyday life. ICT has upgraded human's life such as transforming handwritten letters to electronic mails, in-store shopping to online shopping, and traditional classroom to online learning [1]. This fast-paced technology advancement took place due to the impact of the Covid-19 pandemic which has also increased the e-commerce trends worldwide, including Malaysia. The extensive growth of e-commerce is because of the fact that people are mostly adapting to the new norm which is keeping a safe social distancing and staying indoors to stop the spreading of the virus [2, 3]. The gradual development of ICT in human life since years ago also indicates the significance of technologies in improving the quality of life. With the aid of devices, the productivity and performance of an organization are able to grow positively. However, despite the economic growth in developing countries like Malaysia, people who are much older, or the elderly group, are found to be at risk of poverty in later life [4].

An improper retirement planning is one of the factors that contribute to the rising number of poverty issues among retired elderlies. Due to the increasing life expectancy among the elderly, the retirement planning is very crucial to support their living expenses after retirement [5]. In fact, their only source of income after retirement will be the money that they earn from the past few years of working and the monthly allowance from their family members who are still working. Therefore, to overcome the issues and help the elderlies to reduce their financial burden, a mobile-based learning platform called 'BizGuru' is proposed. Setting up a small business is a great idea to gain income at late age. With the knowledge and guidance on digital marketing provided in BizGuru, the elderlies could promote their businesses on various online platforms. A digital marketing is used to allow the potential customers to learn about the product or service offered. According to Sharma [6], digital marketing is proven to be more effective than traditional marketing due to the growing number of people using digital technologies to shop. Therefore, companies that do not include digital marketing strategies in their businesses such as using blog to market their products are at loss [7].

To fulfill the 4th goal of Sustainable Development Goals (SDG) on promoting lifelong learning opportunities [8], the development of mobile-based learning platform is to encourage the elderlies to gain knowledge despite their old age. This learning platform also helps to create the sense of belonging among the elderlies in the digitalized society. This mobile-based learning platform consists of learning materials on digital marketing in the form of texts, visuals, and audios. With this proposed application, it helps to provide the people interested in digital marketing, the skills and knowledge which they could implement in businesses. Somehow, it will encourage the elderlies to open a small business to help them gain side income, and further to eliminate poverty issues at old age and next, to fulfill the 1st goal of SDG on eradicating poverty. This article is organized in such a way that the following section gives some related work, whereas the outline of our proposed BizGuru is given in Sect. 3 thereon. The Conceptual Process Flow and Model of BizGuru are described in Sect. 4, respectively. Section 5 gives explanations on findings and discussions. Finally, the conclusions are drawn and remarks are made in the last section.

2 Related Work

The development of a mobile-based learning application for elderlies are quite few in numbers. This is due to the needs to be catered for elderlies in terms of UX design are a bit complex as they have poor health conditions. However, these are several past studies that can be found relatable with the proposed application.

2.1 Smartphone for Seniors (S4S) Project

In year 2013, Barros et al. [9] had developed a health-related mobile-based learning application for older adults where they can learn to work out for fall prevention. They conducted the usability test at a local adult care centre for 3 sessions, as an improvement on the app was needed after each session to get the best results. This study shows that visual design is one of the most crucial elements in designing a mobile learning application for older adults. People who have their way of perceiving things, including older adults that are not familiar with the use of a mobile application, have a higher tendency to perceive the functional buttons wrongly. To overcome this problem, the use of icons to represent the text in a button is recommended in this study. It will help to improve the cognitive affordance of an element in a mobile application and avoid errors. Therefore, buttons with short text descriptions and representative icons will be included in the new mobile based learning application to improve the older adults' experience with technology.

2.2 Enhancing Islamic Knowledge via Short Messaging System (SMS)

Alkasirah and Nor [10] had conducted experimental research to explore the potential of mobile-based learning using SMS, based on Adult Learning Theory (Andragogy). The learning topic is about 'waqafa' or known as 'waqf', an Islamic knowledge. The adult learners were tested using the Solomon Four Group Design method and 'waqf' questionnaire whereby the learners were divided into 4 groups (2 groups are control groups, and another 2 groups are the experimental groups). Only the control groups received the SMS regarding 'waqf' knowledge. A pretest and posttest were conducted and the results have shown that the experimental group manages to score higher marks in the posttest which supports the use of mobile learning in education. Besides that, another finding in this research is that adult learners are self-directed as mentioned in the Adult Learning Theory (Andragogy). Hence, to develop a mobile-based learning application for adult learners, developers need to design content with easy comprehension such as avoiding the use of high vocabulary words. Using high vocabulary words could break the adult learners' self-esteem and lead to the lack of use of the mobile-based learning application.

2.3 eVideo Mobile

Patzer et al. [11] had researched a game-based work- related e-learning course which is a digital media in the hospitality industry known as, eVideo. The eVideo used to be web-based learning only. However, with the advancement of technology nowadays that supports mobile-based learning, eVideo is redesigned to be supported by smartphones. The eVideo mobile consists of learning materials including audio and subtitles which are developed by referring to Universal Design for Learning (UDL) and exercises with multiple-choice questions. A pilot study was conducted, and the results have shown that the adult learners were able to adapt to the transformation from web- based to mobile-based learning. Besides, eVideo mobile also made the learning contents accessible anytime and anywhere. However, the downside of the mobile-based is due to the size of the screen compared to web-based. Therefore, to overcome this problem, the new mobile-based learning should be designed using texts with an appropriate size and suitable font for adult learners.

2.4 The SenApp Project

Leen-Thomele et al. [12] had developed SenApp to evaluate the mobile learning concept on older adults in France and Germany. The SenApp focuses on educating older adults on how to use the Skype application and e-mail. There is a total of 12 short learning materials that explain the communication application. A pilot study was conducted and a total of 40 participants (20 learners from each country) was involved. The older adult learners were required to fill out the questionnaires after finishing the learning materials module. The learning materials were rated good by the learners although they still require some improvement. Older adults' learner with low education background gives a low rating on the SenApp because they have a lack of experience in e-learning and the learning materials are quite hard to be understood. This leads them to spending a shorter time on learning due to low self-confidence and motivation. Therefore, to build a new mobile-based learning application for older adults, the variables such as different target users' educational background should be considered. By reducing the amount of learning material for mobile-based learning, the older adult learners could have a better learning experience with less stress and help them to cope with technology advancement.

2.5 iPractice: Tablet-Based Home Practice Program in Aphasia Treatment

Kurland et al. [13] had conducted a pilot test on the effectiveness of a tablet-based home practice program in aphasia treatment. The application used is known as 'iPractice' which is specifically designed for iPad users. The participants recruited in this study are 55–81 years of age, with language deficiency due to past accidents that had affected the brain responsible for linguistics. Since the participants are at old age, the mobile application is designed with simplicity to provide ease of use in navigating while using the application for people not familiar with the technology. The 'iPractice' application contains 2 types of interactive books: one with objects and one with actions, and each book contains 20 words for the participants to practice. Half of the number of words (10 words) were trained in the treatment program and the others were not trained. Due to the participants' deteriorating eyesight due to old age, the videos and tasks given were able to be viewed in full screen. The participants were able to view the video unlimited times during self-practice. The minimum required time to practice was 20 min in 5 or 6 days every week. After every 6 months, the participants were required to follow a check-up at the clinic to check the progress at least once. The results of this study were positive, as they managed to improve their linguistic ability, although there were some complaints received from the users due to their boredom in using the application for practicing for a long period of time. Therefore, in the proposed application designed for the elderlies, each learning material should be designed with simplicity to reduce the time taken for learning and to avoid the feeling of irritation, which will lead to the application being unused.

2.6 Reflection

Therefore, by referring from the previous studies relating to mobile-based learning on the elderlies [7–11], it shows that the elderlies are able to gain knowledge despite their old age. In fact, with the mobility of mobile-based learning, it gives the opportunities for learners to learn at any time and any place. However, there are disadvantages of mobile-based learning on this group such as poor eye-sight and lack of familiarity with the devices. Therefore, to overcome this issue, developers need to implement the UX Design knowledge [14] and align with self-directed learning theory [15] to cater for the needs for learners at old age and provide sufficient information for them to navigate with ease while using the proposed mobile application.

3 Methodology

Instructional System Design (ISD) is a structured development of a digitalized learning platform that applies the instructional theory to establish a set of effective instructions to aid the learning process. In other words, it functions to help people gain knowledge successfully using technology devices [16, 17]. A well-designed tool for an educational purpose helps to improve the learner's ability to learn and to become self-directed [18]. Therefore, developers used ISD models as a reference to plan, create high-quality instructions, and build the system. Besides, a well-known ISD model that is commonly used in designing and developing system is ADDIE model [19]. ADDIE model consists of 5 phases: analysis, design, development, implementation, and evaluation. Each of these phases in Fig. 1 gives out a necessary result to move to the following stage of the development process [20, 21].

3.1 Analysis

Target Audience

Target audience is the group of people who are identified to likely be interested in the product offered. BizGuru application design is focused on catering to the needs of old learners. Therefore, the target audience in this study is a group of retired elderlies aged 60 years old and above, regardless of gender and ethnicity. The criteria are that they must be interested to learn new knowledge and own an Android smartphone. A stable internet connection is required to download and install BizGuru. Apart from that, this target group should have a big interest in learning digital marketing skills or improving their business strategy.

Gap Analysis

The elderlies are found to be at risk of poverty at later life [2, 3]. Therefore, to help the elderlies find an alternative to gain income at old age, a mobile-based learning application on digital marketing is proposed. This will allow the elderly to adapt with the current technology and help them earn money using an online platform which is easier and which needs less energy and physical movement.

Desired Outcome

By exposing the elderlies to digital marketing, they may be encouraged to run a small business to support their living expenses. The elderlies can promote their products or services made from the skills or creativity that they had learned personally throughout

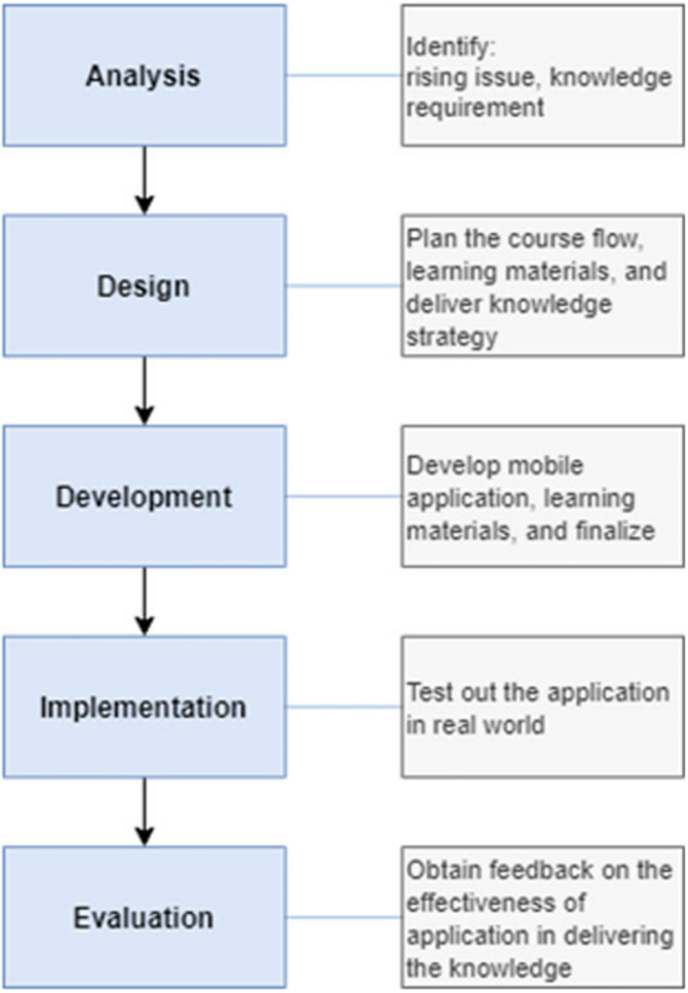
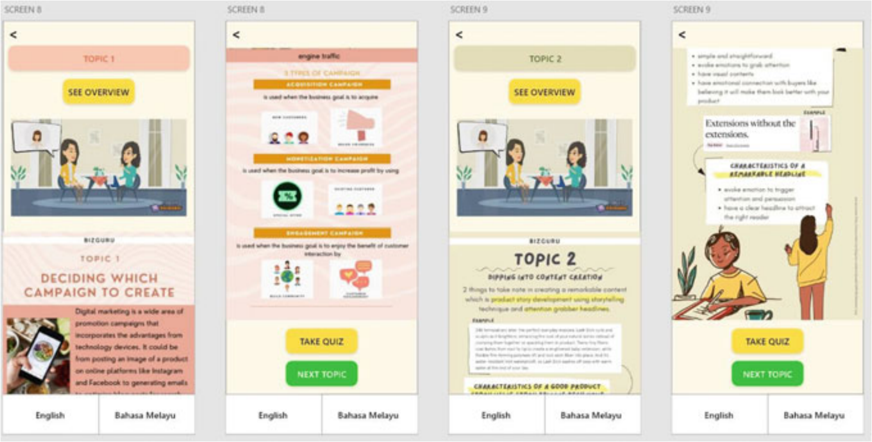

Analysis
Identify:
rising issue, knowledge requirement
Design
Plan the course flow, learning materials, and deliver knowledge strategy
Development
Develop mobile application, learning materials, and finalize
Implementation
Test out the application in real world
Evaluation
Obtain feedback on the effectiveness of application in delivering the knowledge

Fig. 1 ADDIE model flowchart

their long lives. As a result, the number of aged people with financial problems or at risk of poverty can be reduced and this, in the long run, can provide a better quality of ageing life [8]. Besides that, the elderlies would also get to prevent old-age disease by learning new knowledge during their leisure time, as a form of brain exercise.



Digital marketing is a wide area of promotion campaigns that incorporates the advantages from technology devices. It could be from posting an image of a product on online platforms like Instagram and Facebook to generating emails

is used when the business goal is to acquire

is used when the business goal is to increase profit by using

is used when the business goal is to enjoy the benefit of customer interaction by

2 things to take note in creating a remarkable content which is product story development using storytelling technique and attention grabber headlines.

- simple and straightforward
- evoke emotions to grab attention
- have visual contents
- have emotional connection with buyers like believing it will make them look better with your product

Extensions without the extensions.

- evoke emotion to trigger attention and persuasion
- have a clear headline to attract the right reader

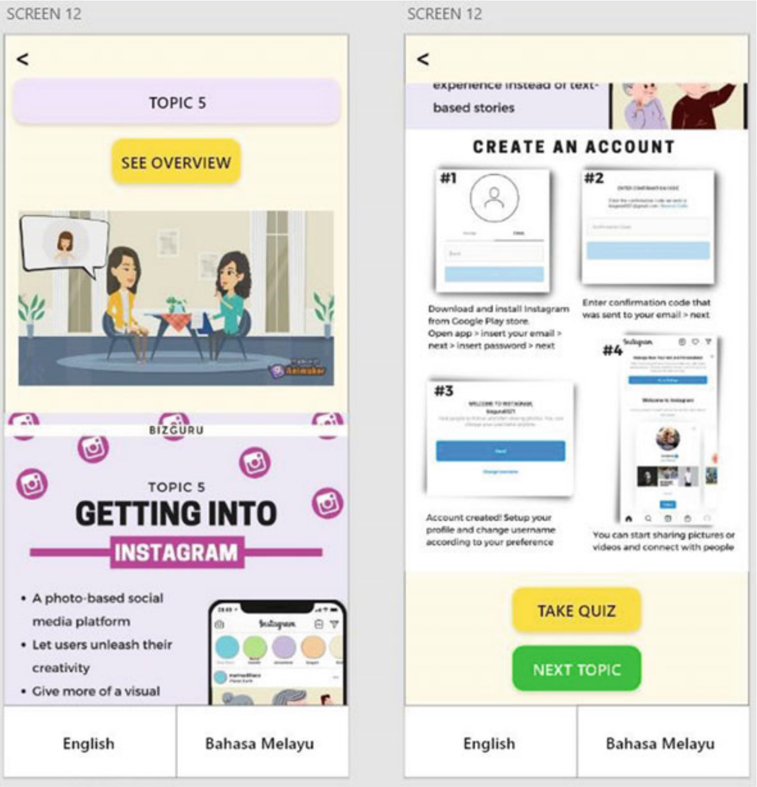
Fig. 2 Learning screen

3.2 Design

The high-fidelity prototype of BizGuru was made using Adobe XD software which is well-known for its impressive user interface (UI) and user experience (UX) designing and tools to build the prototype for mobile applications that allow a smooth process of moving from static mockups to interactive prototypes [22, 23]. The Bahasa Melayu version has a similar interface design with English version except for the language usage in audio, video, and writing. Apart from that, the course purpose, course flow, learning materials, and strategy to deliver the knowledge is figured in this stage. The course flow and learning materials were referred from a well-known book in the digital marketing industry titled ‘Digital Marketing for Dummies’ by Deiss and Henneberry [24]. The strategy to deliver knowledge successfully is by incorporating visual presentation and audios are shown in Fig. 2. Since this proposed mobile-based learning application is meant for elderlies whereby it is common for them to have poor eye sight, the design interface of each screen will include the UX Design principle to ensure that they get the best learning experience [14].

3.3 Development

Ionic5 was used as the programing tool to build the real application. Ionic5 is known for its cost-effective development with high speed performance and functionality, it has an open source feature, and uses simple programming language. This software uses Hyper Text Markup Language (HTML), JavaScript (JS) and Cascading Style Sheet (CSS). Next, Visual Studio Code (VSC) software was chosen as the platform to write and compile all the coding built for BizGuru interface design and functionality. The pages created in VSC consist of a subpage focusing on the layout and placement of element using HTML, another subpage focusing on the design, decorations, and styling using CSS, and subpages focusing on the functionalities using JS. Also, Windows Powershell was also used in giving commands for development purpose such as to include the Android Studio extension in Ionic5 project for testing the developed application using android emulator and build APK file of BizGuru. This APK file is a type of an executable file which is used to compile the mobile application and allow both the distribution and installation of application. Lastly, BizGuru learning materials as shown in Fig. 3, such as infographics and high-quality animation videos for every topic were developed using Canva and Animaker online software.



- A photo-based social media platform
- Let users unleash their creativity
- Give more of a visual

experience instead of text-based stories

Download and install Instagram from Google Play store.
Open app > insert your email > next > insert password > next

Enter confirmation code that was sent to your email > next

Account created! Setup your profile and change username according to your preference

You can start sharing pictures or videos and connect with people

Fig. 3 Application setup with learning assessment

3.4 Implementation

As the cases of Covid-19 pandemic is steadily rising, the data collection procedure had to be transformed into the online method. Therefore, two different websites were created using cloud website builder that is meant for expert evaluation and pilot and usability test. These websites helped to organize the documents and materials needed to conduct the evaluation session. The respective website link was given to the evaluators and participants through e-mail or WhatsApp message. In the website, there is an introduction page for explaining the concept and purpose of study, briefing notes and instruction to perform the required task, informed consent form to explain the terms and conditions after they had agreed to participate in this study, a video demonstration of BizGuru app, a Google Drive link containing an APK file of BizGuru along with the poster guide to install the application, and a link to two different online evaluation forms created using Google Form to collect the ratings and comments given by the evaluators and participants.

3.5 Evaluation

There are three phases of evaluation session done in this study namely expert evaluation, pilot testing, and usability testing. Expert evaluation is a type of assessment that relies on the individual expertise in that particular area of study to discover the potential issues that have arisen in the mobile application [25]. Meanwhile, the pilot testing is an additional step that acts like a rehearsal phase with a smaller number of participants before conducting the main study (such as usability testing) with a larger number of participants [26, 27]. Usability testing is the main research study aiming to test the system developed in real environment which requires the participants to perform, complete a task given, and provide feedback to the researcher [27]. BizGuru application overall performance was evaluated from an expert's and user's point of view. 3 expert evaluators with a strong foundation and knowledge in the technicality and usability of a mobile application interface design were recruited in this study. The instrument used in this evaluation session is known as Mobile Learning Usability Attribute Test (MLUAT) [28, 29]. Other than that, 7 retired elderlies (2 elderlies in pilot test, 5 elderlies in usability testing) had volunteered to participate in this study where they share a common characteristic of being an Android smartphone user. However, these elderlies are not experts in this field of study, as opposed to the expert evaluators. In fact, these elderlies are the target users who will use, learn and gain the knowledge from the mobile learning application developed. The instrument used in this evaluation session is known as System Usability Scale (SUS). The online evaluation form for MLUAT and SUS was created using Google Form with a 5-likert scale answer. The benefit of using 5-likert scale instead of 7 or more Likert scale is because it helps to produce a reliable quantitative data that is easy to be analysed, a universal method of data collection, and ideal method for a long questionnaire with multiple choice. Their ability to understand the functions in the mobile-based learning application was included in the feedback form. The usability and ease of use of the mobile-based learning application was measured throughout the entire interaction process and the feedback on the overall effectiveness of the proposed application to deliver the knowledge given.

4 Conceptual Process Flow and Model of BizGuru

A conceptual model is used to represent the whole process of a system. Flowchart is one of the best ways to show the overall cycle with a clear sequence of actions. Figure 4 shows the flowchart of BizGuru mobile-based learning application.

The BizGuru mobile-based learning application starts off with a splash screen welcoming the learners to use this application. Next, the learners need to choose their preferred language for learning. There are two different languages included in this application which is Bahasa Melayu and English. After choosing the language, the learners will be directed to another screen which shows all the learning topics available in this application ranging from Topic 1 to Topic 5. To start the lesson, the learner can choose any of the learning topics, preferably starting from the first topic as it has a hierarchical order. On each learning topic, the learner will be provided with a short overview to give a brief idea of the learning topic and a video that will help in providing a more detailed explanation. After watching the video, an infographic is provided on each topic which will help the learners to summarize their understanding in visual and textual forms. After the learning process, the learner will be assessed with a short quiz to evaluate their level of understanding. The quizzes will be given after each learning process with a 'Yes' or 'No' answer. Then, the learners will obtain their results on the spot with the feedback that was set up to be provided after the learner tapped on the answer button. If the learner wishes to improve their understanding, they are allowed to retake the learning topic or otherwise, and then proceed to the next learning topic. In the next learning topic, the learning flow is all the same. Lastly, learners are allowed to exit or access the learning materials at any time.

Figure 5 shows the conceptual design of BizGuru mobile-based learning application. The target users of the proposed application are elderlies who seek for a lifelong learning opportunity. Firstly, in the training phase, the learners will access the learning materials by reading a brief overview of the learning topic. Then, as the elderly face some difficulties reading on the small screen for a long period of time, a video explaining on the related topic will be provided after each topic overview. The group will then use their senses to capture the information shown on screen. The information captured through the hearing sense will be kept in the working memory on verbal mode. Meanwhile, the information captured using the eyes will be kept in the working memory on pictorial mode. The valuable information will be interpreted into knowledge and kept in the long-term memory. Next, the learners will undergo

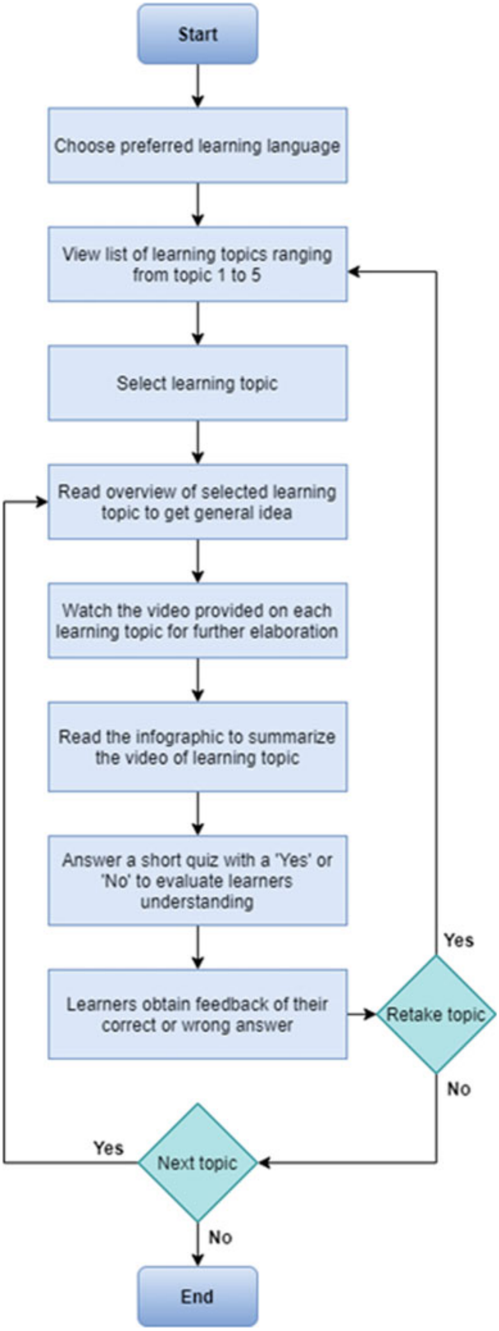

Start
Choose preferred learning language
View list of learning topics ranging from topic 1 to 5
Select learning topic
Read overview of selected learning topic to get general idea
Watch the video provided on each learning topic for further elaboration
Read the infographic to summarize the video of learning topic
Answer a short quiz with a 'Yes' or 'No' to evaluate learners understanding
Learners obtain feedback of their correct or wrong answer
Retake topic
Yes
No
Next topic
Yes
No
End

Fig. 4 Conceptual flow process

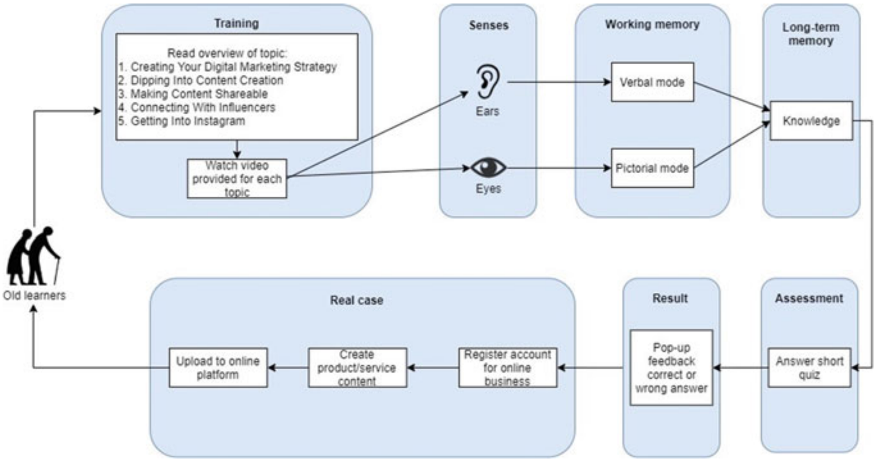

Training
Read overview of topic:
1. Creating Your Digital Marketing Strategy
2. Dipping Into Content Creation
3. Making Content Shareable
4. Connecting With Influencers
5. Getting Into Instagram
Watch video provided for each topic
Senses
Ears
Eyes
Working memory
Verbal mode
Pictorial mode
Long-term memory
Knowledge
Old learners
Real case
Upload to online platform
Create product/service content
Register account for online business
Result
Pop-up feedback correct or wrong answer
Assessment
Answer short quiz

Fig. 5 Conceptual model

an assessment phase which requires them to answer a short quiz to test their understanding and evaluate the effectiveness of the mobile- based learning application in delivering knowledge. These quizzes are provided at the end of each learning topic. The answers chosen will be verified and the learners will obtain their results immediately in the form of pop-up feedback. Lastly, after completing the whole course, the learners can use the knowledge learned to apply to real cases. Learners can start their small online businesses using Instagram and promote their products using the digital marketing knowledge.

5 Findings and Discussion

Evaluation process proves to be a significant step in creating a new application system. The main notion of the system testing is to ensure that the system launched can fulfil the user requirement and satisfaction successfully. The testing must be conducted prior to the proposed application development to ensure that a better product can be obtained before the mass production.

5.1 Subject Matter Expert

The subject matter expert (SME) offers sound knowledge and expertise in a specific subject, business area, or technical area for a project [25, 30]. SME also makes sure that the facts and details are correct so that the project's/program's deliverable(s) will fulfil the stakeholders' needs, also the requirements of the legislation, policies, standards, and best practices. To assess the application, expert received the online link to do the testing on the application. The expert then answered the questionnaire provided. From the finding, the expert evaluator was given the questionnaire to answer the MLUAT evaluation. There are six parts in the evaluation namely visibility status, leaner control and freedom, match between the system and real world, consistency, error prevention, recognition rather than recall and minimize information on screen. The total score from all three expert evaluator feedback was calculated to be 95.93 and the average total score for BizGuru was 4.36. By referring to the 5-likert scale used, low values such as 1 and 2 represent a negative attitude meanwhile higher numbers such as 4 and 5 represent a positive attitude [31]. The average score (4.36) belongs to the higher value group which confirms that BizGuru mobile learning application had acquired a positive attitude from the expert evaluators in the expert evaluation phase, implying a good functionality and usability of the application.

5.2 Usability Evaluation

Usability testing and System Usability Scale (SUS) are the technique and instrument used in the user testing phase- the practice of testing is the extent of the ease of use of the design, to be applied to a group of representative users. It entails an observation of users as they try to complete tasks and receive feedback from the users by way of interviews or questionnaires about user satisfaction on the product's prototype. According to Macefield [27], 5–10 respondents are the least number of respondents required for the usability testing. Nielson stated that elaborate usability tests denote a waste of resources, as the best results come from testing not more than 5 users and running as many small tests as possible [32]. This is due to the fact that when more tests are run on the user, the same results are to be obtained as the previous users where the first study with 5 participants is considered sufficient to find 85% of the usability problems. Therefore, 5 participants from elderly community are chosen at random to be part of the usability testing [27]. The feedback obtained will be gathered and analysed to be further improved in the application development. The SUS score per user was calculated to be 72.5, 77.5, 70, 72.5, and 75. All five users gave a rating of Agree [4] in question number 1 where they admitted that they would use BizGuru frequently. On the other hand, they also gave a significantly low rating on items depicted in questions 8 and 10 where they disagreed to the statement that BizGuru application was cumbersome to use, and they also disagreed to the statement whereby they needed to learn a lot of things before using BizGuru application. These questions are illustrated as negative questions, thus they received low ratings, depicting that BizGuru is easy to use. Besides, the average SUS score for BizGuru was calculated to be 73.5 (Grade B) which is considered as a good application, although it still requires an improvement on certain areas.

5.3 Related Theory and Methodological

The Self-directed learning and Andragogy theory was embedded in developing this application. Both theory defines adult learning as unique and different from general learning. It remarks that the elderly was classed as independent learners since they have a sense of responsibility from their life experiences. By looking at the facts, adults spend more time acquiring specific skills, which can be performed through reading, listening, observation, reflection, and exercise [15]. BizGuru 1.0 is an efficient device that combines learning material related to quail farming in a single platform. It is believed to be more efficient than the traditional website or book as the user can readily download the application and learn independently, which is referred from the theories embedded. The SDL theory describes a process where an individual takes the initiative in learning, including the learning materials and strategies, and evaluates the outcomes. It is opposite to teacher-directed learning or "pedagogy," which is considered an information transfer influenced by the outer in determining the learning outcome.

According to Farage et al. [14], the age factor plays a role in the physical changes of a person; therefore, a design guideline is referred to achieve the satisfaction level of users. In this study, to meet the elderly needs, the universal design principle related to the elderly is referred to in measuring the achievement of the development application. The most common decline in the elderly is the visual field; thus, color choices are essential in accommodating age-related visual impairment. The long-wavelength or "warm" colors are preferred to convey information that differs from the background. Simplicity is the word best defining the visual display, where the essential information is presented in a noticeable and uncrowded field. Short and precise learning materials matter as the degree of working memory in the elderly is shorter than in young adults. An application with complex information will create an overload of information to be processed for the elderly.

ADDIE, an Instructional System Design method [19] was embedded in the development of BizGuru 1.0. It is intended to solve a problem by developing and evaluating IT artefacts ranging from software, formal logic, or natural language. In this study, the issue of poverty among the elderly is chosen. ADDIE involves a process that enables an understanding of the problem addressed and the feasibility of the possible solution. BizGuru was made to disseminate information on digital marketing efficiently and to ease the learning for the target user (the intention and usefulness). In the hope that this application would help the target user to gain income from initiating their quail farming and business (the benefits). It aligned with the ADDIE method concerning the artifacts' intention, usefulness, and benefits. The expertise and target user were involved implicitly during the evaluation to receive direct feedback and to test the perception of the usefulness of proposed mobile based learning application. The most time-consuming process in this study is design and development, which is carried out non-stop even after evaluation. The feedback received during the evaluation will be considered, and changes to the application will be made.

6 Conclusion and Future Work

BizGuru is developed from the analysis of the weaknesses or disadvantages found from previous mobile learning application studies conducted from 2014 to 2019. The improvement has been made to provide a better learning experience for the old learners which also help to boost their cognitive skills and well-being by experiencing what is termed as 'quality ageing'. Besides that, the development of BizGuru which focuses on tending to the needs of the aging learners also helps to fulfil the 4th goal of Sustainable Development Goals (SDG) that is to provide a life-long learning opportunity [8]. As this mobile learning application promotes a digital marketing knowledge, the retired learners can take this opportunity to make it as an alternative to gain side income to continue supporting their living expenses in later life using technology devices. This benefit also helps to fulfil the 1st goal of SDG that is to eradicate poverty issues among the elderly group. Besides, there are a few limitations found in this study- for one, conducting the usability testing remotely amid the Covid-19 pandemic had caused the application designer to be unable to observe the users directly during the testing and evaluation process to get a more detailed remark and experience. There was also a language barrier issue, as one of the users in the usability testing phase had commented that it would be better if there is a Chinese version in the BizGuru mobile learning application. This is due to the differences of races available in Malaysia and the daily language usage by the participants. Hence, future studies are recommended to improve the functionality of the mobile learning application by connecting to a database which will allow a bigger space for learning contents and also record the learners' progress. By implementing a database architecture [33], it will also allow the application developers to monitor the learners' learning curves based on their achievements in the quiz section. Apart from that, the future mobile learning application should improve the cloud security features by adding specific mechanism for a slot registration and log-in page [34]. This will instill the learners with a sense of belonging and security, as their learning progress is kept confidential to themselves without exposure to other learners. Not all learners are comfortable sharing their achievements with other learners, especially the competitive and introvert ones. Lastly, by having an authentication page, it allows multiple users per device, and this means that the number of learners and users of the respective mobile learning application can increase.

References

1. Rothe F-F (2020) Rethinking positive and negative impacts of 'ICT for development' through the holistic lens of the sustainable development goals. Inf Technol Dev 26:653–669. https://doi.org/10.1080/02681102.2020.1756728
2. Galhotra B, Dewan A (2020) Impact of COVID-19 on digital platforms and change in E-commerce shopping trends. In: 2020 fourth international conference on I-SMAC (IoT in social, mobile, analytics and cloud) (I-SMAC), pp 861–866

3. Winarsih, Indriastuti M, Fuad K (2021) Impact of Covid-19 on digital transformation and sustainability in small and medium enterprises (SMEs): a conceptual framework. In: Barolli L, Poniszewska-Maranda A, Enokido T (eds) Complex, intelligent and software intensive systems. Springer International Publishing, Cham, pp 471–476
4. Masud J, Hamid T, Haron SA (2015) Measuring poverty among elderly Malaysians
5. Mohidin R, Jamal AAA, Arokiadasan CG et al (2013) Revisiting the relationship between attitudes and retirement planning behavior: a study on personal financial planning. Int J Multi Thought 3:449–461
6. Sharma DU, Thakur PKS (2020) A study on digital marketing and its impact on consumers purchase. Int J Adv Sci Technol 29:13096–13110
7. Bala M, Verma D (2018) A critical review of digital marketing. Social Science Research Network, Rochester
8. Poverty Eradication | Poverty Eradication. https://www.un.org/development/desa/socialperspectiveondevelopment/issues/poverty-eradication.html. Accessed 6 Jun 2022
9. de Barros AC, Leitão R, Ribeiro J (2014) Design and evaluation of a mobile user interface for older adults: navigation, interaction and visual design recommendations. Proc Comput Sci 27:369–378. https://doi.org/10.1016/j.procs.2014.02.041
10. Alkasirah NAM, Nor NMM (2019) Potential usage of mobile learning via short messaging system (SMS) for enhancing Islamic knowledge of adult learners
11. Patzer Y, Lambertz J, Schulz B, Pinkwart N (2018) Mobile online courses for the illiterate: the eVideo approach. In: Miesenberger K, Kouroupetroglou G (eds) Computers helping people with special needs. Springer International Publishing, Cham, pp 379–383
12. Leen-Thomele E, Hetzner S, Held P (2016) Mobile learning concepts for older adults: results of a pilot study with tablet computers in France and Germany. In: Zhou J, Salvendy G (eds) Human aspects of IT for the aged population. healthy and active aging. Springer International Publishing, Cham, pp 319–329
13. Kurland J, Wilkins AR, Stokes P (2014) iPractice: Piloting the effectiveness of a tablet-based home practice program in aphasia treatment. Semin Speech Lang 35:51–64. https://doi.org/10.1055/s-0033-1362991
14. Farage MA, Miller KW, Ajayi F, Hutchins D (2012) Design principles to accommodate older adults. Glob J Health Sci 4:2–25. https://doi.org/10.5539/gjhs.v4n2p2
15. Roberson DN (2004) The nature of self-directed learning in older rural adults. Ageing Int 29:199–218. https://doi.org/10.1007/s12126-004-1017-0
16. Shin S, Bae Y (2015) A study on the hierarchical instructional system design of software education by school system. J Korean Assoc Inf Educ 19:533–544. https://doi.org/10.14352/jkaie.2015.19.4.533
17. Shminan AS, Choi LJ, Sharif S (2020) AutiTEACCH: Mobile-based learning in a structured teaching approach for autistic children caregivers. In: 2020 international conference on informatics, multimedia, cyber and information system (ICIMCIS), pp 259–264
18. Uz R, Uzun A (2018) The influence of blended learning environment on self-regulated and self-directed learning skills of learners. Eur J Educ Res 7:877–886
19. Valverde-Berrocoso J, Fernández-Sánchez MR (2020) Instructional design in blended learning: theoretical foundations and guidelines for practice. In: Martín-García AV (ed) Blended learning: convergence between technology and pedagogy. Springer International Publishing, Cham, pp 113–140
20. Magliaro SG, Shambaugh N (2006) Student models of instructional design. Educ Tech Res Dev 54:83–106. https://doi.org/10.1007/s11423-006-6498-y
21. Castro MDB, Tumibay GM (2021) A literature review: efficacy of online learning courses for higher education institution using meta-analysis. Educ Inf Technol 26:1367–1385. https://doi.org/10.1007/s10639-019-10027-z
22. Figliolia AC, Sandnes FE, Medola FO (2020) Experiences using three app prototyping tools with different levels of fidelity from a product design student's perspective. In: Huang T-C, Wu T-T, Barroso J et al (eds) Innovative technologies and learning. Springer International Publishing, Cham, pp 557–566

23. Stoeva M (2021) Model and prototype of interactive assistant for compliant interface development—MayUI tool. In: 2021 international conference automatics and informatics (ICAI), pp 295–300
24. Deiss R, Henneberry R (2017) Digital marketing Fd, 1st edn. For Dummies, Hoboken
25. Perez RS, Fleming Johnson J, Emery CD (1995) Instructional design expertise: a cognitive model of design. Instr Sci 23:321–349. https://doi.org/10.1007/BF00896877
26. Shminan AS, Choi LJ, Barawi MH et al (2021) InVesa 1.0: the conceptual framework of interactive virtual academic advisor system based on psychological profiles. In: 2021 13th international conference on information and communication technology and system (ICTS), pp 112–117
27. Macefield R (2009) How to specify the participant group size for usability studies: a practitioner's guide. J Usability Stud 5:34–45
28. Fetaji M, Fetaji B (2011) Comparing developed MLUAT (mobile learning usability attribute testing) methodology with qualitative user testing method and heuristics evaluation. In: Proceedings of the 12th international conference on computer systems and technologies. Association for Computing Machinery, New York, pp 516–523
29. Arain AA, Hussain Z, Rizvi WH, Vighio MS (2016) Evaluating usability of M-learning application in the context of higher education institute. In: Zaphiris P, Ioannou A (eds) Learning and collaboration technologies. Springer International Publishing, Cham, pp 259–268
30. Le Maistre C (1998) What is an expert instructional designer? Evidence of expert performance during formative evaluation. ETR&D 46:21–36. https://doi.org/10.1007/BF02299759
31. Norman G (2010) Likert scales, levels of measurement and the "laws" of statistics. Adv in Health Sci Educ 15:625–632. https://doi.org/10.1007/s10459-010-9222-y
32. Bangor A, Kortum PT, Miller JT (2009) Determining what individual SUS scores mean: adding an adjective rating scale. Undefined
33. Mishra A, Jaiswal A, Chaudhari L, Bodade V (2022) Health record management system—a web-based application. J IoT Soc Mob Anal Cloud 3:301–313
34. Shakya S (2020) Survey on cloud based robotics architecture, challenges and applications. JUCCT 2:10–18. https://doi.org/10.36548/jucct.2020.1.002

Development of Secure Cloud-Based Healthcare Management Using Optimized Elliptic Galois Cryptography

V. Gokula Krishnan, D. Siva, S. MuthuSelvi, T. A. Mohana Prakash, P. A. Abdul Saleem, and S. Mary Rexcy Asha

Abstract The ever-increasing amount of e-medical data poses a security risk because of technological advancements in the healthcare business. An unstructured and large amount of unstructured data is generated by the healthcare data management system because of the wide variety of data formats that are used to capture patient information. Branches of hospitals can also be found in different parts of a city or state. Health information on patients that is kept in multiple places must be merged from time to time for research purposes. Cloud-based healthcare management systems can be an effective solution for storing and managing health care data more

V. Gokula Krishnan (✉)
Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Thandalam, Chennai, Tamil Nadu 602105, India
e-mail: gokul_kris143@yahoo.com

D. Siva
Department of Computer Science, Faculty of Science and Humanities, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu 600089, India
e-mail: d.siva885@gmail.com

S. MuthuSelvi
Department of CSE, Vel Tech Multi Tech Dr Rangarajan Dr Sakunthala Engineering College, Avadi, Chennai, Tamil Nadu 600062, India
e-mail: muthuselvis@veltechmultitech.org

T. A. Mohana Prakash
Department of CSE, Panimalar Engineering College, Poonamallee, Chennai, Tamil Nadu 600123, India
e-mail: tamohanaprakash@gmail.com

P. A. Abdul Saleem
Department of CSE, CVR College of Engineering, Mangalpally, Hyderabad, Telangana 501510, India
e-mail: drsaleemprincipal@gmail.com

S. Mary Rexcy Asha
Department of IT, Panimalar Engineering College, Poonamallee, Chennai, Tamil Nadu 600123, India
e-mail: rexcyasha@gmail.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*, Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_38

effectively. However, security is the most pressing issue with a cloud-based healthcare system. Elliptic Galois Cryptography (EGC) is used in this study to encrypt medical data files, and the value of the Galois field is determined using the Mayfly Algorithm. As a result, the proposed model is referred to as a "optimal EGC". Use of the elliptic curve over a Galois field in elliptic curve cryptography reduces rounding errors. The healthcare data is protected in terms of both confidentiality and integrity when it is shared via the health cloud. Experiments have shown that the ideal solution can be computed more quickly in terms of file upload and download speeds as well as key generation and generation time. Additionally, it protects healthcare data from being tampered with during transmission via the health cloud.

Keywords Cloud computing · Security · Key generation · Mayfly algorithm · Elliptic Galois cryptography · Medical data

1 Introduction

When it comes to health care, the wearable technology is kindly spreading its tentacles to include not only all walks of hominid life, but also challenge the stowed huge health care data [1]. There are many different kinds of health care data, and they are made at a high rate, making it challenging to keep them locally. So the need for medical media applications like multimedia email, presentations, high quality audio and video sharing and shared papers has increased exponentially [2, 3]. In the healthcare industry, all patient records must be stored in the Cloud for future reference. This study examines the day-to-day operations of the health care business. Computational and processing challenges are plaguing the current health care business. Physical storage, security and medical errors are inherent issues in the traditional healthcare industry. It is critical to keep patient records safe since they include sensitive information. Patient data is being compromised by a number of issues in the current system. It takes up a lot of memory space, which is not cost-effective [4, 5].

In order to protect patient information, the cloud offers a high level of security. Prescription retrieval is easy because the patient's info is stored in the cloud, so they may access it whenever they want [6]. Because the data is stowed in the cloud, anyone with a mobile device, such as a smartphone or PDA, can access it without requesting specific permission. One of the most dynamic sectors of the information technology business is the healthcare sector, where cloud computing is becoming increasingly important [7, 8]. Internet-enabled devices can access health-care information throughout the world thanks to cloud computing technology. The medical community can also benefit from the exchange of resources and information with other leading researchers in the same subject around the globe. To improve and develop the current health care industry, this study is being conducted. Anuradha et al. [9] Despite the advantages of cloud-based health care schemes, many doctors and healthcare institutions are reluctant to utilize them because of the risk of data breach. Also, because of the subtle nature of the data being kept and retrieved, numerous health care organizations are avoiding public cloud and installing private cloud services in its place [10].

The electronic health annals must be securely transferred via networks in order to protect patient confidentiality and data integrity. In [11, 12], the drawbacks of using a cloud database to store health information are discussed. For the protection of electronic health records during transmission, a cryptosystem is typically needed. Security of user data has been achieved by the employment of conventional methods such as the Rivest–Shamir–Adleman (RSA)-based system [13, 14], and the encryption of user data [15]. Large keys and complicated computations make it difficult to use such systems on mobile devices, which is another drawback of these approaches. The use of Elliptic Curve Cryptography (ECC) in cryptosystems has grown in recognition and use during the past several years. In order to achieve great scalability and efficiency, ECC reduces the key complexity by using smaller key lengths. In order to store and share health data in the cloud, users must encrypt their data before uploading it to the servers. The health cloud will benefit from this research because of the optimized EGC implementation, which will allow it to provide better healthcare services while maintaining the integrity of patient data.

In order to build an Access Control List (ACL), the TTP-CS receives healthcare data, a list of CUs, and the necessary criteria from the data owner. Later, the encrypted data is transferred to the HC on behalf of the CU for storage. If the CU is interested in accessing health data files, the TTP-CS will get a download request. The following are a few of the methodology's most significant benefits:

- Stronger encryption methods ensure the safety of patient data in the health cloud.
- To ensure data security and speed, the health cloud uses an EGC mechanism that is tuned for scalability and uploading speed.
- It provides a high level of protection for data from insider threats.

The rest of this paper is prearranged as follows. Section 2 delivers a comprehensive review of the relevant scholarly literature. With the help of an overall system design, the proposed methodology is clarified in Sect. 3. Section 4 details the proposed system's performance evaluation, and Sect. 5 closes the article with recommendations for future research.

2 Related Works

In order to protect against smart health threats, Zhang et al. implemented CP-ABE (Ciphertext Policy Attribute Based Encryption). Smart healthcare's application of CP-ABE brings with it a unique set of challenges [16]. It was created to address these issues: a smart health access control system that takes privacy into account. In PASH, only the name attribute is made public, while the value of the access policy attribute is hidden in encrypted smart health records. In addition, attribute values typically contain more private information than other types. In this decryption test, PASH is able to successfully decrypt SHR (it requires few bilinear-pairings).

Mobile Healthcare Social Networks are plagued by privacy concerns (MHSN). MHSN profile matching and data sharing are planned by Huang et al. [17] in the cloud. Identity Based Broadcast Encryption (IBBE) is used to outsource encrypted data to the cloud (IBBE). In addition, the doctor's group receives data fast and safely. Using attribute-based conditional data re-encryption, the doctor's referral is disseminated throughout the network to another doctor. A new enciphered text is generated from the encrypted one (without leaking the sensitive information).

While integrating and exchanging E-health information, this book sought to address security and privacy concerns by providing a solution for Internet applications. Bao and colleagues [18] have presented a signal scrambling technique based on the application layer. To protect patient information, a minuscule amount of data is used to scramble the original. VOLU It uses either a random generator or a piece of data to derive the small data.

Masood et al. [19] established a six-step architecture for measuring the patient's physiological characteristics in Sensor Cloud Infrastructure (SCI). It begins with a preliminary selection, followed by an assessment of the patient's physiological parameters and a security analysis. Finally, it estimates the functioning of the system. Cloud computing is a promising tool for healthcare data security. It's a requirement, along with other security measures, while communicating electronically. Mbonihankuye et al.'s [20] leading strategy is the Health Insurance Probability and Accountability Act (HIPAA): [20]. Different analytical and conservational procedures can be used to ensure that healthcare data is properly recorded and kept.

Data leaks and attacks on the cloud distributor may occur when the medical data is being published. The AFBS WOA algorithm, created by Thanga Revathi, et al. [21], combines AFBSO (Adaptive Fractional Brain Storm Optimization) with the Whale Optimization technique to address this issue [20, 21] (WOA). A new AFBS WOA algorithm generates the key matrices coefficients needed to retrieve a corrupted database and keep patient information private in the cloud. The secret key was calculated using a fitness function that incorporated utility and privacy considerations. A secure database can be built by multiplying the input database by a key matrix created by Tracy–Singh using the Tracy.

Kumar et al. [22] extremely difficult to constantly monitor the central storage of health records that are vulnerable to security risks. For this reason, in order to protect confidential patient information, this study uses a block chain technology and a digital signature with authentication to protect it, as well as a cloud-based model to ensure the information's authenticity and reliability. Traditional methods for preserving medical records were studied and compared to the model presented in the study, in terms of response time and the cost of storing and retrieving records.

Smys [23] new technologies, such as sensor networks and smart monitors, have altered this picture by leveraging mobile devices and internet services. This has improved practical healthcare through predictive modeling and the acquisition of more detailed individual measurements. A large amount of data allows researchers to analyse patterns [24] and trends in order to provide solutions that improve medical treatment while keeping costs down, while also ensuring that human lives are not put at risk. The survey on the accuracy and predictive power of big data analysis in the health care system is presented in this study.

3 Proposed System

An outline of a way for safely transferring healthcare data between cloud systems is provided here.

3.1 Architecture Overview

The following entities make up an efficient healthcare system based on EGC's overall architecture (Fig. 1):

HC: Users can store, update, and back up healthcare data using cloud services provided by the HC All cloud services are supported by the HC's server, which houses all of the healthcare data. The health cloud's data had to be protected from a variety of dangers. Encryption of data in the health cloud ensures the privacy of patient records.

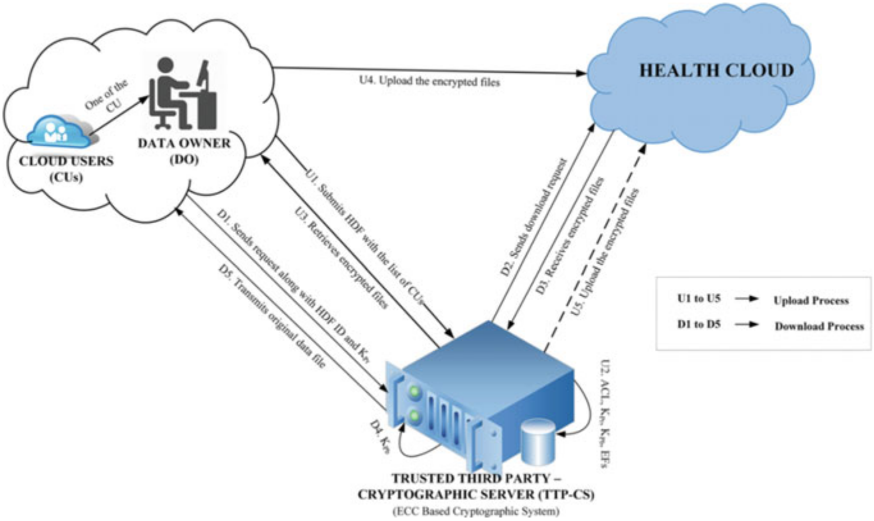

HEALTH CLOUD
One of the CU
CLOUD USERS (CUs)
DATA OWNER (DO)
U4. Upload the encrypted files
U1. Submits HDF with the list of CUs
U3. Retrieves encrypted files
D1. Sends request along with HDF ID and K_{Pr}
D5. Transmits original data file
D2. Sends download request
D3. Receives encrypted files
U5. Upload the encrypted files
U2. ACL, K_{Pr}, K_{Pu}, EFs
D4. K_{Pr}
U1 to U5 → Upload Process
D1 to D5 → Download Process
TRUSTED THIRD PARTY – CRYPTOGRAPHIC SERVER (TTP-CS)
(ECC Based Cryptographic System)

Fig. 1 Architecture of optimized EGC-based secure health cloud

TTP-CS: TTP-CS is the third-party-owned trusted entity that performs the cryptographic process outside of the cloud. The ECC algorithm has been modified in this mechanism's design to ensure the security of sensitive healthcare records. In order to ensure that healthcare data may be shared in a secure manner, it is responsible for data confidentiality, integrity, key management.

CUs: The health cloud's clients are the cloud's users (such as researchers, analysts, physicians, and others). Registration of CUs with the TTP-CS is required in order to carry out security services. Only one CU will own each data file, while all other CUs will be consumers of that information.

3.2 System Model

For the safe transfer of medical data files to and from the cloud, this architecture supports asymmetric or public key cryptography. The TTP-CS receives a list of CUs and a patient health information file from the DO. It is as a result of this that TTP-CS generates two random 256-bit keys, the public key (K Pb) and private key (K Pr). An asymmetric key algorithm can be made to run for a shorter or longer period of time using several techniques. K Pb and K Pr are generated by using the SHA-256 hash function on a random number RN. For the encryption and decryption of healthcare data, it is used further. After the encryption or decryption procedure, no one has access to the complete key. For each CU, TTP-CS generates a unique K Pb that can be freely shared and used in the encryption process. However, K Pr is only known to the decryption unit and is not shared with any other units. Security objectives can be achieved by implementing these cryptographic operations.

3.2.1 Loading a File to HC

The TTP-CS receives the request for encryption when a CU wants to upload healthcare data to the health cloud. Access privileges are mentioned in the HDF and the cloud user's list. Access to HDF may be Read-only or Read–Write depending on the CUL's permissions for each CU. The TTP-CS creates the Access Control List (ACL) for healthcare data using CUL. The data owner will inform the novel CUL to the TTP-CS while the HDFs are being shared with a new center. Otherwise, it only transmits the center ID of the last remaining center in the chain of transmission. ACLs for each HDF are created and maintained by the TTP-CS once the encryption requisition has been received. The ACL is made up of data about files, such as the file's ID, size, and owner ID, as well as metadata describing how the item was shared. The TTP-CS makes K Pr and K Pb for each CU after constructing the ACL. The HDF is then encrypted with an EGC-optimized encryption technique.

Elliptic Galois Cryptography

Because of its foundation in elliptic curve theory, ECC is usually referred to as the public key encryption method. Instead of using conventional methods, the keys are created by utilizing the features of elliptic curve equations. EGC is employed in the proposed project. Elliptic curves over Galois fields (Fa) are used to improve calculation efficiency and eliminate rounding mistakes. It is possible to determine the Galois field's value by utilizing the Mayfly algorithm's best answer for the ideal value.

Mayfly Algorithm

The Mayfly method is used in this work to maximize CNN's learning rate. To put it another way, Zervoudakis and his colleagues have presented a variation on PSO that incorporates the best of PSO, GA and FA. Because it has been demonstrated that PSO requires some modifications to ensure an optimal point when performing in high-dimensional spaces, researchers trying to improve the performance of the PSO algorithm using techniques like crossover and local search now have a powerful hybrid algorithmic structure based on the behavior of mayflies. A possible solution to the problem can be found by examining the mayfly's location in the search space. The following is the flowchart for the algorithm. In the beginning, two groups of mayflies, one for each sex, are randomly formed. Each mayfly is randomly placed in the problem space as a potential solution and its performance is evaluated using the predetermined objective function, f, which is represented by the vector x =(x 1,…,x d) (x). The velocity of a mayfly is defined as the change in its position, and the flying direction of each mayfly is a dynamic interaction of individual and communal flying experiences. It is also possible for each mayfly to modify their trajectory toward their personal best (pbest) and the best position obtained by any swarm mayfly to date (gbest).

(a) **Movement of male mayflies**
It follows that the location of each male mayfly in a swarm is determined by both its own knowledge and that of its neighbors, as the males gather in swarms.

(b) **Movement of female mayflies**
Female mayflies do not form swarms, unlike their male counterparts. For the purpose of mating, they prefer to fly toward males.

(c) **Mating of mayflies**
Using the operator, two mayflies' mating process is depicted as: One parent is chosen from the male, while the other is chosen from the female one. They attract each other in the same manner as parents attract their children. A random process or a fitness function can be used to make the selection. Likewise, the most beautiful woman is paired with the most attractive man. As a result of the mating, the following two children are born:

$$offspring1 = L * male + (1 - L) * female \tag{1}$$

$$offspring2 = L * female + (1 - L) * male \tag{2}$$

For example, L is a random variable that falls within a certain range for male and female parents. The starting velocities of offspring are set to 0 at the beginning of the game. The Mayfly Algorithm (MFA) can be stated in a pseudo code that shows the basic processes.

Algorithm 1: Pseudo Code of MFA

```
$Objective\ function\ f(x), x = (x1, \ldots, xd)^T$
$Initialize\ the\ male\ mayfly\ population\ x_i (i = 1, 2, \ldots, N)\ and\ velocities\ v_{mi}$
$Initialize\ the\ female\ mayfly\ population\ y_i (i = 1, 2, \ldots, M)\ and\ velocities\ v_{fi}$
$Evaluate\ solutions$
$Find\ global\ best\ gbest$
**Do While** $stopping\ criteria\ are\ not\ met$
    $Update\ velocities\ and\ solutions\ of\ males\ and\ females$
    $Evaluate\ solutions$
    $Rank\ the\ mayflies$
    $Mate\ the\ mayflies$
    $Evaluate\ the\ offspring$
    $Separate\ offspring\ to\ male\ and\ female\ randomly$
    $Replace\ worst\ solutions\ with\ the\ best\ new\ ones$
    $Update\ pbest\ and\ gbest$
**End while**
$Post - process\ results\ and\ visualization$
```

The encrypted files as E f1 and f2 are the end result of this process. In this case, E f1 is the product of a random number k and a point on the elliptic curve Pt c that is randomly selected. HDF, k, and the public key, K Pb, are added together to form E f2. KPb is included into each CU's ACL for the next step in the procedure. The integrity of each encrypted file is safeguarded by the HMAC signature and key generated and stored by the TTP-CS. This information is sent to the person who requested it: the center ID, the encrypted files (E f1 and E f2), and their K Pr. Whereas, just the center ID and K Pr are transmitted across a Secure Socket Layer to the rest of the CUs (SSL). After the encryption procedure, a secure overwrite separates K Pr and K Pb from the TTP-CS. It is up to DO or TTP-CS to upload the encrypted files (E f1 and E f2) after they have been received (on behalf of CU).

The key generation procedure begins as soon as the encryption center is activated or the encrypted file is submitted. You have two options when it comes to uploading files: In either case, the DO can be promptly posted to the HC, as previously explained, or the TTP-CS can upload the file on behalf of the CU to the HC, which has the authority delegation. It is possible to upload a single file of medical data to the Health Center by following the steps listed below:

U1	The TTP-CS receives the healthcare data file from the physician and the list of users
U2	The ACL, private key, and public key for the physician are generated by TTP-CS. Next, an efficient EGC encryption mechanism is used to protect the data
U3	Physicians can get their TTP-CS private key, centre ID, and encrypted files by using the TTP-CS
U4	The physician upload the encrypted files straight to the health cloud
U5	TTP-CS uploads encrypted files to the health on behalf of the doctor in special cases

3.2.2 Downloading a File from the HC

It is either essential for the TTP-CS to receive an authentication request from the CU or for the DO to download encrypted files directly from the HC and then submit a decryption request. A locally upheld ACL authenticates the CU's authorization from the HC. The TTPCS obtains K Pb from the ACL. The requesting CU will receive an access forbidden message if the ACL does not contain the K Pb. Because each CU has its own K Pb, no CU can use the K Pr of another CU. As a result, the decryption procedure can begin after the TTP-CS verifies the file's integrity. Depending on whether or not the TTP-CS receives a valid K Pr, the decryption operation will either succeed or fail.

After a successful ECC decryption, the HDF is sent to the relevant CU over an SSL connection. The secure overwriting approach eventually removes K Pr and K Pb from the TTP-CS. The TTP-CS can also be used to download files on behalf of the CU, same like the file uploading process. TTP-CS receives this request for decryption along with login credentials, as previously stated. Once the TTP-CS has confirmed that the CU for the specified file is genuine, it will forward this request on to the HC for processing. Further transmission of encrypted data takes place via TTP-CS, with the HC acting as a conduit. The rest of the process is the same as described previously. Here is an example of how to obtain a medical data file from the HC:

D1	The TTP-CS receives requests from the CU
D2	In order for TTP-CS to send a download request to the health cloud, ACL verification is required
D3	The health cloud sends encrypted files to TTP-CS
D4	TTP-CS retrieves from the ACL
D5	It is sent to the appropriate CU with the original data file in it

3.2.3 File Restore

ACL and key generation are not performed while recovering a file, unlike when uploading a file. TTP-CS receives a restore request from CUs (who have already downloaded the file) if any modifications have been made. True or false, TTP-CS verifies whether or whether CU has WRITE access to a file. The TTP-CS computes the keys if a valid request for file restoration has been made. Additionally, the file is encrypted before being subjected to the HMAC algorithm. They're either re-encrypted and transmitted through email or uploaded to the HCI server. Finally, the K Pr and K Pb are left out of the equation.

The proposed model delivers the subsequent features to the healthcare data: Healthcare data must be protected from insider threats by preventing unauthorized access within the center. Secure sharing of healthcare data among the center.

3.2.4 Security Analysis

1. **Eaves dropping**
   The patient receives the private key from the certificate authority via a secure connection. As a result, hackers will be unable to access the encrypted data.
2. **Replay attack**
   The property and the secret key used to encrypt the files can be found in the tree structure. The EGC cryptography algorithms have been implemented by the doctors to their fullest potential. The optimization mechanism identifies the EGC values in order to identify the best solutions. As a result, a replay assault on the keyword provided by the patient is ruled out. Even if the hacker knows the characteristics and ciphertext, the secret key is not fixed in the EGC, therefore he can't calculate it.
3. **Masqurade and man in the middle attack (MIM)**
   Hackers can't use a masqurade attack because the properties are utilized to encrypt the files. The properties of the file must be known by the hacker if he wants to hack it. The hacker must know the property before he or she can change ciphertext files. Files are transmitted via cryptography rather than the MIM since we've utilized an EGC that operates on points instead than bytes.

4 Results and Discussion

Using an Intel Core i5-6200U CPU clocked at 2.40 GHz and 8.00 GB of RAM, the proposed solution is put into practice on a Windows 10 64-bit OS system. The HC, TTP-CS, and CUs are the three main components described in the system model. Uses JPBC v.2.0.0 Java Pairing Based Cryptography library for communication between entities. Both elliptic curve and pairing procedures can be implemented with the

Table 1 Computation time for finding the value of EGC with different iterations

No. of iterations	WOA	*AFBS_WOA* [21]	Mayfly
5	1.594	1.534	1.494
10	1.741	1.606	1.598
15	1.888	1.798	**1.645**
20	2.193	2.110	**2.001**
25	**2.356**	**2.190**	2.129

Table 2 Key generation time

FS (MB)	Methodologies (time in second)			
	AES	ECC	EGC	Proposed optimized EGC
10	1.594	1.534	0.004	0.00212
20	1.741	1.606	0.00425	0.00235
30	2.321	1.684	0.00476	0.00286
40	1.888	1.799	0.005	0.00302
50	1.952	1.866	0.00512	0.00328
60	2.193	1.923	0.0055	0.0035
70	2.286	2.034	0.00598	0.00398
80	2.694	2.129	0.00632	0.00427
90	2.827	2.388	0.00664	0.00463
100	2.887	2.545	0.00697	0.00499

help of its functions. It is possible to communicate between the entities thanks to the Java libraries. SSL encrypts all data sent and received. It was tested using the Cloudsim toolkit and evaluated in terms of key generation time, file upload and download times, and the time it took to discover EGC's value. Performance analyses of proposed mayfly algorithm are tabulated on Tables 1, 2, 3 and 4.

4.1 Performance Analysis of Proposed Mayfly Algorithm

See (Table 1).

4.2 Performance Analysis of Proposed Optimized EGC

See (Tables 2, 3 and 4).

Table 3 Time taken for uploading the encrypted files and downloading the decrypted files

FS (MB)	Methodologies (time in second)							
	AES		ECC		EGC		Proposed optimized EGC	
	UL	DL	UL	DL	UL	DL	UL	DL
0.1	1.4	0.99	1.48	1.15	0.80	0.80	0.70	0.70
0.5	1.48	1.03	1.89	1.31	0.94	0.96	0.80	0.82
1	2.06	1.48	2.90	1.85	1.24	1.18	1.20	1.24
10	14.95	9.90	14.59	10.45	6.43	6.48	5.60	5.68
50	58.56	35.57	60.37	35.90	9.01	10.24	8.25	8.78
100	112.41	59.14	115.15	61.59	17.39	20.68	16.35	18.98
500	492.03	229.81	872.09	400.21	33.24	39.25	31.10	38.22

Table 4 Speed of file uploading

File size (MB)	Uploading speed (Mb/s)
0.1	11.5
0.5	12
1	11.9
10	12.92
50	12.5
100	13
250	13
500	13

5 Conclusion

Increasing e-health productivity is now possible because to cloud computing-based health clouds, which allow medical professionals to access patient records from anywhere at any time, on any device. Secure data exchange between general practitioners, medical providers, and insurance companies is a critical concern for any healthcare company. In order to deal with this problem, encryption technologies are used to safeguard critical healthcare data. EGC-based encryption is utilized for data security in the proposed health cloud architecture. TTP-CS is also responsible for the encryption and decryption operations. Using the EGC model surpasses other existing systems in terms of key generation time, file upload time, file download time and uploading speed. EGC has a smaller key size, which makes key administration more simpler. The results demonstrate that EGC-based approach is a promising choice for safe healthcare data sharing in the health cloud. Because it hasn't been developed to handle image-based data yet, this encryption approach can only be used to protect plaintext. It is possible that this problem will be resolved in the future. Also implement the security mechanism in the cloud that are federated and then compare its efficiency with the existing methods.

References

1. Jayaraman I, Stanislaus Panneerselvam A (2021) A novel privacy preserving digital forensic readiness provable data possession technique for health care data in cloud. J Ambient Intell Humaniz Comput 12(5):4911–4924
2. Manne R, Kantheti SC (2021) Application of artificial intelligence in healthcare: chances and challenges. Curr J Appl Sci Technol 40(6):78–89
3. Qiu H, Qiu M, Liu M, Memmi G (2020) Secure health data sharing for medical cyber-physical systems for the healthcare 4.0. IEEE J Biomed Health Inform 24(9):2499–2505
4. Rushanan M, Rubin AD, Kune DF, Swanson CM (2014) Sok: security and privacy in implantable medical devices and body area networks. In: 2014 IEEE symposium on security and privacy. IEEE, pp 524–539
5. Sun Y, Lo FPW, Lo B (2019) Security and privacy for the internet of medical things enabled healthcare systems: a survey. IEEE Access 7(2019):183339–183355
6. Timothy DP, Santra AK (2017) A hybrid cryptography algorithm for cloud computing security. In: 2017 International conference on microelectronic devices, circuits and systems (ICMDCS). IEEE, pp 1–5
7. Banos O, Villalonga C, Damas M, Gloesekoetter P, Pomares H, Rojas I (2014) Physiodroid: combining wearable health sensors and mobile devices for a ubiquitous, continuous, and personal monitoring. Sci World J
8. Abdullah A, Ismael A, Rashid A, Abou-ElNour A, Tarique M (2015) Real time wireless health monitoring application using mobile devices. Int J of Comput Netw Commun (IJCNC) 7(3):13–30
9. Anuradha M, Jayasankar T, Prakash NB, Mohamed Yacin Sikkandar, Hemalakshmi GR, Bharatiraja C, Sagai Francis Britto A (2021) IoT enabled cancer prediction system to enhance the authentication and security using cloud computing. Microprocess Microsyst 80(2021):103301
10. Jaiswal K, Sobhanayak S, Turuk AK, Bibhudatta SL, Mohanta BK, Jena D (2018) An IoT-cloud based smart healthcare monitoring system using container based virtual environment in edge device. In: 2018 international conference on emerging trends and innovations in engineering and technological research (ICETIETR). IEEE, pp 1–7
11. Nirabi A, Hameed SA (2018) Mobile cloud computing for emergency healthcare model: Framework. Proc Int Conf Comput Commun Eng:375–379
12. Kumar V, Bhardwaj A (2020) Deploying cloud-based healthcare services: a holistic approach. Int J Serv Sci Manag Eng Technol (IJSSMET) 11(4):87–100
13. Blakley GR, Borosh I (1979) Rivest-Shamir-Adleman public key cryptosystems do not always conceal messages. Comput Math Appl 5(3):169–178
14. Gola KK, Gupta B, Iqbal Z (2014) Modified RSA digital signature scheme for data confidentiality. Int J Comput Appl 106(13)
15. Ali A, Pasha MF, Ali J, Fang OH, Masud M, Jurcut AD, Alzain MA (2022) Deep learning based homomorphic secure search-able encryption for keyword search in blockchain healthcare system: a novel approach to cryptography. Sensors 22(2):528
16. Zhang Y, Zheng D, Deng RH (2018) Security and privacy in smart health: efficient policy-hiding attribute-based access control. IEEE Internet Things J 5(3):2130–2145
17. Huang Q, Yue W, He Y, Yang Y (2018) Secure identity-based data sharing and profile matching for mobile healthcare social networks in cloud computing. IEEE Access 6:36584–36594

18. Bao S-D, Chen M, Yang G-Z (2017) A method of signal scrambling to secure data storage for healthcare applications. IEEE J Biomed Health Inform 21(6):1487–1494
19. Masood I, Wang Y, Daud A, Aljohani NR, Dawood H (2018) Towards smart healthcare: patient data privacy and security in sensor-cloud infrastructure. Wireless Commun Mobile Comput 2018(Nov 2018), Art. no. 2143897
20. Mbonihankuye S, Nkunzimana A, Ndagijimana A (2019) Healthcare data security technology: HIPAA compliance. Wireless Commun Mobile Comput 2019(Oct 2019), Art. no. 1927495
21. Thanga Revathi S, Gayathri A, Kalaivani J, Christo MS, Pelusi D, Azees M (2021) Cloud-assisted privacy-preserving method for healthcare using adaptive fractional brain storm integrated whale optimization algorithm. Secur Commun Netw
22. Kumar D, Smys S (2020) Enhancing security mechanisms for healthcare informatics using ubiquitous cloud. J Ubiquitous Comput Commun Technol 2(1):19–28
23. Smys S (2019) Survey on accuracy of predictive big data analytics in healthcare. J Inf Technol 1(02):77–86
24. Zervoudakis K, Tsafarakis S (2020) A mayfly optimization algorithm. Comput Ind Eng 145:106559

A Review of Mobile Computation Offloading Techniques

M. Jyothirmai, Kesavan Gopal, and M. Sailaja

Abstract More and more people are increasingly using multimedia on their mobile devices, such as smartphones, tablet PCs and smart watches as a result of technological improvements. The most significant elements of a mobile device are the battery life, memory, bandwidth, and CPU performance. When such computationally extensive tasks are performed on a mobile device, the battery quickly drains. However, offloading such tasks to a proxy and executing those results in significant power savings in mobile devices. We compare the ways of offloading to a proxy from a mobile device in this study based on power usage, energy, and execution time. A thorough examination of the offloading process is also presented. The findings show a significant reduction in the amount of energy consumed by mobile devices.

Keywords Mobile devices · Computation offloading · Task partitioning

1 Introduction

Because of the growing popularity of using mobile devices e.g., smartphones, smart watches, and electronic tablets in people's daily lives, demand for multimedia services (e.g., voice over IP (VoIP), web-browsing, video-watching, and file-downloading) is skyrocketing, resulting in an explosion in mobile traffic [1]. Multimedia apps that need a lot of resources, such as 3D video games, are becoming increasingly popular on mobile phones.

M. Jyothirmai (✉) · M. Sailaja
JNTUK, Kakinada, Andhra Pradesh, India
e-mail: jyothimunjamp@gmail.com

K. Gopal
LPU, Phagwara, Punjab, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_39

Mobile devices will continue to develop even as their hardware and mobile networks evolve and improve. Always be resource-scarce, insecure, with erratic connectivity, and limited energy. Mobile applications are growing increasingly popular, and there has been a large increase in mobile subscriptions, indicating that mobile cloud applications and services are in high demand among mobile device users.

However, because mobile terminals have limited resources, executing these applications directly on them does not meet the user's expectations. As a result, computation is required on mobile devices and the performance and functionality are currently limited.

To summarize, smartphones offer multi-core CPUs, better screen resolution, more memory, additional sensors, and radios, as well as a vast array of apps. These factors combined place a significant strain on the battery's energy consumption [2].

Much research has been done to save energy to increase the life of batteries [2–5]. Computation offloading is a popular energy-saving strategy for mobile devices, in which applications make use of resource-rich infrastructures by transferring computation to these infrastructures. It was also discovered that transferring processing to the cloud via networks can help lower smartphone power usage.

Mobile data offloading is a viable solution for addressing the aforementioned issues by leveraging complementary and novel networking approaches to transmit mobile data that was originally intended for cellular networks. Mobile data offloading is projected to play a significant role in mobile networks in the future, given the continued and rapid growth of mobile traffic. Offloading may help mobile devices conserve energy and function better. This, however, is typically dependent on several factors, including network bandwidths and data transfer rates.

Various infrastructures and techniques have been proposed to improve offloading. They address issues such as user transparency, privacy, security, mobility, energy efficiency, and so on. The goal of this work is to introduce readers to compute offloading research for mobile devices. The motives, strategies, technology enablers, and designs for compute offloading are discussed in this study.

This document addresses existing limitations and research directions for algorithmic mechanisms and associated infrastructures, as well as serves as a collective reference for them. The following is how the paper is structured: In Sect. 2, a brief history of enabling technologies is described. Section 3 outlines infrastructures and tools designed to solve the issues of offloading, as well as two objectives for offloading: lower execution time and conserve energy. Section 4 explains why offloading will become more significant in the future, and Sect. 5 gives the conclusion.

2 Mobile Task Offloading

In this section, we first cover the basics of mobile cloud off-loading systems before going into the primary obstacles and issues that come with making off-loading decisions.

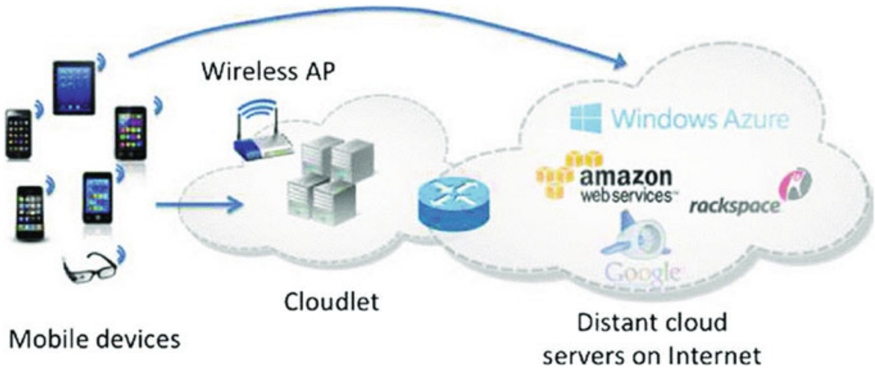


Fig. 1 Architecture of the offloading process

Figure 1 depicts an off-loading system that uses a proxy server to deliver multimedia material to a portable device. This system is made up of three components: content servers, proxy servers, and mobile devices [6]. A mobile or handheld device is any networked resource, such as a handheld (personal digital assistant or PDA), a gaming device, or a wireless security camera. Material servers keep track of multimedia and database content and respond to queries by delivering data (such as images) to clients.

All communication between mobile devices and servers is relayed through proxy servers. Proxy servers are powerful servers that, among other things, can compress and decompress images, transcode video in real time, access/provide directory services, and deliver services based on a rule base. As a result, mobile devices negotiate security, service quality, and content delivery with proxy servers.

In turn, the proxy servers make requests the image/video/data stream to content servers according to the needs of the user It's worth noting that mobile devices can also be used to produce and deliver data to other mobile devices via proxy servers connected to the network.

The abbreviations used are shown in Table 1.

Table1 Abbreviations

PDA	Personal Digital Assistant
MEC	Multi-access Edge Computing
SPEA	Strength Pareto Evolutionary Algorithm
EFFORT	Energy Efficient Framework for Offload Communication
MCC	Mobile Cloud Computing
MCM	Mobile Cost Monitor
WLAN	Wireless LAN

3 Major Criteria in Offloading

Metrics: Most offloading decisions are based on a certain criterion. On one hand, energy, cost, and storage are criteria that should be minimized, while performance, robustness, and security are criteria that should be maximized [7].

Application Partitioning: Using the collected data, the offloading decision-making module makes a decision based on the metrics module (i.e., minimizing or maximizing some criteria), and then the partitioning module is called to divide the classes that make up an application into local and remote partitions, with the former running locally on the mobile device and the latter offloaded to a dedicated cloud server [8].

Load balancing: At any time, the edge node or datacenter may be overwhelmed, causing delays and, in some cases, inaccurate results [9].

Scalability: Real-time applications are frequently used by multiple users at the same time, which is managed by separate algorithms that are responsible for a large number of user requests. The offloading process, on the other hand, must be optimized in order for the application to scale up without causing harm [10].

Energy consumption trade-off: Because the offloading procedure consumes energy and bandwidth, the offloading decision must be made in light of this trade-off [10].

Availability: Mobile devices must always connect to the edge/cloud, which is difficult owing to network coverage gaps, congestion, limited bandwidth, and other network-related issues [10].

Security: Tasks and users' data are transported via the network throughout the offloading process, increasing the risk of data theft and misuse. To address this problem, it is important to enlist the help of a third party [11].

Decision-making: Determining whether or not to offload a work is difficult due to a variety of factors such as delay, energy, and payment expenses [12].

Flexibility: When connecting to the edge, this poses a big issue [13].

Resource utilization: When compared to cloud infrastructure, the quantity of resources on the edge is limited. As a result, resource utilization management is essential to get the most out of limited resources while avoiding system overhead [14].

4 Classification

Offloading can be divided into two categories: data offloading and computation offloading [15]. Data offloading is a technique for transferring data from a mobile device with limited storage and capability to a cloud repository. Data offloading from multi-mobile devices to multi-MEC servers was proposed by Zhang [16].

He proposed a game-centric pricing structure utilizing MEC to prioritize data offloading based on mobile device and resource allocation. Xu [17] proposed a time-efficient offloading strategy for intelligent sensors in edge computing, while maintaining privacy. An improved version of the Strength Pareto Evolutionary Algorithm (SPEA2) is utilized to jointly optimize average time consumption and average privacy entropy. Sun [18] constructed an optimization problem with weighting parameters in order to maximize the total efficiency of user computation. He also employed the iterative and gradient descent methods to demonstrate that the suggested strategy outperforms established methods.

Computation offloading is a technique for processing intensive applications remotely in order to take advantage of strong resources in cloud servers, bypassing the CPU and battery limits on mobile devices [19]. Applications can run with reduced latency on mobile user equipment thanks to partial compute offloading [20].

Full and partial offloading are the most common offloading modes. The entire task is offloaded without partitioning in full offloading mode [21–24]. In partial offloading mode, the task is partitioned. As a result, some parts are processed locally while others are done remotely [20, 25–28].

By offloading compute-intensive operations to proxy servers, computation offloading is an effective technique to ensure user service quality [29]. The primary goal of computation is to shorten the service's response time, increase the quality of the service Furthermore, when the mobile device doesn't have the processing power, the computation can be moved to a proxy server or a cloud data center in order to increase the system's overall performance. To do so many components of the computation offloading decision must be taken into account, such as maximizing performance and minimizing energy use.

5 Offloading Decision

Before computation offloading, a number of questions must be answered.

(1) When to offload? The task scheduler must be able to determine the offloading time slot for various scenarios and constraints.

Because of the necessity for additional data transfers, which might increase time and/or energy consumption when task-related data is exchanged, remote execution on a cloud server is not always an advantageous option [8]. When the time and energy saved from offloading is insufficient to cover the additional communication expenses between the mobile device and the cloud, we can choose to run the app locally rather than offload it to the cloud. As a result, mobile cloud offloading is a viable option, but it is not required. We must determine the best moment to offload, such as when the wireless network is available, the amount of transmission data is low, or the amount of processing is high.

Rahmati and Zhong [30] proposed a computation offloading method that allows smart mobile devices to dynamically adjust computational speeds based on calculation demands, lowering energy consumption and computation time. They improve the processing speed, transmission power, and offload rate on smart mobile devices to reduce energy consumption and application execution delays. For single-server and multi-server scenarios, they created a new compute offloading approach.

There is a basic trade-off between mean energy usage and mean response time for diverse applications [31]. Offloading, in which cloud services are available with minimal network latencies, can provide better service by reducing response time. Miettinen and Nurminen [32] proposed offloading operation by switching between multiple transmission technologies, as well as examining the trade-off between energy usage for Wi-Fi search and transmission efficiency when the Wi-Fi network was accessible. Energy efficient delayed network selection has been used to optimize the trade-off between energy usage and data transmission delay by postponing data transmission until the device finds an energy-efficient network [33].

Some research was conducted to determine whether it is preferable to offload computational processes to a dedicated server rather than conducting them on the mobile device [34]. Mobile cloud offloading has the potential to reduce mobile device execution time and energy consumption, but the savings must outweigh the higher communication costs between the device and the cloud. [35]. Because the energy overhead of data transport may outweigh the energy savings from lower CPU usage, a performance seeking offload cannot guarantee energy savings [36].

When a wireless network is available, regardless of network quality, all traffic is promptly offloaded to the remote cloud [37]. When there isn't a high-quality network available right now, the offloading procedure might be postponed until a suitable network becomes available [38]. We can opt not to offload the work to the cloud when the link is bad or the amount of data is huge, rather than transferring it directly to the cloud.

(2) What to offload? If a task can be offloaded, the task scheduler must be able to determine whether it may be offloaded and, if so, what should be offloaded. Is it better to discharge in stages or all at once?

Cloud data center workloads must be offloaded to edge devices and edge servers for edge computing to reduce service latency and network bandwidth utilization. As a result, workloads in cloud data centers must be partitioned, with some of them destined for edge devices and servers [39]. Cuervo et al. [40] recommended offloading and caching on the edge server instead of transferring the centralized database to the client to enhance application latency. Furthermore, filtering a high number of server requests through web proxies can reduce server workload dramatically.

To accomplish a specific performance goal, such as the shortest reaction time or the lowest energy consumption, it should be established which portions of the task should be stored on the cloud server and which should be kept on the mobile device [39]. Chen et al. [40] proposed a network caching technique that uses the storage capacity of multiple network devices to reduce network traffic.

Some operations should always be performed locally on the mobile device, either because transferring relevant data would take too long and consume too much energy, or because these functions require access to local components (e.g., camera, GPS, user interface, accelerometer, or other sensors) [41]. To alleviate the strain of rapidly rising demand for caching and computing services, Lin et al. [42] proposed a new information-centric heterogeneous network design for content caching and computation. Malik [43] proposed that data-intensive edge computing applications be updated in the main database, eliminating the need for a copy to be sent to the server. They also provided a wide-area replication mechanism for delivering dynamic material while utilizing the capabilities of edge computing.

(3) What is the optimum site to outsource workload execution and where should it be done?

It's critical to locate the greatest cloud service for offloading, i.e., the ideal spot to unload. By distributing partitioned tasks to mobile devices and proxy servers, task offloading can be done. The selection of targeted edge devices and edge servers involves the optimization of several objectives such as performance, energy, network bandwidth, and data privacy protection.

To save energy, for example, energy-intensive activities are offloaded to cloud servers, and data-intensive jobs are offloaded to edge servers to reduce latency and network traffic [39]. In [44], an Energy efficient framework for offload communication (EFFORT) in MCC was used for communication offloading, resulting in a significant reduction in energy consumption.

Flores and Srirama [45] devised a decision tree-based method for creating an offloading strategy in which all computation-intensive offloading decision operations are sent to a distant server for execution.

An application's components can be deployed on numerous application processing nodes, such as a mobile device, cloudlet, or cloud, implying that many offloading destinations and targets are possible [46]. In mobile cloud computing systems, Flores and Srirama [45] introduced a context-aware offloading method and a computation offloading approach. To estimate task execution costs, it supplied a general cost estimates model for cloud resources, including execution time and energy usage.

A Mobile Cloud Middleware (MCM) was designed as a bridge between the mobile device and the cloud in [8], with the purpose of handling asynchronous delegation of mobile tasks to cloud resources and minimizing the time it takes to transfer duties from mobile devices to the cloud. In [47], they developed an adaptive offloading technique based on Lyapunov optimization that decides where each application activity should be performed so that energy consumption is minimized while latency is kept to a minimum.

Effective offloading selections are made after assessing where offloading will increase system performance or generate the highest gains.

(4) How to offload?

> Using one or more of the available wireless networks, mobile cloud offloading moves heavy computing from mobile devices to powerful cloud servers. Offloading work to a dedicated resource can be accomplished in a variety of ways, including using a cellular connection or a sporadic WLAN hotspot [48].

In most study papers [49, 50] it is assumed that cellular networks are always available to mobile users, whereas WiFi network availability varies by location to make decision-making easier. On their mobile devices, users move in and out of a WiFi coverage area.

Chen et al. [51] suggested an online task offloading strategy to reduce the time it takes for mobile apps to run. They observed that load balancing methods might be used to dump jobs onto the cloud, maximizing parallelism between mobile and cloud for parallel tasks.

Khan et al. [52] investigated the design of mobile edge computing computation offloading mechanisms in 5G heterogeneous networks and proposed the EECO framework as a multi-device energy-saving computation offloading framework to reduce computing task energy consumption during computation offloading and execution.

Offloading decisions can be made in a fixed, static way, or offloading can be done dynamically in response to the application's activity [58]. A good offloading choice is made by calculating the optimal time to offload under various device parameters, such as available bandwidth, amount of data to transfer, and energy consumption, and then picking the appropriate component to offload by separating a specific application into local and remote parts and determining the best place to offload under various cloud resource conditions.

6 Computation Offloading

Resource-intensive calculations can be offloaded by shifting them to a separate processor such as a cluster, grid, or cloud. Offloading to a coprocessor can speed up a variety of tasks, including image watermarking, image compression, and mathematical computations. By offloading processing to an external platform across a network, it is feasible to get around a device's physical limitations, such as its constrained computational power, storage space, and energy. Computation offloading model is shown in Fig. 2.

Table 2 summarizes the previous work done in computation offloading. We present the contribution of the work done by using different partitioning algorithms with the goal of prolonging smartphone battery life or improving energy efficiency.

The volume of mobile data is expected to expand fast, as seen in Fig. 3. Large volumes of multimedia data are generated by mobile platforms, and the majority of this material is kept online on cloud servers. When millions of cameras, microphones, GPS, and a variety of other sensors are connected, the amount of data generated is

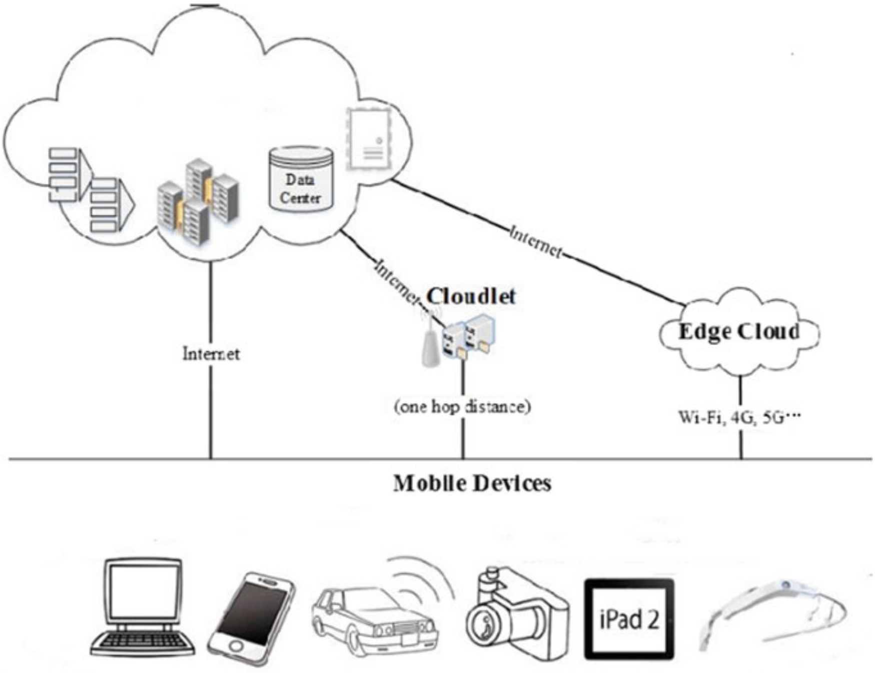


Fig. 2 Computation offloading model

enormous. As the number of connected devices such as smartphones, tablets, laptops and sensors increases, so will the demand for more features.

These factors suggest that mobile computing rates will not keep pace with the growth of data and the processing demands of applications. On one hand, there has been a significant increase in the types and amounts of mobile data, as well as the computational requirements of mobile applications.

On the other hand, the computing capabilities of the devices, those that receive and store data and deliver user applications—are unlikely to keep up. Computation offloading is a natural answer to this issue.

7 Offloading Techniques

The major issue with smartphones that needs to be fixed is their high battery consumption. By shifting mobile applications to the cloud and solving the issues there, the

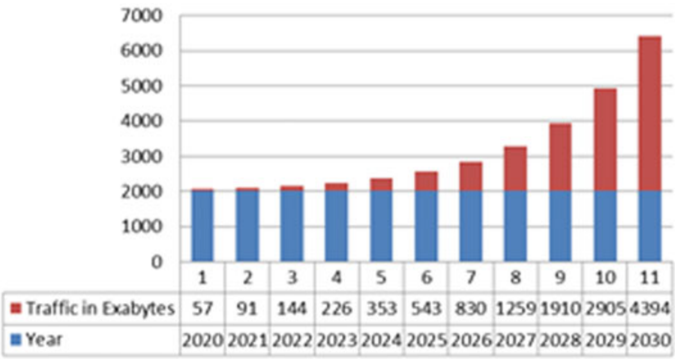
Table 2 Literature review

Year	Authors	Contribution
2020	Elgendy [21]	For mobile-edge computing, an efficient and secure multi-user multi-task compute offloading model with guaranteed performance in latency, energy, and security
2019	Nguyen [[9]	Maximize computation offloading and resource allocation to reduce the weighted sum of all mobile users' energy consumption
2019	Sun [18]	A combined computation approach that combines two schemes, local computing and data offloading
2019	Kuang [25]	For MEC systems with several separate jobs, there is a dual problem of partial offloading scheduling and resource allocation
2018	Ning [20]	An Iterative Heuristic MEC Resource Allocation (IHRA) algorithm is used to make a dynamic offloading decision
2018	Akherfi [13]	Highlights current offloading frameworks and computation offloading strategies, as well as their most important challenges, and assesses them
2015	Tao and Yaling [3]	They attempt to determine which components of a mobile application should be offloaded to the cloud for a given set of computational components
2015	Wu [8]	Offloading computation-intensive elements of mobile apps to a capable cloud server is an effective technique to reduce a resource-constrained mobile device's struggle with resource-hungry mobile apps, hence improving the device's performance
2014	Xia and Feng [2]	Reduces the execution time of an application running on a smartphone by offloading computing to the cloud. This improves the energy economy of smartphones and improves the application's performance
2014	Zhang [5]	Offloading service that allows a mobile application to easily offload some of its work from mobile terminals to the cloud
2013	Wu [7]	Based on the tradeoff between reducing execution time and extending the battery life of mobile devices, a novel adaptive offloading strategy is presented and examined
2012	Kovachev [1]	Allows for the expansion of android application execution from a mobile client to the cloud in an adaptable manner
2006	Kejriwal [6]	A method of partitioning the watermarking, embedding, and extraction algorithms and offloading some tasks to a proxy server

energy consumption issue could be solved. This section explains some computational offloading models.

MCC

Mobile cloud computing technique uses a cloud-based copy of the communication method. A message is delivered from the program to the cloud-based clone, which then sent it to the service. That service has two ends, one of which is on the cloud and the other on a mobile device. The cloud service pulls the data from the cloud copy and then sends it to the smartphone. The cloud-based service recognizes the



	1	2	3	4	5	6	7	8	9	10	11
🟥 Traffic in Exabytes	57	91	144	226	353	543	830	1259	1910	2905	4394
🟦 Year	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030

Fig. 3 Growth in mobile traffic

communication with the mobile device. The mobile device selects the appropriate application and sends the data to it. As a result, there will be only one service running on mobile devices rather than a service for each application.

Clone Cloud

The Clone Cloud framework was developed with the goal of extending the battery life and enhancing performance on mobile devices by offloading resource-intensive components to cloud servers.

In this approach, the partitioning step combines static program analysis with program profiling to create a set of off loadable components while adhering to certain requirements, such the requirement that methods that employ mobile sensors should be run locally.

MEC

In mobile edge computing, compute-intensive operations are offloaded either so that mobile devices can use less power or because they cannot be completed in time due to hardware limitations. MEC servers allow for the consideration of novel applications as possibilities for offloading, resulting in bandwidth savings and scalability.

MAUI

In order to calculate the energy consumption during code execution based on the number of CPU cycles needed to run it. In MAUI the mobile device is profiled and a straightforward linear model is created.

8 Conclusion

We review the existing work in the subject of mobile compute offloading in this paper. We also go through the current frameworks for computation offloading, as well as the many strategies utilized to improve the capabilities of smartphone devices using cloud resources.

We've noticed that current offloading frameworks are still dealing with various issues. We feel that pursuing other options, such as adopting a middleware-based design with an optimal offloading mechanism, could improve the frameworks already available and deliver more efficient and adaptable solutions.

Offloading strategies that take advantage of cloud computing will become increasingly relevant as various linked devices become more widely deployed. Applications on these linked devices will begin to be created with offloadable computing in mind, and this type of application design will benefit from the methodologies and solutions discussed in this paper.

Hence we conclude that computation offloading can be used to reduce energy consumption in mobile devices. However, this technique has its own limitations. In order to achieve maximum efficiency, the device should be able to perform certain tasks at low power levels. This means that the device needs to be able to run at lower clock speeds than normal. Another limitation is that the device cannot use multiple cores simultaneously.

References

1. Kovachev D, Yu T, Klamma R (2012) Adaptive computation offloading from mobile devices into the cloud. In: 2012 IEEE 10th international symposium on parallel and distributed processing with applications. IEEE
2. Xia F et al. (2014) Phone2Cloud: exploiting computation offloading for energy saving on smartphones in mobile cloud computing. Inf Syst Front 16(1):95–111. Elissa K, Title of paper if known, unpublished
3. Tao Y, Zhang Y, Ji Y (2015) Efficient computation offloading strategy for mobile cloud computing. Proc IEEE Int Conf Adv Inf Net Appl (AINA 2015)
4. Elgazzar K, Martin P, Hassanein HS (1989) Cloud-assisted computation offloading to support mobile services. IEEE Trans Cloud Comput 4(3):279–292. Young M (1989) The technical writer's handbook. University Science, Mill Valley, CA
5. Zhang Z, Lim H, Lee HJ (2014) An efficient framework for computation offloading in mobile cloud computing. Int Conf Future Inf Commun Eng 6(1)
6. Kejariwal A et al. Energy efficient watermarking on mobile devices using proxy-based partitioning. IEEE Trans Very Large Scale Integr (VLSI) Syst 14(6):625–636
7. Wu H, Wang Q, Wolter K (2013) Tradeoff between performance improvement and energy saving in mobile cloud offloading systems. Proc Int Conf Commun Workshops (ICC):728732
8. Wu H (2015) Analysis of offloading decision making in mobile cloud computing. Ph.D. dissertation, Department of FB Mathematik und Informatik, Freie Universitaet, Berlin, Germany
9. Nguyen PD, Ha VN, Le LB (2019) Computation offloading and resource allocation for backhaul limited cooperative MEC systems. In: Proceedings of the 90th vehicular technology conference (VTC2019-Fall), pp 1–6
10. Aazam M, Zeadally S, Harras KA (2018) Offloading in fog computing for IoT: review, enabling technologies, and research opportunities. Future Gener Comput Syst 87:278–289
11. Wang F, Diao B, Sun T, Xu Y (2020) Data security and privacy challenges of computing offloading in FINs. IEEE Network 34(2):14–20

12. Behera SR, Panigrahi N, Bhoi S, Sahani A, Mohanty J., Sahoo D, Maharana A, Kanta LP, Mishra P (2020) A novel decision making strategy for computation offloading in mobile edge computing. In Proceedings of 2020 international conference on computer science, engineering and applications (ICCSEA), pp 1–5
13. Akherfi K, Gerndt M, Harroud H (2018) Mobile cloud computing for computation offloading: Issues and challenges. Appl Comput Informat 14(1):1–16
14. Peng K, Zhao B, Xue S, Huang Q (2020) Energy- and resource-aware computation offloading for complex tasks in edge environment. Complexity
15. Alqarni MM, Cherif A, Alkayal E (2021) A survey of computational offloading in cloud/edge-based architectures: strategies, optimization models and challenges. KSII Trans Internet Inf Syst (TIIS) 15(3):952–973
16. Zhang T (2017) Data offloading in mobile edge computing: a coalition and pricing based approach. IEEE Access 6:2760–2767
17. Xu Z et al. (2019) A time-efficient data offloading method with privacy preservation for intelligent sensors in edge computing. EURASIP J Wireless Commun Netw 2019(1):1–12
18. Sun H, Zhou F, Hu RQ (2019) Joint offloading and computation energy efficiency maximization in a mobile edge computing system. IEEE Trans Veh Technol 68(3):3052–3056
19. Khan MA (2015) A survey of computation offloading strategies for performance improvement of applications running on mobile devices. J Netw Comput Appl 56:28–40
20. Ning Z et al. (2018) A cooperative partial computation offloading scheme for mobile edge computing enabled Internet of Things. IEEE Internet Things J 6(3):4804–4814
21. Elgendy IA et al. (2020) Efficient and secure multi-user multi-task computation offloading for mobile-edge computing in mobile IoT networks. IEEE Trans Netw Serv Manag 17(4):2410–2422
22. Guo H, Liu J, Qin H (2018) Collaborative mobile edge computation offloading for IoT over fiber-wireless networks. IEEE Network 32(1):66–71
23. Valentino R, Jung W-S, Ko Y-B (2018) A design and simulation of the opportunistic computation offloading with learning-based prediction for unmanned aerial vehicle (uav) clustering networks. Sensors 18(11):3751
24. Luo J et al. (2019) QoE-driven computation offloading for edge computing. J Syst Architect 97:34–39
25. Kuang Z et al. (2019) Partial offloading scheduling and power allocation for mobile edge computing systems. IEEE Internet Things J 6(4):6774–6785
26. Saleem U et al. (2020) Latency minimization for D2D-enabled partial computation offloading in mobile edge computing. IEEE Trans Veh Technol 69(4):4472–4486
27. Qin M et al. (2017) Service-oriented energy-latency tradeoff for iot task partial offloading in mec-enhanced multi-rat networks. IEEE Internet Things J 8(3):1896–1907. Mach P, Becvar Z (2017) Mobile edge computing: a survey on architecture and computation offloading. IEEE Commun Surv Tuts 19(3):1628–1656, 3rd Quart
28. Nir MPS (2014) Scalable resource augmentation for mobile devices. Ph.D. dissertation, Department of Systems and Computers Engineering, Carleton University, Ottawa, ON, Canada
29. Wang Y, Sheng M, Wang X, Wang L, Li J (2016) Mobile-edge computing: partial computation offloading using dynamic voltage scaling. IEEE Trans Commun 64(10):4268–4282
30. Rahmati A, Zhong L (2007) Context-for-wireless: Context-sensitive energy-efficient wireless data transfer. In: Proceedings of 5th international conference mobile systems applications services, pp 165–178
31. Lin Y-D, Chu ET-H, Lai Y-C, Huang T-J (2013) Time-and-energyaware computation offloading in handheld devices to coprocessors and clouds. IEEE Syst J 9(2):393405
32. Miettinen AP, Nurminen JK (2010) Energy efficiency of mobile clients in cloud computing. In: Proceedings of 2nd USENIX conference hot topics in cloud computing, p 4
33. Gember A, Dragga C, Akella A (2012) ECOS: practical mobile application offloading for enterprises. In: Proceedings of international conference mobile systems application services, p 4

34. Mehmeti F, Spyropoulos T (2013) Performance analysis of 'on-the-spot' mobile data offloading. In: Proceedings global communication conference (GLOBECOM), pp 1577–1583
35. Mehmeti F, Spyropoulos T (2014) Is it worth to be patient? Analysis and optimization of delayed mobile data offloading. In: Proceedings INFOCOM:2364–2372
36. Jiang C et al. (2019) Toward computation offloading in edge computing: a survey. IEEE Access 7(2019):131543–131558
37. Yuan C, Chen Y, Zhang Z (2004) Evaluation of edge caching/offloading for dynamic content delivery. IEEE Trans Knowl Data Eng 16(11):1411–1423
38. Wu H, Knottenbelt W, Wolter K, Sun Y (2016) An optimal offloading partitioning algorithm in mobile cloud computing. In: Proceedings international conference quantitative evaluation systems, pp 311–328
39. Chen M, Hao Y, Qiu M, Song J, Wu D, Humar I (2016) Mobility-aware caching and computation offloading in 5G ultra-dense cellular networks. Sensors 16(7):974
40. Cuervo E et al (2010) MAUI: making smartphones last longer with code offload. In: Proceedings of the 8th international conference mobile systems applications services, pp 49–62
41. Zhou, Yu FR, Chen J, Kuo Y (2017) Resource allocation for information-centric virtualized heterogeneous networks with in-network caching and mobile edge computing. IEEE Trans Veh Technol 66(12):11339–11351
42. Lin Y, Kemme B, Patino-Martinez M, Jimenez-Peris R (2007) Enhancing edge computing with database replication. In: Proceedings 26th IEEE international symposium reliable distributed systems (SRDS):45–54
43. Malik SUR et al. (2021) EFFORT: energy efficient framework for offload communication in mobile cloud computing. Softw Pract Experience 51(9):1896–1909
44. Wu H, Huang D (2014) Modeling multi-factor multi-site risk-based offloading for mobile cloud computing. In: Proceedings 10th international conference networks service management (CNSM), pp 230–235
45. Flores H, Srirama SN (2014) Mobile cloud middleware. J Syst Softw 92:8294
46. Wang C, Li Z (2004) A computation offloading scheme on handheld devices. J Parallel Distrib Comput 64(6):740–746
47. Hyytiä E, Spyropoulos T, Ott J (2015) Offload (only) the right jobs: Robust offloading using the Markov decision processes. In: Proceedings 16th international symposium world wireless, mobile multimedia networks (WoWMoM), pp 19
48. Kim Y, Lee K, Shroff NB (2014) An analytical framework to characterize the efficiency and delay in a mobile data offloading system. In: Proceedings of the 15th international symposium on mobile ad hoc networking computing, pp 267–276
49. Mehmeti F, Spyropoulos T (2016) Stay or switch?: Analysis and comparison of delays in cognitive radio networks with interweave and underlay spectrum access. In Proceedings 14th ACM international symposium mobility management wireless access (MobiWac), pp 9–18
50. Jia M, Cao J, Yang L (2014) Heuristic offloading of concurrent tasks for computation-intensive applications in mobile cloud computing. In: Proceedings IEEE conference computing communication workshops (INFOCOM WKSHPS), Toronto, ON, Canada, pp 352–357
51. Chen X, Jiao L, Li W, Fu X (2016) Efficient multi-user computation offloading for mobile-edge cloud computing. IEEE/ACM Trans Netw 24(5):2795–2808
52. Khan MA (2015) A survey of computation offloading strategies for performance improvement of applications running on mobile devices. J Netw Comput Appl 56:28–40

Study of the Impact of Sybil Attack in VANETs Using F2MD

T. Pavithra, B. S. Nagabhushana, and Suchismitha Das

Abstract According to a survey, 137,726 accidents have occurred in the year 2018 alone. In order to avoid such accidents and also to provide other services, research on Vehicular Adhoc networks started. VANETs have hardware called OnBoardUnits (OBU) situated on vehicles and Fixed infrastructure situated on the Roadside called RoadSideUnits (RSU). Standards have been developed that allows communication between vehicle to vehicle and Vehicles to RSUs. Also, separate Bandwidth standard has been defined called Dedicated Short Range Communication (DSRC) to provide communication during emergency. However, we are yet to identify the associated risk of privacy and security in such communication. Then only successful implementation of Intelligent Transportation System can be achieved. Hence study of various attacks and its impact on Vehicular network is very much important. In this paper, we will perform the simulation of VANET and study the various performance metrics by introducing Sybil attack and analyse its impact over the network.

Keywords Vehicular adhoc NETwork (VANET) · Framework for misbehavior detection (F2MD) · SUMO (simulation of urban mobility) · TraCI (traffic control interface) · ITS (intelligent transportation system) · MBD (misbehavior detection) · Vehicles in network simulation (Veins) · Objective modular network testbed (Omnet++)

T. Pavithra (✉)
Department of Electronics and Communication Engineering Government, SKSJT Institute, Bangalore, India
e-mail: pavithratnaik@gmail.com

B. S. Nagabhushana
Department of Electronics and Communication Engineering, BMS College of Engineering, Bangalore, India

S. Das
Tata Consultancy Services, Bangalore, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_40

1 Introduction

According to a recent survey, India reports maximum number of road accidents globally, with almost 1.5 lakh people getting killed and more than 4.5 lakh people getting into minor injuries, annually. The biggest reason for accidents is found to be the speed. One of the main goals of VANET is to ensure safety of the users. VANET is a type of MANETS (mobile Adhoc networks) in which vehicles are the mobile nodes which are served by the Road Side Units (RSUs) which act as the Access Points (APs). Each vehicular nodes can freely move in the network and nodes can communicate with each other with single-hop or multi-hop [2]. The main information used by the driver from VANET for guidance are traffic information, road information, and collision [1]. With proper interpretation and post processing messages, drivers become aware of the situation and the can take necessary decisions to prevent road accidents. At the same time, there is a possibility of inducing false messages or network may be flooded with false message to make the network unavailable to the genuine users by inducing some attacks in the network. Therefore, it is necessary to identify attacks and eliminate such attacking nodes from the network.

In this paper we discussed a type of attack called Sybil attack. The Sybil attack is named after the book Sybil, a case study of a woman with multiple personality disorder. Sybil attack is an impersonation attack and in this type of attack, an attacker node creates an illusion among other nodes of a network as a set of nodes by taking false identities, or by creating new identities at the worst case. It is very easy to execute such attacks in adhoc network environment of which VANETs is a representative case. Here, an attacker node will send messages with different pseudonyms to other nodes present in the network. The nodes sending spoofed messages to other nodes are attacking nodes and nodes receiving the spoofed information are called Sybil nodes. All other types of attacks get easily emulated in the presence of Sybil attack. Such cases may just be creating an illusion of a traffic jam by using some fabricated non-existent nodes or an illusion of accident happened so that other vehicles can change their route [3].

The detection of Sybil attacks is quite difficult in VANETs. A node that is participating actively in VANET can take a fake identity and cause network disruption. Such an action is called misbehaviour in VANETs and it creates a safety issue in the network. Misbehaviour Detection (MBD) in Vehicular networks is a field of research which focus on detection of intrusion of nodes that causes trouble to the transportation safety. The Algorithms developed to counter MBD could either be sensor-based or Vehicle to Vehicle and Vehicle to RSU based. Hence, it is necessary to simulate the VANETs and evaluate the developed MBD algorithms using simulation results [4]. The work presented herewith is a contribution in this specific space.

2 Literature Review on Sybil Attack

Sybil attack was first introduced by Douceur [5] during his study on security issues in peer-to-peer networks. Then, Karlof and Wanger proved that Sybil attack cab pose a serious threat to network routing in case of Wireless Sensor Networks [6]. Grover et al. have analysed the effect of the Sybil attack on the performance of VANET by measuring the packet delivery ratio, throughput, packet drop and number of collisions with variation in node mobility by performing simulation. They executed different types of Sybil attacks over the vehicular network and the results obtained after simulating the network showed that the Quality of Service provided by the Vehicular network degrades if Sybil attack happens in the network and hence study of Sybil attack is important in VANETs [7]. They have proposed an approach for distributed Sybil attack, in which, nodes exchange the observed details with their neighbours cooperatively and help RSU in the detection of attack.

Kanwalpreet Singh et al. have proposed an interesting technique for detection of Sybil attack. This technique makes use of signal strength and vehicles need to be in monitor mode. In this technique, the RSU will flood the ICMP messages through the network. The vehicles will send their signal strength to their nearest RSU while receiving the ICMP messages. The RSU will collect the entire information from the vehicles and exchange this data with other RSUs. If multiple signal strength is received for a specific node, it will be considered as the malicious node in the networks. In order to get a confirmation about the malicious node, the RSUs will send the control packets in the network. After receiving the control packets, vehicles will enter into the monitor mode and start watching their adjacent nodes. Once the malicious node is sensed, the technique of multiple path routing is applied to isolate the malicious nodes from the network. In [8], authors arrived at a conclusion that broadcasting technique is very efficient to select an efficient route from source to the destination. Due to decentralized nature of the network, it is quite easy for a malicious node to join the network. These nodes are responsible for various types of active and passive attacks [9–11].

One of the commonly used mobility and network simulators is Vehicle Network Simulator: VEINS. VEINS is a combination of Objective Modular Network Testbed in C++: OMNET++ and Simulation of Urban Mobility: SUMO. The simulator will pair an OMNeT ++ node for each vehicle present in the simulation and then pairs node movements with movements of vehicles in the road traffic simulator (i.e., SUMO). Then, both the network simulator and mobility generator can run in parallel[13]. This is possible due to a bidirectional coupling by a standardized connection protocol, called the Traffic Control Interface (TraCI). Unlike to NS2, it does not require any Vehicular mobility generator separately and TraCI helps in communication between Omnet++ and SUMO.

One of the well-known simulation frameworks which allows us to evaluate the efficiency of Misbehavior Detection algorithms in Vehicular networks is Vehicular Reference Misbehaviour (VeReMi). It is an extension of VEINS and it has a dataset that consists of transmission data from both misbehaving nodes and genuine nodes. This can be used for the evaluation of Misbehaviour Detection algorithms offline. VeReMi also provides five attacks based on position and four algorithms for Misbehaviour Detection. In [4], authors have proposed simulation framework named Framework for Misbehaviour Detection: F2MD based on VEINS. F2MD allows performing the studies on a large set of vehicles and can be simulated using a large road network.

F2MD can perform Local Misbehaviour Detection running a Misbehaviour Detection system by each entity to cope with insider attackers. In our paper, we are using veins-F2MD to evaluate the effect of Sybil attack on VANET and local detection of Sybil attack using MBD algorithm.

3 Sybıl Attack

Sybil attack is a type of network attack where the attacker creates large number of fake identities associated with a single node to give an illusion of large number of nodes in a network[15]. These fake identities are called Pseudonyms. An attacker uses these pseudonyms to perform an attack such as creating an illusion of congestion in the network. There are multiple ways to create the fake identities for a sybil node. Also, the type of sybil attacks are different such as distributed storage, routing and data summarization.

A type of Sybil attack includes generating the vehicles called ghost vehicles that will be used to perform attack. The sequence of actions will be used to perform the sybil attack include:

- Here, the speed, direction and Position of the ghost vehicle will be decided based on the target vehicle.
- Many number of pseudonyms will be generated and maintained. One pseudonym will beused for each ghost vehicle.
- Attacker vehicle's beaconing frequency will be increased based on the number of ghost vehicles.

4 Attacker Model

In order to perform the simulation, we are making the following assumptions.

Insider Attacker: The attacker will possess the required credentials to communicate in a C-ITS. Therefore, any node in the network can be converted to a malicious node.

Active: Here, the attacking node will participate actively in C-ITS communication and sends beaconing data. He can even modify the data as he can access the data traversing through the network.

Pseudonym certificates access: Here, the assumption is made that, the attacker has full access to Pseudonym certificate.

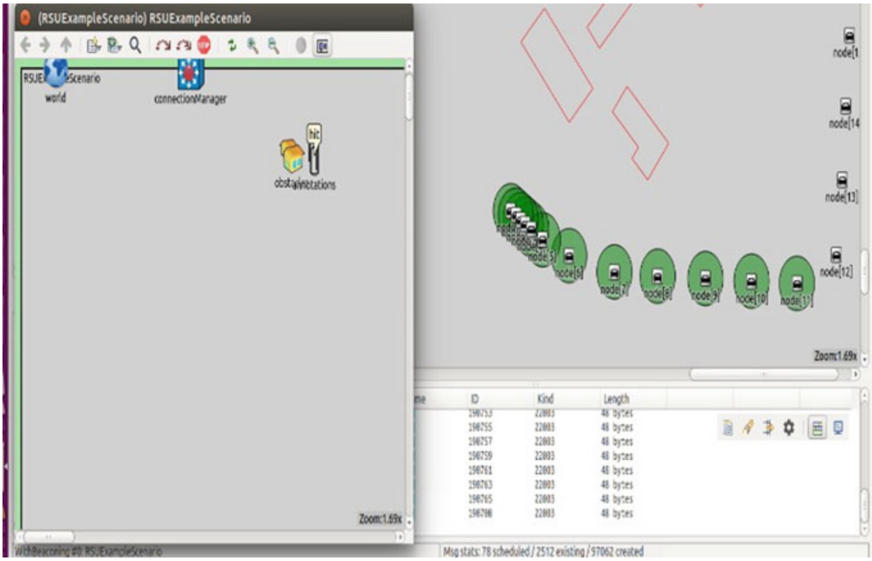
5 Sımulatıon

5.1 Evaluation

Veins is built on: OMNET++ (discrete event simulator) and SUMO (traffic simulator). SUMO will provide the traffic patterns for the realistic map and OMNET++ will provide different modules such as Application layer, Dedicated Short Range Communication, and a Physical layer to provide the realistic network behaviour. A small patch called Traffic Control Interface (TraCI) is used to establish a connection between OMNET++ and SUMO.

To evaluate the effect of Sybil attack in VANETs, we created a map using OSM Web Wizard in SUMO as shown in Fig. 1.

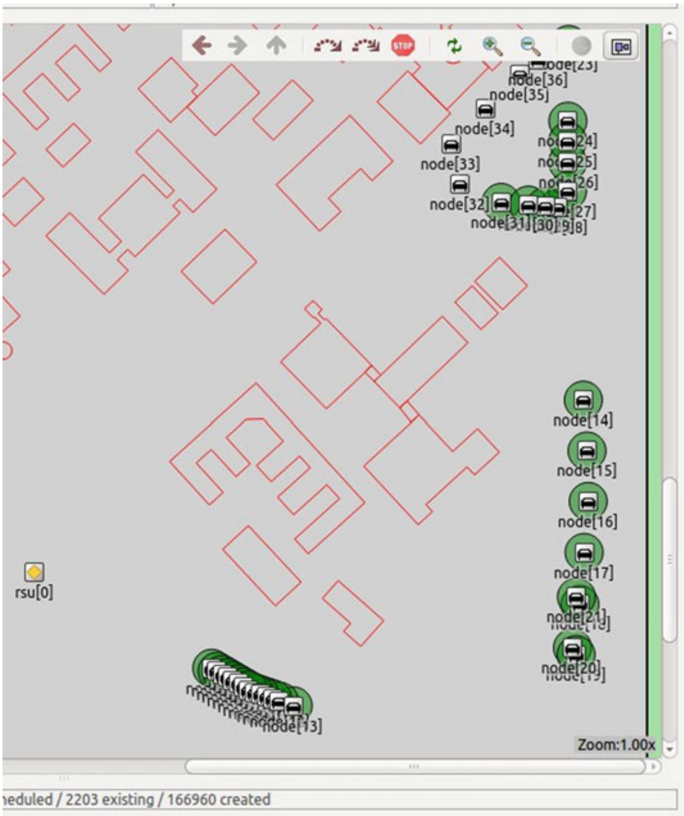
Firstly, we have introduced random vehicles in the default map of the veins network. When nodes are moving through the network, accident event is introduced at a random Location and time. The nodes send the event details to their neighbouring



...me	ID	Kind	Length
	190753	22003	48 bytes
	190755	22003	48 bytes
	190757	22003	48 bytes
	190759	22003	48 bytes
	190761	22003	48 bytes
	190763	22003	48 bytes
	190765	22003	48 bytes
	190766	22003	48 bytes

Msg stats: 78 scheduled / 2512 existing / 97062 created

Fig. 1 Map created using OSM web wizard in SUMO



ıeduled / 2203 existing / 166960 created

Fig. 2 Vehicles diverting routes to avoid congestion

nodes through broadcasting. Upon reception of this information, vehicles wait for some time and divert to some other route so that they can avoid the congestion as shown in Fig. 2.

Then, we injected the malicious nodes by 10%, 30%, 50%, and 70% the network to understand the effect caused by such attackers over the network.

The below table 1 shows the simulation parameters and its corresponding values.

5.2 Performance Evaluation Metrics

For the evaluation of the performance over Vehicular networks in presence of Sybil attack, the following metrics have been considered [12].

Table 1 Simulation setup parameters

Attributes	Value
Network simulator	Omnet++ 5.4.1
Traffic simulator	Sumo 1.0.1
Framework V2x simulator	Veins-F2md
No. of vehicles	–
No. of malicious nodes	0, 10, 30, 50, 70%
No. of RSU	09
Simulation area	8 × 8 km
Communication range	1200 m
Simulation time	50 s

5.2.1 Content Delivery Ratio (CDR)

Content refers to the information that a vehicle want o convey.Since, VANETs have been developed with the interest of avoiding accidents by sending the sensitive data, delivery of the information to the intend vehicle is very important. Hence, the calculation of Content Delivery Ratio gives the major information about how efficiently data is sent/received over a network. CDR indicates the number of messages that are successfully received by the genuine vehicles.

Let,

MR represent the the number of messages received.

MPRE represent the number of messages that are expected to be received in the network,

CDR can now be calculated as:

$$\mathrm{CDR} = \mathrm{MR}/\mathrm{MPRE}$$

We require this ratio to be high to assume a network to be efficient.

5.2.2 Packet Loss Ratio (PLR)

When the CDR is low, it indicates that network is not receiving as much information as needed which clearly says that some data is leaking. Therefor we need to calculate the amount of information lost. PLR gives the information about the number of messages that are lost due to the misbehaviour action of malicious nodes. If MT represents the total number of messages and out of MT, if ML number of messages are dropped in the network, then PLR can be calculated as:

$$\mathrm{PLR} = \mathrm{ML}/\mathrm{MT}$$

In general, MT indicate the messages received by both genuine and malicious nodes. If MR represents the number of messages received by genuine nodes and ML represents the number of messages dropped at the Malicious nodes, then the total number of messages MT can be written as:

$$\mathrm{MT} = \mathrm{MR} + \mathrm{ML}$$

In result section we are going to find how these paraameters get affected with increase in Sybil attack.

6 Results

In this section we will see how CDR and PLR values vary with variation in the attacking vehicle density.

As we can see in Fig. 3, the content delivery ratio is reduced with an increase in Attack Percentage.CDR was 100% without any attack and gradually reduces with increase in the percentage of attack. Therefore, Content Delivery Ratio can be considered as one of the parameter to identify that attack is being introduced in the network.

Packet Loss Ratio, as shown in Fig. 4 increases and False message also increases as shown in Fig. 5 with an increase in attack percentage.

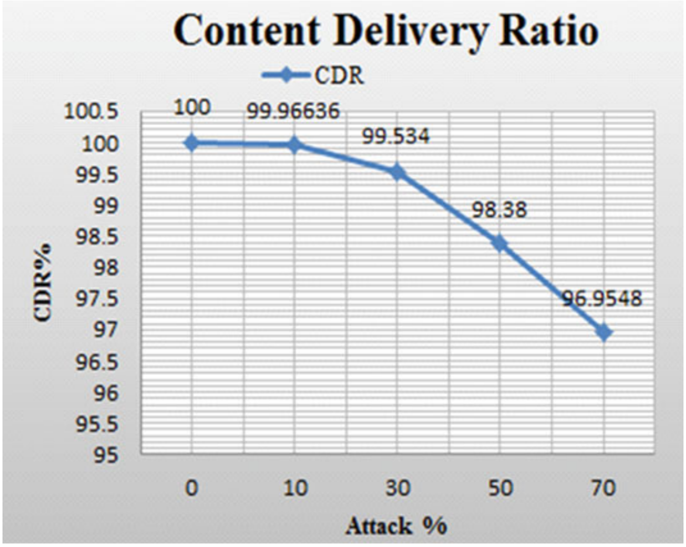


Fig. 3 Attack percentage versus content delivery ratio

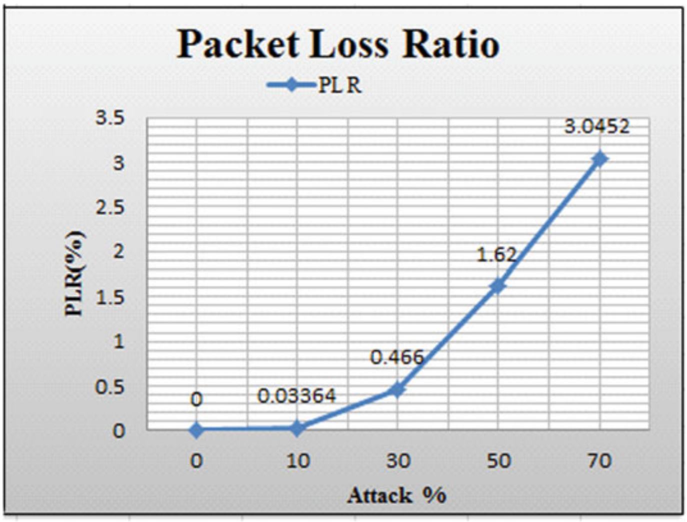


Fig. 4 Attack percentage versus packet loss ratio

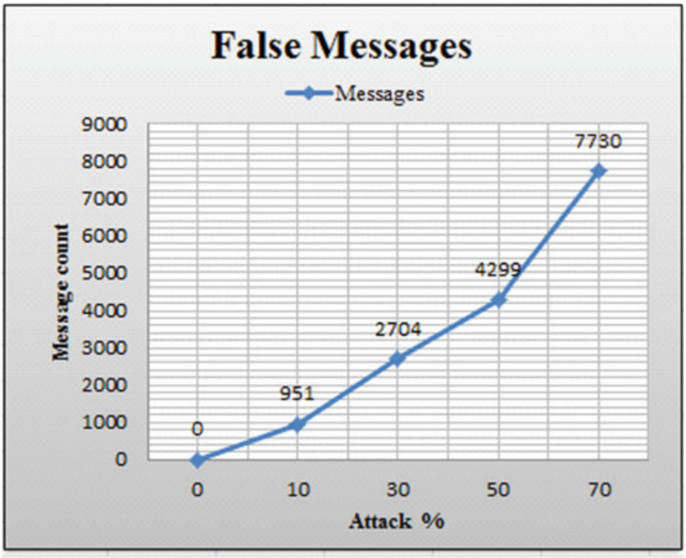


Fig. 5 Attack percentage versus false message

7 Conclusion

From our simulation result, it is clear that sybil attack affects the network performance by reducing the Content Delivery Ratio and by incresing the Packet loss and False messages. But the security in VANETs is very much important. Many researcher are working on these issues to learn how security can be provided in VANETs. Author of [14] proposed a method involving cloud computing to improve the network performance and The approximate member query filter is utilised to avoid interaction between attackers.

Inorder to avoid VANET attcks so as to keep the performance of the network high, we need to detect the attack as soon as it is introduced in the network. This requires developing a trained model which can detect the and in turn, developing a trained model requires too much of data both with network attack and without any attack in the network. Therefore, In our next paper, we will use the available data from different research institutes and learn how different Machine Learning models help us in identifying the attack and which model gives better classification accuracy in identifying the attack.

References

1. Jain M, Saxena R (2017) Overview of VANET: requirements and its routing protocols. In: International conference on communication and signal processing, India, pp 6–8
2. Samara G, Al-Salihy WAH, Sures R (2011) Security analysis of vehicular Ad-Hoc networks (VANET), National advanced IPv6 Center, Universiti Sains Malaysia, Penang, Malaysia; *School of Computer Science, Universiti Sains Malaysia, Penang, Malaysia. In: 2010 Second ınternational conference on network applications, protocols and services. network security & ıts applications (UNSA), vol 3, no 6
3. Rahbari M, Jamali MAJ (2011) Effıcıent detectıon of sybıl attack based on cryptography ın VANET, Department of Computer Science, Shabestar Branch, Islamic Azad University, Shabestar, Iran. Int J Netw Secur Appl (UNSA) 3(6)
4. Kamel J, Ansari MR, Petit J, Kaiser A, Jemaa IB, Urien P (2013) Simulation framework for misbehavior detection in vehicular networks. IRT SystemX, Palaiseau, France; Telecom Paris Tech, Paris, France; #OnBoard Security , Inc., Wilmington, USA
5. Patil HK, Chen TM Chapter 18—Wireless sensor network security: the ınternet of things, pp 317–337
6. Karlof C, Wagner D (2003) Secure Routing in Wireless Sensor Networks: Attacks and Countermeasures. Ad Hoc Netw 1:293–315
7. Grover J, Kumar D, Sargurunathan M, Gaur MS, Laxmi V (2010) Performance evaluation and detection of Sybil attacks in vehicular Ad-Hoc networks. Commun Comput Inf Sci 89:473–482
8. Kanwalprit Singh, Harmanpreet Kaur, "Evaluation of Technique for detection of Sybil Attack on VANET"
9. https://veins.car2x.org/
10. https://doc.omnetpp.org/omnetpp/manual/
11. https://sumo.dlr.de/docs/
12. Sommer C, German R, Dressler F (2011) Bidirectionally coupled network and road traffic simulation for ımproved IVC analysis. IEEE Trans Mob Comput (TMC) 10(1):3–15
13. Weber J, Neves M, Ferreto T (2021) VANET simulators: an updated review. J Braz Comput Soc 27:8. https://doi.org/10.1186/s13173-021-00113-x
14. Neelaveni R (2019) Performance enhancement and security assistance for VANET using cloud computing. J Trends Comput Sci Smart Technol 1(1):36–45. https://doi.org/10.36548/jtcsst.2019.1.004

15. Pavithra T, Nagabhushana BS (2020) A survey on security in VANETs. In: 2020 Second ınternational conference on ınventive research in computing applications (ICIRCA), pp 881–889

Aatmanirbhar Sanchar: Self-Sufficient Communications

Jay Jhaveri, Abhay Gupta, Prem Chhabria, Neeraj Ochani, Sharmila Sengupta, Mrs. Sunita Suralkar, and Shashi Dugad

Abstract In the light of recent war crimes and data piracy conspiracies, privacy is of utmost importance to an organization and even to an individual. The majority of the population is dependent on third-party services for their daily communication. Albeit these major corporations advertise "secure" means of chat transfer, they install various kinds of backdoors to sell the user's data to advertisers. Under the notion of going "Aatmanirbhar" i.e., Make in India, we have developed an indie solution without incorporating any third-party services or APIs. "Aatmanirbhar Sanchar" aims at providing users with a real-time off-the-grid, secure, and anonymous messaging service. It features an End-to-End encrypted transmission of messages and data files likewise. This is achieved by combining the open-source AES algorithm with a self-developed XOR encryption process.

Keywords Messaging · Security · Privacy · Self-sufficient · Aatmanirbhar · Self-hosting · AES · Hash-based message authentication code verification

J. Jhaveri (✉) · A. Gupta · P. Chhabria · N. Ochani · S. Sengupta · Mrs. S. Suralkar
Computer Engineering, Vivekanand Education Society's Institute of Technology, Mumbai, India
e-mail: 2018.jay.jhaveri@ves.ac.in

A. Gupta
e-mail: 2018.abhay.gupta@ves.ac.in

P. Chhabria
e-mail: 2018.prem.chhabria@ves.ac.in

N. Ochani
e-mail: 2018.neeraj.ochani@ves.ac.in

S. Sengupta
e-mail: sharmila.sengupta@ves.ac.in

Mrs. S. Suralkar
e-mail: sunita.suralkar@ves.ac.in

S. Dugad
Tata Institute of Fundamental Research (TIFR), Mumbai, India
e-mail: shashi@tifr.res.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_41

(HMAC) · Indigenous private server · Scalability · Cross plat-form · Throwaway · Anonymous

1 Introduction

In today's world, privacy and security are of utmost importance to an individual. Let me elaborate: Data Privacy and Data Theft are the hot debate in the World-Wide mass media at the moment, but have you ever wondered what exactly it is? Have you ever questioned how the so-called "Free" applications are kept afloat? They pay their bills by selling that very data you unknowingly give them while using their "free" services. This borderline stolen data is then used for targeted, personalized advertisements and much worse. To combat this, we are developing anonymous communication software without the use of any third-party services, hence maximizing the privacy of an individual. To put it simply, secure messaging is a way of safely exchanging documents between users, healthcare providers, organizations, and their customers.

2 Motivation

Currently, all other messaging services are hosted on Third-party cloud platforms, mainly Google and AWS cloud services. Let us consider WhatsApp for example, which has been recently acquired by The Facebook (META) team. After this acquisition, WhatsApp updated its privacy policy which gave access to Facebook to collect private information on its users causing many controversies and heated discussions in the IT industry around the globe.

There is a dire need for us to focus on these privacy problems faced by users using these "free" applications like WhatsApp and Facebook messenger. The companies owning these applications do not take adequate security measures in handling the user data but drive their marketing/advertising agenda through the data provided by their users.

Further, In the light of recent events, amidst the Russian-Ukraine War, there are major sanctions placed by the west on Russia, disabling them from using multiple western applications. Even their banking apps were restricted leading to a major downfall in their economy. Now, India has developed a native solution for the banking sector called the Unified Payments Interface, also known as UPI. Why not take this spirit and create an indie chat application?

These were mainly our inspiration to create India's very own messaging service without utilizing any kind of third-party services. In collaboration with the Tata Institute of Fundamental Research (TIFR), we have built a secure communication, cross-platform messaging application wherein a user can exchange vital information with other users and groups of users without being concerned about any kind of data leak or data monetization.

3 Literature Review

Cohn-Gordon et al. [1] in "A Formal Security Analysis of the Signal Messaging Protocol" explained that Signal Protocol is a private messaging protocol that provides instant messaging encryption to applications such as Skype, Facebook Messenger, and WhatsApp among many others, with more than 1 billion active users. The signal contains unique unfamiliar security features (such as "future privacy" or "post-compromise security"), which are made possible by ratcheting, a process through which session keys are updated with each new message.

Singh et al. [2] in "Blockchain-Enabled End-to-End Encryption for Instant Messaging Applications" presented a blockchain-based E2EE framework for mitigating current messaging application vulnerabilities. During the installation of the application, the end-user device generates a pair of public/private keys and asks its mobile network operator to issue a digital Identifier and store it in the blockchain. The end user can obtain another user's certificate from their chat private server and utilize a ratchet forward encryption process to interact securely with them.

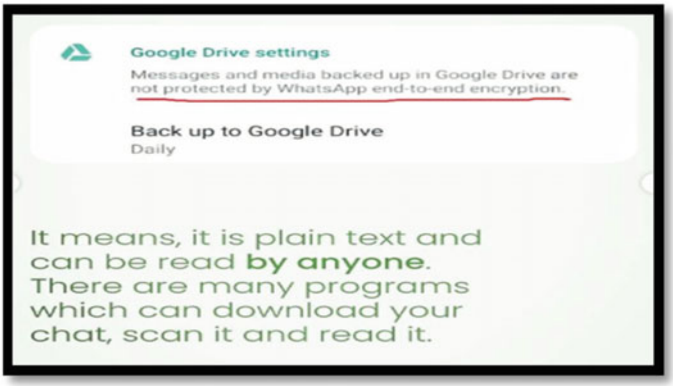
Botha1 et al. [3] in "A Comparison of Chat Applications in Terms of Security and Privacy, ECCWS 2019 18th European Conference on Cyber Warfare and Security" described a gadget that helps people adapt to social life by allowing them to understand domain messages, names, letters used in mailboxes, in daily newspapers, and so on. The major goal of the project is to solve the above problem by using a Raspberry Pi and an OCR sensor to recognize environmental messages automatically and then using TTS to translate those messages into voice or audio for better and easier engagement with society.

Sabah et al. [4] in "Developing an End-to-End Secure Chat Application" presented a chat program that provides end-to-end security, allowing users to safely transmit confidential information without fear of data loss. In addition to the storage protection. This article presents a list of requirements for creating a secure chat application, and the program was created based on these requirements. The suggested chat application was compared to other popular apps based on those criteria, and it was also put to the test as a proof of concept for delivering End-to-End security.

Burak [5] in "Encryption Methods and Comparison of Popular Chat Applications" proposed end-to-end encryption chat solutions that allow users to safely transmit personal information. The paper includes a list of requirements for developing a secure chat application.

Canetti [6] in "Universally Composable End-to-End Secure Messaging" explained all the contemporary widely accepted encryption algorithms in detail and their limitations in the real practical world. It also helped in choosing the most suitable encryption algorithm for this chat application.

Emura [7] in "Membership Privacy for Asynchronous Group Messaging" focuses on a method capable of hiding membership information from the viewpoint of non-group members in a secure group messaging (SGM) protocol, which we call "membership privacy".



Google Drive settings

Messages and media backed up in Google Drive are not protected by WhatsApp end-to-end encryption.

Back up to Google Drive

Daily

It means, it is plain text and can be read **by anyone**. There are many programs which can download your chat, scan it and read it.

Fig. 1 WhatsApp Plain text Backup Proof

4 Lacuna in the Existing System

"Data is the new fuel" and major tech corporations are utilizing every gizmo at their disposal to amass and utilize user data for monetary advantage because their customers' personal and behavioral data is worth millions of dollars if mined to its full potential.

WhatsApp and other Applications [8] provide the option of verifying users' public keys, but the mechanisms used are not robust and pose major session hijacking issues. Besides this, there is no reliable third-party involvement to check the suitability of keys stored on WhatsApp servers [9–13].

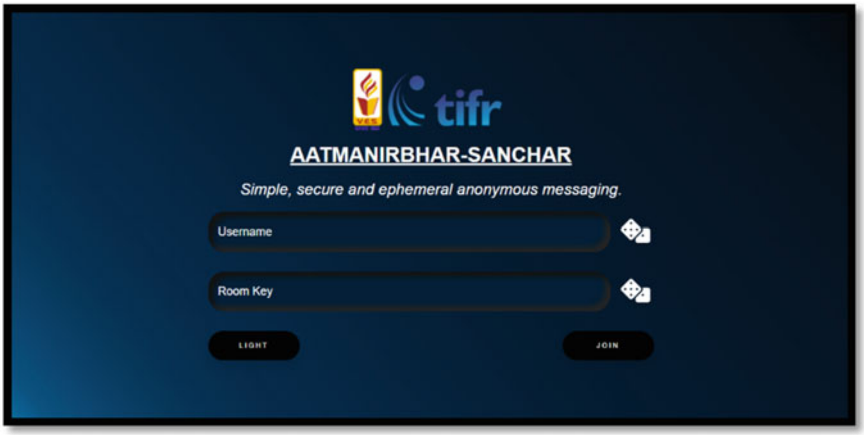
The backup method utilized by WhatsApp does not provide real end-to-end encryption (see Fig. 1). The alternate copy is kept in plain text on the user's cloud, depending on the user's OS, such as iCloud, Google Drive, One Drive, and so on.

5 Methodology

5.1 Joining a Chat Room

When a user visits the web application at http://aatmanirbhar-sanchar.live/, hosted on the private servers at Vivekanand Education Society's Institute of Technology (VESIT), he is greeted with the homepage asking for a Username along with a Room key (RK) (see Fig. 2). The entered username will act as the main identity of that particular user for the ongoing session. The room key is the most significant aspect of the chat application. This key serves a dual purpose:

1. The hash of the key acts as the identity of a particular group chat created using the same.



AATMANIRBHAR-SANCHAR

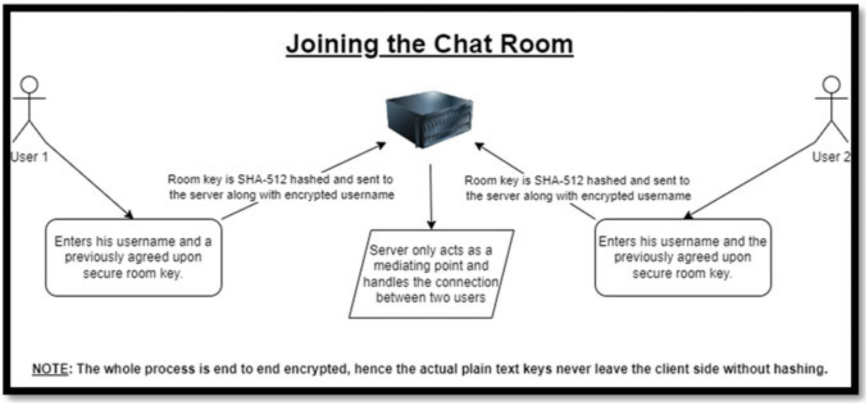
Simple, secure and ephemeral anonymous messaging.

Fig. 2 Home page

2. The key itself is used in the encoding and decoding process of incoming and outgoing messages.

When the user enters a room key (RK), the key first goes through an extensive algorithm to test the strength of the key. If the resulting strength is not up to the standards for secure communication, it will warn the user of a weak key and the user may decide if he wants to proceed or add a new key.

A passphrase is highly recommended instead of a password to ensure utmost privacy while communicating. To keep it user-friendly, a random passphrase generator has been added. Now, when the user clicks the join button, the following processes occur (see Fig. 3):



Joining the Chat Room
User 1
User 2
Room key is SHA-512 hashed and sent to the server along with encrypted username
Room key is SHA-512 hashed and sent to the server along with encrypted username
Enters his username and a previously agreed upon secure room key.
Server only acts as a mediating point and handles the connection between two users
Enters his username and the previously agreed upon secure room key.

NOTE: The whole process is end to end encrypted, hence the actual plain text keys never leave the client side without hashing.

Fig. 3 Joining process

1. The key is first hashed using the SHA-256 algorithm to get ready to be transmitted to the server.
2. *NOTE:* The plain text key never leaves the client side
3. The username, AES encrypted using the RK, is appended along with the hashed RK and is sent to the server.

The user can now share this room key with the intended recipients to begin a secure private communication channel with him/her.

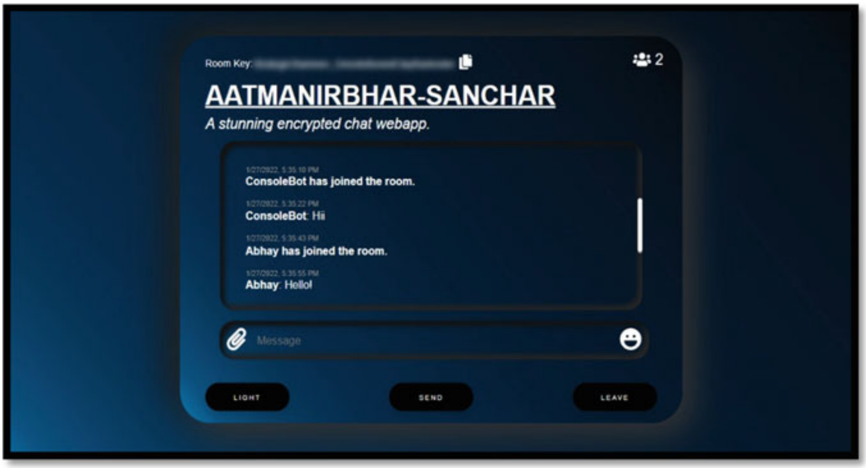
5.2 The Ephemeral Chat Room

When a user joins a chat room using the shared RK, the active user counter is incremented and their encrypted username is broadcasted to all the users active in the room utilizing which a greeting message is displayed (see Fig. 4).

Now every user is subscribed to the following events:

1. Join Response: *Handles a new incoming user.*
2. Chat Response: *Handles incoming text messages.*
3. File Response: *Handles incoming files.*
4. Leave Response: *Handles a user leaving.*

When a user joins a chat room using the shared RK, the active user counter is incremented and encrypted.



1/27/2022, 5:35:10 PM
ConsoleBot has joined the room.

1/27/2022, 5:35:22 PM
ConsoleBot: Hii

1/27/2022, 5:35:43 PM
Abhay has joined the room.

1/27/2022, 5:35:55 PM
Abhay: Hello!

Fig. 4 The chat room

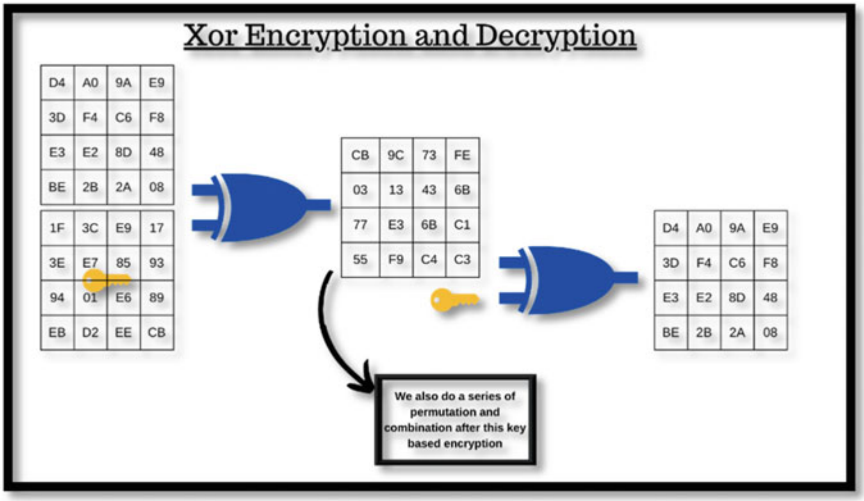
5.3 Encryption Process

Aatma Sanchar uses a double-layered encryption process for achieving enhanced security. The first layer constitutes the self-developed XOR encryption process:

This encryption system is based on the concept that if an object is XOR'ed by the same key twice it will revert to its original state. To make this viable in this innovative era of cybercriminals vs cybersecurity, multiple iterations of permutations and combinations on the original entity take place before further encoding. To put it simply, a text message is first converted to its binary format in the shape of matrices. These matrices are then shuffled and reshuffled to increase protection followed by undergoing the XOR process by the room key 10.

This encryption system is based on the concept that if an object is XOR'ed by the same key twice it will revert to its original state (Fig. 5). To make this viable in this innovative era of cybercriminals vs cybersecurity we have added multiple iterations of permutations and combinations of the original entity. To put it simply, a text message is first converted to its binary format in the shape of matrices. These matrices are then shuffled and reshuffled to increase protection followed by it getting XOR'd by the room key [9].

This encryption layer is followed by the Advanced Encryption Standard (AES-256) algorithm 11 to ensure privacy while maintaining efficiency. From the Graph in Fig. 6, it is clear that while there exist faster encryption algorithms other than AES, as the file size increases (which is a common situation in a messaging platform), AES easily comes out on top. Hence, AES-256 was selected as the second layer in this encryption process.


Xor Encryption and Decryption
We also do a series of permutation and combination after this key based encryption

D4	A0	9A	E9
3D	F4	C6	F8
E3	E2	8D	48
BE	2B	2A	08

1F	3C	E9	17
3E	E7	85	93
94	01	E6	89
EB	D2	EE	CB

CB	9C	73	FE
03	13	43	6B
77	E3	6B	C1
55	F9	C4	C3

D4	A0	9A	E9
3D	F4	C6	F8
E3	E2	8D	48
BE	2B	2A	08

Fig. 5 Second layer encryption using the XOR matrix method

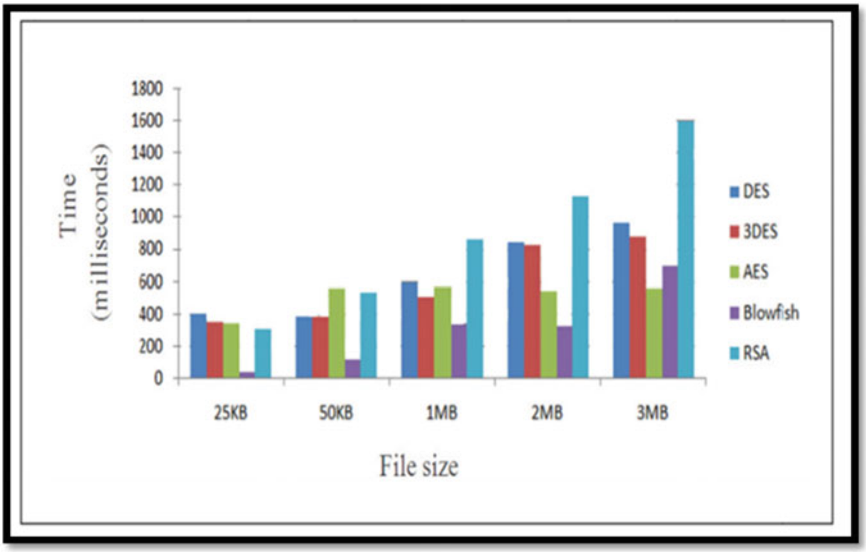


Fig. 6 Second layer encryption using the XOR matrix method

Finally, to ensure the integrity and authenticity of the transmitted message, the process of Hash-based Message Authentication Code verification (HMAC) is also practiced which guarantees tamper-proof messaging 12. This is done by creating a hash-based checksum using the combination of the room key and the data ready to be sent to the sender's client. This checksum value is then recomputed on the receiver's device, and if any discrepancy is detected it indicates that the message was tampered with (Fig. 13).

5.4 Transmitting a Message

The users can either directly type or send a text message using the provided chat box, else a user can also attach files up to the 50 MB limit to be transmitted.

1. Sending a text message (see Fig. 7):
   a. When a user types a message and hits the send button the message is encrypted using the Room Key (RK) utilizing the doubly layered encryption algorithm mentioned before. Further, the SHA-256 hashed RK and the encrypted Username are appended into a dictionary along with the encoded message.
   b. Next, we use the HMAC (hash-based message authentication code) algorithm to ensure the authenticity and integrity of the message being sent. The HMAC is generated using the above-created dictionary and the Room Key (RK).

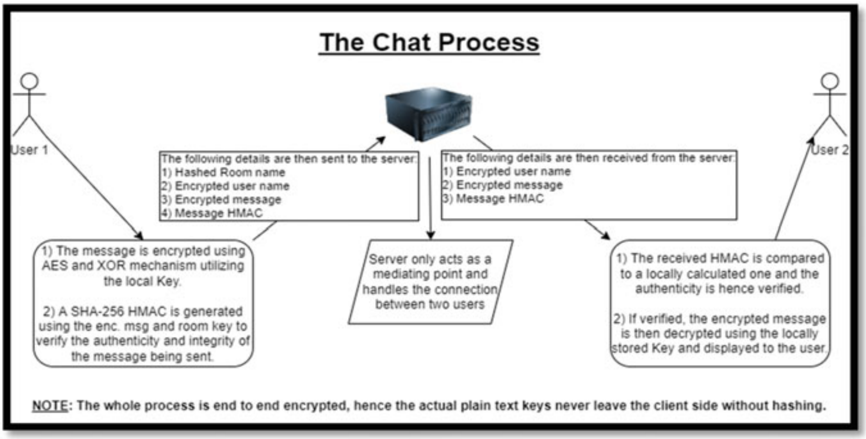

The Chat Process
User 1
User 2
The following details are then sent to the server:
1) Hashed Room name
2) Encrypted user name
3) Encrypted message
4) Message HMAC
The following details are then received from the server:
1) Encrypted user name
2) Encrypted message
3) Message HMAC
1) The message is encrypted using AES and XOR mechanism utilizing the local Key.
2) A SHA-256 HMAC is generated using the enc. msg and room key to verify the authenticity and integrity of the message being sent.
Server only acts as a mediating point and handles the connection between two users
1) The received HMAC is compared to a locally calculated one and the authenticity is hence verified.
2) If verified, the encrypted message is then decrypted using the locally stored Key and displayed to the user.
NOTE: The whole process is end to end encrypted, hence the actual plain text keys never leave the client side without hashing.

Fig. 7 Sending a text message

c. Finally, this encoded dictionary along with the HMAC appended to it is emitted through a socket to the server.

2. Sending a File as an Attachment (see Fig. 8):

a. A user also has the option to send any type of file as long as it is under the 50 MB limit. When he/she selects the file to be uploaded, the file is first converted into its binary (Base 64) format.

b. This binary format is encrypted using the Dual layer encryption process and is stored in the dictionary along with its file name and file type. Finally, as in the text messaging process, an HMAC code is calculated utilizing the above

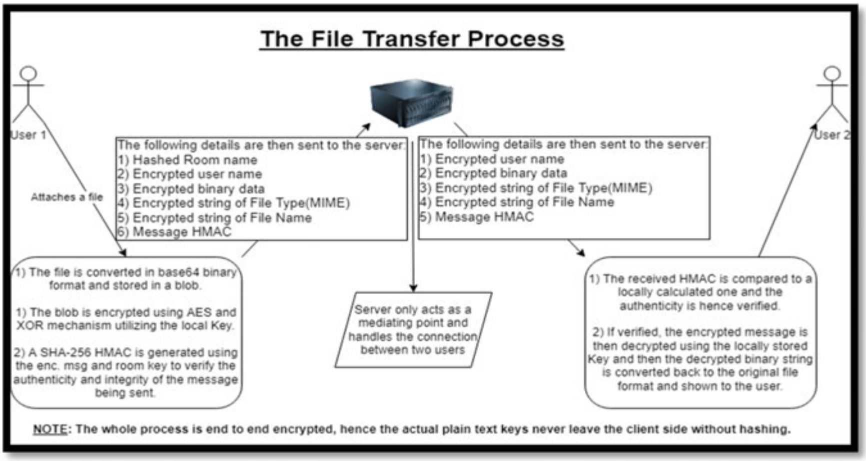

The File Transfer Process
User 1
User 2
Attaches a file
The following details are then sent to the server:
1) Hashed Room name
2) Encrypted user name
3) Encrypted binary data
4) Encrypted string of File Type(MIME)
5) Encrypted string of File Name
6) Message HMAC
The following details are then sent to the server:
1) Encrypted user name
2) Encrypted binary data
3) Encrypted string of File Type(MIME)
4) Encrypted string of File Name
5) Message HMAC
1) The file is converted in base64 binary format and stored in a blob.
1) The blob is encrypted using AES and XOR mechanism utilizing the local Key.
2) A SHA-256 HMAC is generated using the enc. msg and room key to verify the authenticity and integrity of the message being sent.
Server only acts as a mediating point and handles the connection between two users
1) The received HMAC is compared to a locally calculated one and the authenticity is hence verified.
2) If verified, the encrypted message is then decrypted using the locally stored Key and then the decrypted binary string is converted back to the original file format and shown to the user.
NOTE: The whole process is end to end encrypted, hence the actual plain text keys never leave the client side without hashing.

Fig. 8 File transfer process

dictionary and the room key. This is then sent using the same socket method as for a text message.

5.5 *Receiving a Message*

After a user has sent the encrypted message to the server then broadcasts the message to all the users currently connected to that particular room. On the receiver's end before any decryption process can start, the HMAC is recalculated on the client end utilizing the encrypted dictionary and if any disruption is found, an error is displayed in the chat box indicating the message was tampered with (Further explained in the Cryptanalysis part of the paper).

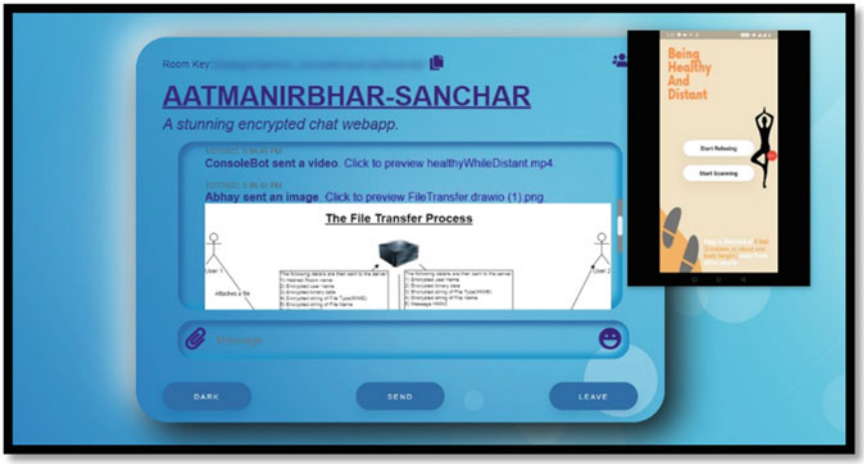
After verifying the HMAC code, the actual decryption process starts:

1. Receiving a text message (see Fig. 7):
   a. The incoming encoded dictionary is first decrypted using the AES algorithm followed by the reverse XOR method.
   b. This decrypted message is shown to the user in the chat box along with the decrypted username of the sender (see Fig. 4).
2. Receiving a file as an attachment (see Fig. 8):
   a. The incoming encoded dictionary is first decrypted using the AES algorithm followed by the reverse XOR method.
   b. This generates the file in its pure binary (Base 64) format. This binary file is then converted according to the MIME type into its original form.
   c. This original form is then converted into a blob link from where the receiver can download the same. If the File type is in a known multimedia format (Music, Image, Video), then the user is also given the option to preview the same within the chat box itself (see Fig. 9).

5.6 *Leaving the Chat Room*

A user can press the "leave" button to securely exit the chat room. Once pressed, the client is unsubscribed from all the live-time events and finally emits an encoded dictionary constituting the hashed room key and the encrypted user name. This username is then broadcasted to every active user with a message indicating that this person left the chat room.

Once left, the user cannot access the chat history and for enhanced security, the chats are never stored on the local device. If every user leaves a particular chat room, the session is destroyed in instantly.


Room Key: [redacted]

AATMANIRBHAR-SANCHAR

A stunning encrypted chat webapp.

ConsoleBot sent a video. Click to preview healthyWhileDistant.mp4

Abhay sent an image. Click to preview FileTransfer.drawio (1).png

Fig. 9 File preview in the chat room (*light mode*)

5.7 *Experimental Environment Details*

The following tools and technologies have been used in the development of Aatmanirbhar Sanchar:

- Front-End Development:

  React JS
  HTML/CSS

- Backend Development (server-side):

  Languages used:

  – NodeJS
  – Socket.IO

  Compatible Operating System:

  – Ubuntu (16.04)
  – Windows 10

  Requirements:

  – A Public Static IP to host the messaging application.
  – Android Studio for mobile applications (Auto File Sync).
  – Maps and Google Sheets API

6 Applications

6.1 In Large Organizations as a Quick, Secure Chat Platform

Our chat application being ephemeral does not store any of the chat data on the local client machine or the cloud server. This, apart from making our product more secure from any kind of unauthorized access to the local machines, also saves a lot of vital storage space in the cloud servers that can be essential for more important subjects. Being a throwaway chat application, it can be reused N number of times without any load on the server or the client.

6.2 Communications Where Security is of National Importance

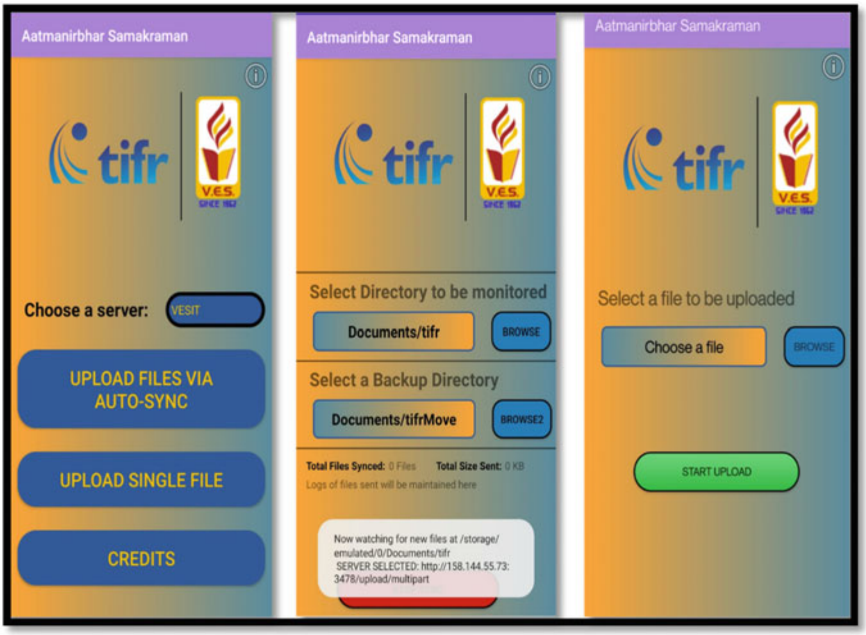
In matters where national security is of utmost importance, one cannot simply rely on the external, untrustworthy third-party application for secure communication [11]. Our chat application has up-to-date encryption algorithms along with our self-developed XOR encryption process ensuring utmost protection without using any kind of third-party material. For instance, during a hostage situation, the officers in command can safely plan a rescue operation along with their subordinates without sacrificing the vital game plan to anyone intercepting the conversation.

6.3 **Aatmanirbhar Samakraman***: File Auto-Synchronization App*

Using our encryption process, we have also developed an auto synchronization file app useful in situations where the user repeatedly stores important readings in the format of a file in a selected directory and wants to securely upload the same to a remote server.

The user first selects a specific directory to be continuously monitored by the app. He/she then selects another directory where he wants these files to be moved once uploaded to the server. He then presses the start syncing button to activate the background process (see Fig. 10).

Now every time a new file is stored in that specific directory it triggers the application and the file are automatically uploaded to the remote servers and the locally stored file is moved to the second directory. This transfer of the file to the server takes place using the Multi-Part technology after being encrypted with our XOR-encryption process.



Total Files Synced: 0 Files Total Size Sent: 0 KB

Logs of files sent will be maintained here

Now watching for new files at /storage/emulated/0/Documents/tifr
SERVER SELECTED: http://158.144.55.73:3478/upload/multipart

Fig. 10 Aatmanirbhar Samakraman android application

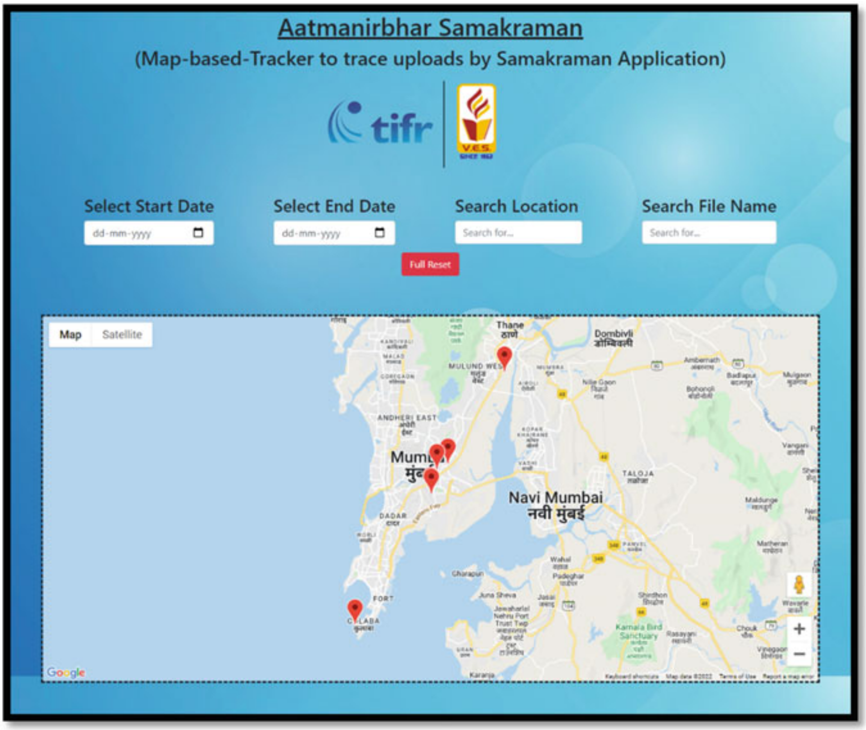
Along with the file, the current location coordinates of the user are also sent which is then used to display a tracking history on a web-based map UI for easy analysis: (http://file.aatmanirbhar-sanchar.live/) (see Fig. 11).

7 Results

In the end, we were successful in developing a secure, private, ephemeral chat application and deployed it on a private server hosted at our college, the VESIT campus. The web application is completely free from third-party services and is fully built upon open-source libraries.

We also developed an encryption system from scratch. This XOR encryption process is nothing but dividing the binary data into matrices followed by shuffling and reshuffling of the data and finally the data being XOR'd by a predefined key (see Fig. 5).

Finally, the output generated from the XOR method is passed to the Advanced Encryption Standard (AES-256) Algorithm in turn ensuring utmost security.



Aatmanirbhar Samakraman

(Map-based-Tracker to trace uploads by Samakraman Application)

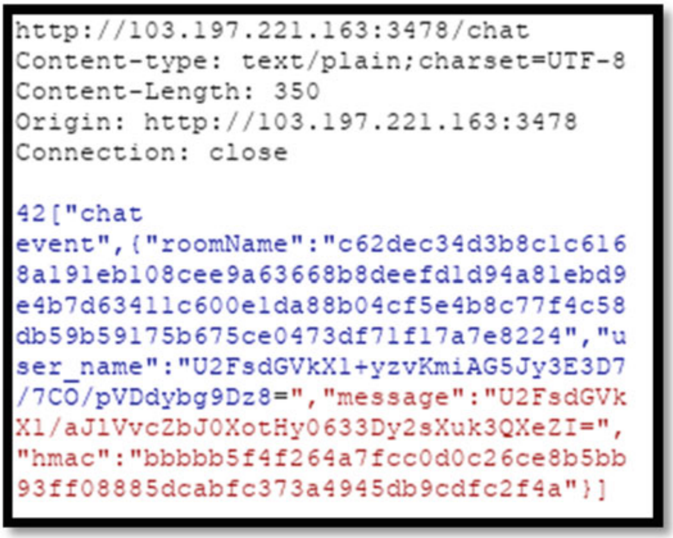
Fig. 11 Aatmanirbhar Samakraman live-map tracking

Finally, to ensure the integrity and authenticity of the transmitted message, the process of HMAC (hash-based message authentication code) verification is also practiced which guarantees tamper-proof messaging. To test the security of our product, we have tried the following Cryptanalysis techniques:

1. Snooping
   a. An intruder listens to traffic between two machines on a network in a snooping attack. We prevent this attack by only transmitting everything in an encrypted format. To an outsider, everything will look like gibberish (see Fig. 12).

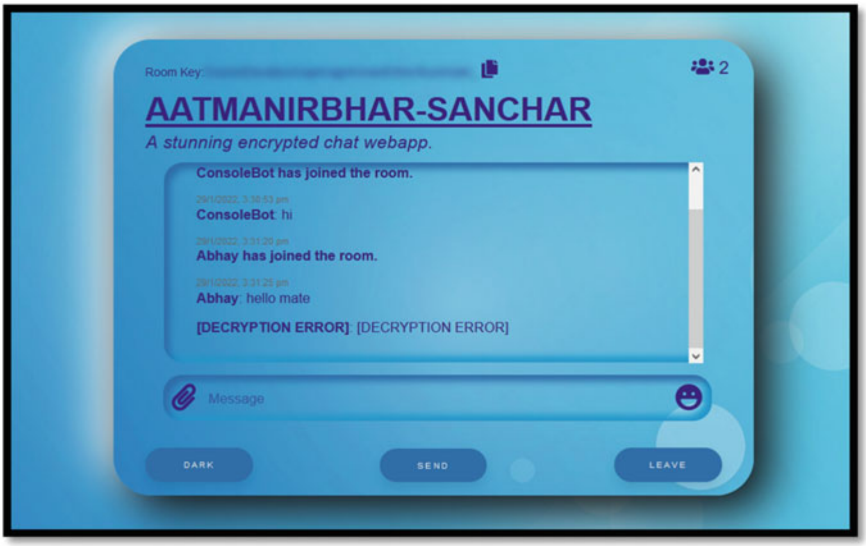
2. Man in the middle attack (MITM)
   a. An attacker intercepts a message/key sent between two communicating parties through a secure channel in this sort of attack and tries to alter them. We can detect any tampering in the incoming messages due to the HMAC verification process. If any alterations are detected, a "decryption error" is shown to all the users indicating a tampered message (see Fig. 13).

3. Server attack
   a. In the case wherein an intruder has successfully gained access to the remote server would cause no harm to the privacy of the users. This is achieved due



```
http://103.197.221.163:3478/chat
Content-type: text/plain;charset=UTF-8
Content-Length: 350
Origin: http://103.197.221.163:3478
Connection: close

42["chat
event",{"roomName":"c62dec34d3b8c1c616
8a191eb108cee9a63668b8deefd1d94a81ebd9
e4b7d63411c600e1da88b04cf5e4b8c77f4c58
db59b59175b675ce0473df71f17a7e8224","u
ser_name":"U2FsdGVkX1+yzvKmiAG5Jy3E3D7
/7CO/pVDdybg9Dz8=","message":"U2FsdGVk
X1/aJ1VvcZbJ0XotHy0633Dy2sXuk3QXeZI=",
"hmac":"bbbbb5f4f264a7fcc0d0c26ce8b5bb
93ff08885dcabfc373a4945db9cdfc2f4a"}]
```

Fig. 12 Intercepted message as visible to an intruder



Room Key: [redacted]

2

AATMANIRBHAR-SANCHAR

A stunning encrypted chat webapp.

ConsoleBot has joined the room.

29/1/2022, 3:30:53 pm

ConsoleBot: hi

29/1/2022, 3:31:20 pm

Abhay has joined the room.

29/1/2022, 3:31:25 pm

Abhay: hello mate

[DECRYPTION ERROR]: [DECRYPTION ERROR]

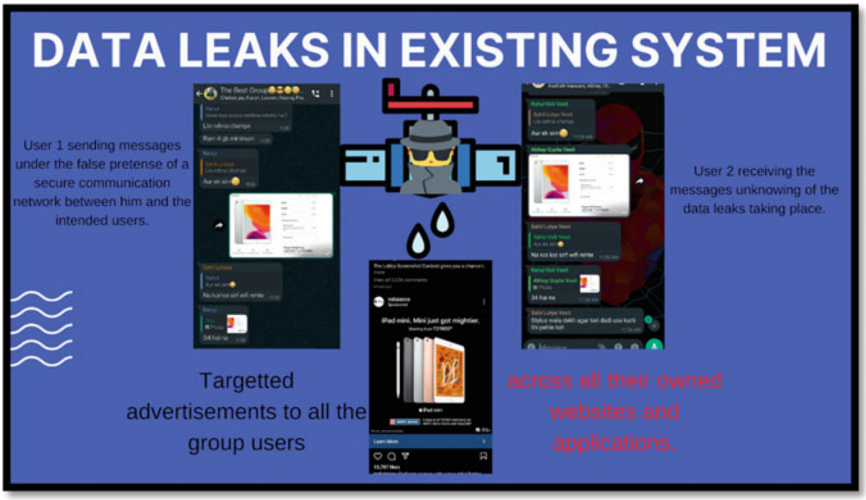
Fig. 13 Error is shown indicating message tamper

to the fact that no vital information is stored or even transmitted to the server without it being dually encrypted. So, the worse this intruder can do is shut down the server itself.

4. TCP-SYN flood attack
   a. TCP-SYN flood is a type of DDOS attack where the attacker starts pinging the server continuously from several different IP addresses by not providing Full information which is required by the server. Due to this, the server has to disconnect the running applications and wait for the partially opened connection started by the mugger, which can take enough resources to render the system unusable to authorized congestion.

7.1 Comparison of Results with Existing Systems

One of the major differences between Aatmanirbhar Sanchar and other similar applications is the promise of keeping your data safe and keeping the application open source for anyone to verify its' contents. The majority of the existing chat applications keep their source code proprietary and hide data mining loopholes in their terms and conditions (Fig. 14.).



DATA LEAKS IN EXISTING SYSTEM

User 1 sending messages under the false pretense of a secure communication network between him and the intended users.

User 2 receiving the messages unknowing of the data leaks taking place.

Targetted advertisements to all the group users

across all their owned websites and applications.

Fig. 14 Data leaks in existing software

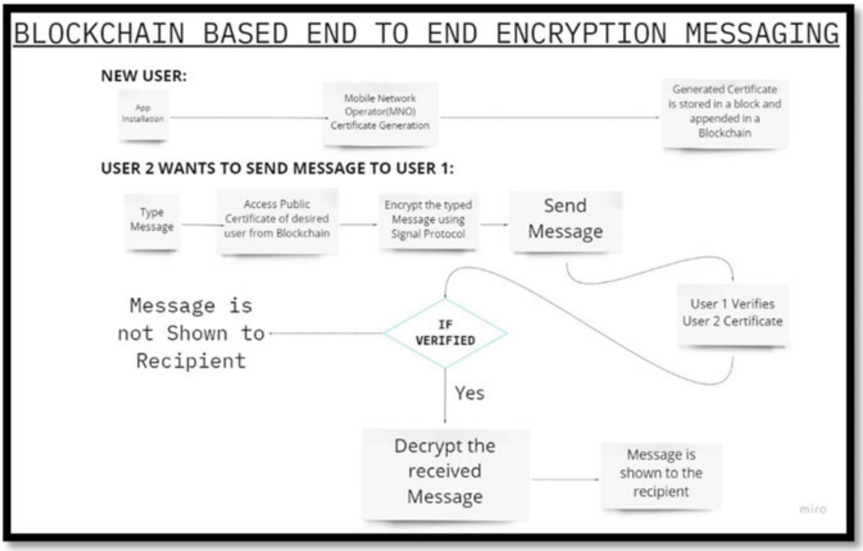

BLOCKCHAIN BASED END TO END ENCRYPTION MESSAGING
NEW USER:
App Installation
Mobile Network Operator(MNO) Certificate Generation
Generated Certificate is stored in a block and appended in a Blockchain
USER 2 WANTS TO SEND MESSAGE TO USER 1:
Type Message
Access Public Certificate of desired user from Blockchain
Encrypt the typed Message using Signal Protocol
Send Message
User 1 Verifies User 2 Certificate
Message is not Shown to Recipient
IF VERIFIED
Yes
Decrypt the received Message
Message is shown to the recipient
miro

Fig. 15 Decentralized chat application

8 Future Scope

8.1 Converting the Application to a Decentralized WEB3 Application (See Fig. 15)

Now, instead of a centralized server, we can also theoretically use a blockchain network to convert the application into a web3 decentralized application.

Blockchain is a type of database. In this database, data is stored in the form of blocks and these blocks are chained together to form a blockchain. When each block in the chain is added to the chain, it is given a precise timestamp.

It is extremely difficult to modify the contents of a block after it has been put into the blockchain. This is because each block has its own hash in addition to the previous block's hash.

So, in this, if a new user installs the app for the first time a certificate will be generated which is then stored in a block and appended to the blockchain. This certificate will be used as a public key to the blockchain system.

If a user wants to send a message to a particular person the certificate of the recipient is accessed from the blockchain. The message will be encrypted using the signal protocol before being sent. Now, user 1 verifies user 2's certificate, and once verified, user 2 will be allowed to decrypt the message otherwise an error should be thrown.

9 Conclusion

Secure and Private communication is a serious issue in today's world. One is not able to communicate with their loved ones without being spied upon by the "Mark Zuckerbergs" of the world. There is a serious lack of an indie Secure Communication application that can be freely and securely used by public and private organizations.

Hence, we are very excited to present to the world, "Aatmanirbhar Sanchar: An Ephemeral, Anonymous, Secure Chat Application" based on a custom-based encryption algorithm that can be easily hosted on one's very own private servers without any external eyes watching over.

Conflict of Interest Statement On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Cohn-Gordon K, Cremers C, Dowling B, Garratt L, Stebila D (2020) A formal security analysis of the signal messaging protocol. J Cryptol 33(4):1914–1983. https://doi.org/10.1007/s00145-020-09360-1
2. Singh R, Tewari H (2021) Blockchain-enabled end-to-end encryption for instant messaging applications. https://arxiv.org/abs/2104.08494
3. Botha JG, Van 't Wout MC, Leenen L (2019) A comparison of chat applications in terms of security and privacy. In: 18th European conference on cyber warfare and security. University of Coimbra, Portugal
4. Sabah N, Kadhim JM, Dhannoon BN (2017) Developing an end-to-end secure chat application. Int J Comput Sci Netw Secur
5. Burak M (2021) Encryption methods and comparison of popular chat applications. Adv Artif Intell Res 52–59
6. Emura K, Kajita K, Nojima R, Ogawa K, Ohtake G (2022) Membership privacy for asynchronous group messaging. National Institute of Information and Communications Technology (NICT), Japan.
7. Canetti R, Jain P, Swanberg M, Varia M (2022) Membership privacy for asynchronous group messaging. National Institute of Information and Communications Technology (NICT), Japan
8. Marlinspike M, Perrin T (2016) The X3DH key agreement protocol. Signal
9. Whatsapp whitepaper (2021) WhatsApp encryption overview. Whatsapp. https://www.whatsapp.com/security/WhatsApp-Security-Whitepaper.pdf
10. Perrin T, Marlinspike M (2016) The double ratchet algorithm. Signal
11. Daemen J, Rijmen V (1999) AES proposal: Rijndael. Rijndael Block Cipher
12. Bellare M, Canetti R, Krawczyk H (1996) Message authentication using hash functions—the HMAC construction. RSA Lab CryptoBytes
13. Hess A (2015) Encryption and cyber security for mobile electronic communication devices. Fed Bur Inv

A Meta Heuristics SMO-SA Hybrid Approach for Resource Provisioning in Cloud Computing Framework

Archana and Narander Kumar

Abstract Cloud computing is an up-to-date model for distributing information processing utility and provides a large amount of resources through the internet. The major challenges affecting a cloud computing environment include resource provisioning and security. In this paper, we focused on resource provisioning mechanisms using Meta-heuristics techniques such as spider monkey optimization (SMO) and simulated annealing (SA). A simulated annealing algorithm helps to give a fine solution along with statistical promises for uncovering the best solution, yet it cannot notify whether the best solution is found. So it requires another method to overcome this drawback. This paper presents the Spider Monkey Optimization algorithm with Simulated Annealing (SMO-SA) to generate the best fitness value possible. The aim of the proposed hybrid algorithm is to generate the minimum fitness value by combining spider monkey optimization with simulated annealing to provision the resources dynamically. This paper also presents the step-by-step mathematical working of our proposed hybrid algorithm by applying it to the relevant data set and calculating the speedup factor as well as mean square error (MSE) value along with fitness value, which shows the effective impact of our proposed SMO-SA algorithm.

Keywords Cloud environment · Spider monkey optimization · Simulated annealing · Resource provisioning · Meta-heuristics · Fitness value · Speedup factor · Mean square error

Archana · N. Kumar (✉)
Department of Computer Science, Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow, UP, India
e-mail: nk_iet@yahoo.co.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_42

1 Introduction

Nowadays, the cloud environment is a progressive conception that provides greater utilization of numerous assets, and it involves cloud service providers to furnish the assets to a cloud customer [1–3]. So the major concern is to provide resources appropriately. Resource provisioning is a mechanism that plays a challenging task in the cloud computing environment [4, 5]. Here some new search nature-inspired meta-heuristics techniques are available, which can use to provide the resources appropriately.

1.1 Spider Monkey Optimization

Swarm intelligence includes many optimization techniques, and its recently developed and popular technique is SMO, which establishes its place in the group of optimization techniques. SMO is categorized as a nature-inspired stochastic optimization method formed on the conception of fission–fusion social structure (FFSS). It is similar to other swarm intelligence techniques where each swarm updates its position by information sharing and continuous learning. SMO generates fitness value for each spider monkey to show how near the food source is. The global head is considered the finest solution of the entire group, and the local head represents each group's best solution. It involves six stages: local head stage, global head, global head learning stage, local head learning stage, local head resolution stage, and global head resolution stage [6, 7].

1.2 Simulated Annealing

An Annealing is a procedure in statistical mechanics where the system is gradually cooled to reach a position of low energy where they are very strong. Simulated annealing is based on the concept of annealing, and Kirkpatrick developed it in 1983. This algorithm has a greater impact on handling optimization problems using the annealing technique. We used an objective function in this method and attempted to minimize the solution. Here calculating the probability that the metal will leap to a larger energy level is given by

$$S = e^{(-\Delta c/T)} \tag{1}$$

Δc = change value of the objective function
$T = temperature.$
In the SA procedure, randomly generate the initial solution.

The system's temperature plays a similar effect as the temperature of the physical annealing process. SA may be at the starting phase, most worsening moves select but only improving ones are allowed [8].

By integrating the features of the SMO and SA approaches, this paper proposed a hybrid method (SMO-SA) to efficiently perform resource provisioning in the cloud environment.

2 Related Work

A maximum VM placement with minimum power consumption (MVMP) technique is proposed to increase earned benefits through cloud server provisioning. This technique also attempts to reduce the power budget. Here simulated annealing approach is employed to resolve a bi-objective optimization problem [9].

A hybrid algorithm of the meta-heuristic approach like Ant Colony optimization and simulated annealing is proposed for dynamic provisioning resources. This proposed ACO-SA algorithm is presented in multitier applications to schedule jobs with reduced costs [10].

A three-layered model is proposed to minimize the consumers' load power generation system based on the framework of the cloud environment [11]. A modified PSO algorithm is proposed to improve resource utilization. This proposed EPSO technique also maximizes the quality parameters and reduces the cost. This simulated annealing was incorporated with PSO to overcome the optimal local issue [12].

An artificial bee colony and ACO algorithm are used for dynamic resource provisioning. This is proposed to minimize the time optimization of provisioning and minimize the number of resources in multitier clouds. The result showed that ACO worked faster than other meta-heuristic algorithms [13].

A constrained continuous optimization problem solved by proposed modified version of SMO method named as constrained spider monkey optimization algorithm. This algorithm also contrasts with other constrained versions of the artificial bee colony, particle swarm optimization, and differential evolution [14].

A hybrid algorithm of meta-heuristic methods like particle swarm optimization and simulated annealing for provisioning the resources in the multitier applications. Here also showed the simulation results to compare the PSO-SA algorithm with the PSO and SA algorithm in multitier applications [15]. The experimental results showed better performance in comparison to other latest advanced algorithms. An Ageist Spider Monkey Optimization (ASMO) method is also proposed as a new form of SMO [16].

A detailed review analysis of SMO is presented to show the effective impact of SMO variants in real-world optimization problems [17]. Improved spider monkey optimization (ISMO) technique is proposed to improve the charge of convergence. This proposed approach is used in local leader stage to update the solution location [18].

A modified spider monkey optimization method is proposed to refine the convergence rate. In this algorithm global search capability improved by using the metropolis basis from simulated annealing [19]. A SMO is proposed for optimized resource allocation and energy utilization, Green Cloud scheduling model is suggested. The proposed approach was tested on different parameters by performing cloud simulation. The simulated results showed the proposed approach's effective energy consumption, response time and resource utility [20].

An ARPS (Autonomic resource provisioning and scheduling) framework is designed to encounter resource allocation problems. This framework is also joined with Spider money optimization to resolve a multi-optimization problem. Simulation results presented a better outcome of the proposed method in a cloud computing environment [21, 22].

A Three-layered model is proposed to minimize the cloud and fog framework's generating system power and consumer load. Analysis proves that PSO-SA performs stronger than RR [23]. A hybrid approach with cat swarm optimization and simulated annealing (CSM-CSOSA) has been proposed to schedule the resources dynamically [24].

3 Proposed Hybrid Algorithm (SMO-SA)

We suggested a hybrid method of SMO and Simulated Annealing (SMO-SA) for provisioning the resources appropriately. SMO use as a local and global seeking, and after finding gBest and lBest we employ SA to find all over gBest. Specifically, lBest and gBest have more chances to improve fitness value. The numerical working is given to state the working of the proposed SMO-SA algorithm. In this proposed algorithm, we used the following mathematical formulations to perform this Proposed SMO-SA, which are as follows.

3.1 Mathematical Formulation

Population initialization

$$popinit = rand(lower\ limit, upper\ limit) \tag{2}$$

By Eq. (2), here randomly initialize the population to perform the proposed SMO-SA algorithm. $popinit$ represents the initialization of the population.

Location update in local head stage

Perturbation rate. It requires updating the monkeys' position in the local leader phase. If the randomly generated number is greater or equal to the selected perturbation rate, then update the position, otherwise, accept the previous one.

$$perturbation\ rate \in (0.1, 0.9) \tag{3}$$

$$\begin{aligned}\text{SM}\ new_{ij} = SM_{ij} &+ \text{rdno}(0, 1) \times (LH_{kj} - SM_{ij}) \\ &+ \text{rdno}(-1, 1) \times (SM_{rj} - SM_{ij})\end{aligned} \tag{4}$$

Here SM new_{ij} represents the new position of the monkey. SM_{ij} shows the jth dimension of the spider monkey (SM), rdno represents the random number between the range, LH_{kj} shows the local leader position from the jth dimension of the kth group, SM_{rj} represents the randomly selected SM from the kth group.

Function value. In this proposed algorithm following equation helps to calculate the function value.

$$function\ value\ (fv) = x1^2 + x2^2 \tag{5}$$

Here $x1^2$, $x2^2$ represents the dimensions of each monkey.

Fitness value

$$\text{fitv} = \begin{Bmatrix} \frac{1}{1+\text{fv}_\text{i}}, & \text{if } \text{fv}_\text{i} \geq 0 \\ 1 + \text{absfv}_\text{i}, & \text{if } \text{fv}_\text{i} < 0 \end{Bmatrix} \tag{6}$$

By using Eq. (6) calculate the fitness value of each monkey fitv shows the fitness value and fv_i shows the function value of each monkey.

Location update in global head stage

Probability. Probability requires updating the position of the global head stage. If the randomly selected number is less than the probability, then update the SM position otherwise, consider the previous one.

$$probability\ (P) = 0.9 * \left(\frac{fitv_i}{maxfit}\right) + 0.1 \tag{7}$$

Here $fitv_i$ shows the fitness value of each monkey and $maxfit$ represents the better fitness of among all monkeys.

Metropolis criterion

$$\Delta c = fitv_{new} - fitv_{previous} \tag{8}$$

$$T = \sum \frac{fitv}{SM} \tag{9}$$

$$S = e^{(-\Delta c \div T)} \tag{10}$$

Here S represents the Metropolis criterion. Δc shows the positive change in fitness value and $fitv_{new}, fitv_{previous}$ represent the calculated new fitness value and previous fitness value. T shows the temperature value, which minimizes at the end. $fitv$ Shows the fitness value of all monkeys, and SM shows the total number of monkeys. If a new solution is not better than the previous one, then apply this criterion to decide whether to accept or reject the solution in this proposed algorithm. If the random generated number is less than S, then accept the point, otherwise refuse the point.

Reduction factor and random solution. In this proposed SMO-SA algorithm randomly generate the random and reduction factor.

3.2 Algorithm of Proposed Hybrid Approach (SMO-SA)

The flow chart of proposed hybrid (SMO-SA) algorithm is shown in Fig. 1.

1. Initialize population and perturbation rate.
2. Evaluate position, function value, and fitness value.
3. Pick out global and local heads.
4. Location update by local head stage.
5. SM $new_{ij} = SM_{ij} +$ rdno(0, 1) $\times \left(LH_{kj} - SM_{ij}\right) +$ rdno(−1, 1) $\times \left(SM_{rj} - SM_{ij}\right)$
6. Location update by global head stage
7. SM $new_{ii} = SM_{ij} +$ rdno(0, 1) $\times (GH_j - SM_{ij}) +$ rdno(−1, 1) $\times \left(SM_{rj} - SM_{ij}\right)$
8. Now applying SA on gBest and lBest value to search around the gBest and generate reduction factor(c) and calculate the temperature
9. $T = \sum \frac{fitv}{SM}$
10. Generate a new random solution and calculate the objective fitness value
11. If new solution < previous solution
12. Then accept a new point
13. Otherwise, apply the metropolis criterion
14. (to accept or reject the current point)
15. S $= e^{(-\Delta c/T)}$
16. If random value <S
17. Then accept a new point
18. If $\Delta c\ is\ negative$
19. We accept the current point and increase the iteration
20. j = j + 1, check if the end of the iteration
21. Stop
22. Otherwise, go to step 3.
23. Check if gBest needs to migrate and get the output.

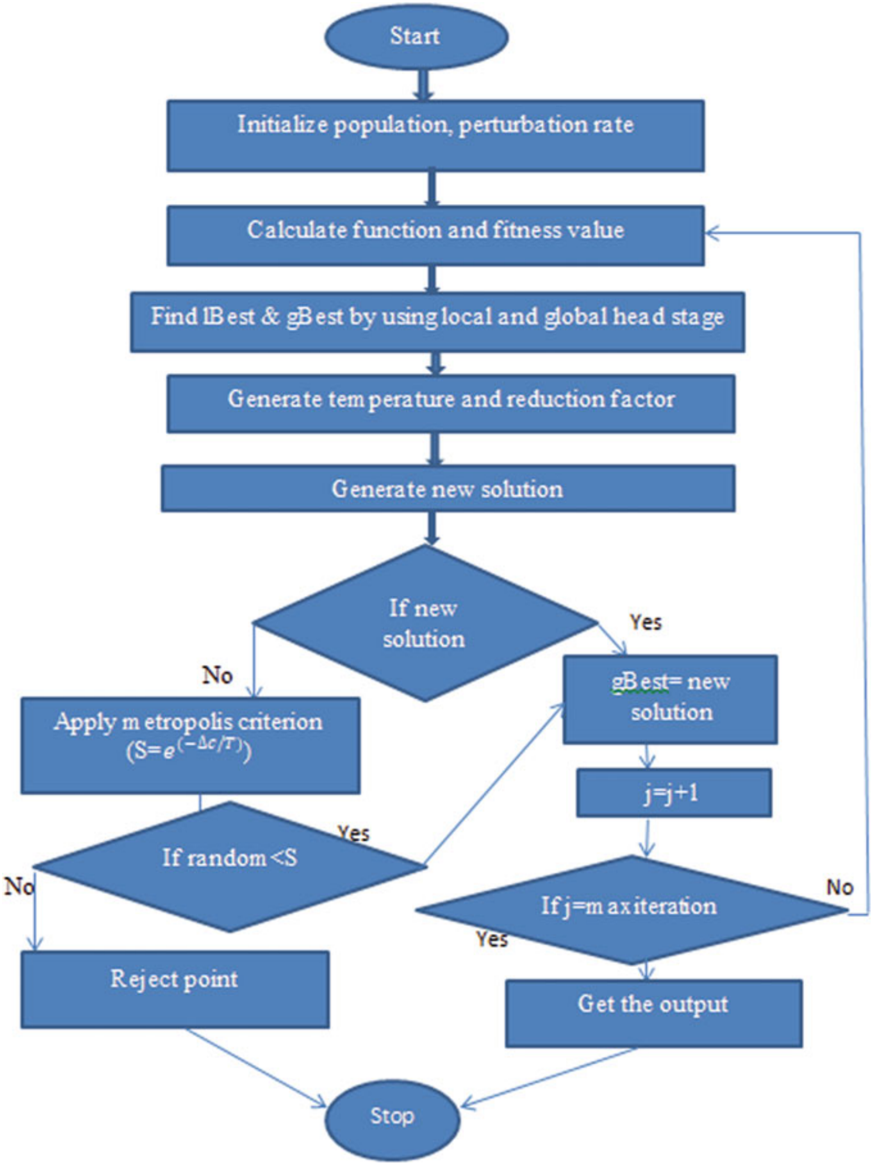

Start
Initialize population, perturbation rate
Calculate function and fitness value
Find lBest & gBest by using local and global head stage
Generate temperature and reduction factor
Generate new solution
If new solution
Yes
No
gBest= new solution
Apply metropolis criterion
(S=$e^{(-\Delta c/T)}$)
j=j+1
If random <S
Yes
No
If j=maxiteration
No
Yes
Reject point
Get the output
Stop

Fig. 1 Flow chart of proposed hybrid (SMO-SA) algorithm

4 Working Example of the Proposed Hybrid Algorithm (SMO-SA)

Here we showed the step-by-step working of our proposed hybrid algorithm SMO-SA on data set [25].

Objective function = $x_1^2 + x_2^2$.
Population size (N) = 10
Dimensions of each monkey (D) = 2
Max. no. of groups = N/10 = 10/10 = 1
Global head limit $\in \left\{\frac{N}{2}, 2N\right\} = \{5, 20\}$ let global head limit = 10
Local head limit = 2 × N = 2 × 10 = 20
pr ∈ (0.1, 0.9) let pr = 0.6.
Table 1 shows the initialization of parameters.

Now calculate the fitness function

if $fv_i(x) \geq 0$, then $\frac{1}{1+fv_i(x)}$.
if $fv_i(x) < 0$.
then $1 + abs\ fv_i(x)$.

Tables 2 and 3 show the calculation of the fitness function. Here minimum fitness or best fitness value is 0.0016, which is SM 3, position (14.4, 20).

Location updated stage

Local head stage

Table 1 Initialization

SM No.	x_1	x_2
T_1	10	4.5
T_2	4	7
T_3	14.4	20
T_4	9	8
T_5	14	5
T_6	2.25	2
T_7	3.5	7.2
T_8	2.5	1.12
T_9	1	1.75
T_{10}	3.6	8.33

Table 2 Calculated function value

SM No.	1	2	3	4	5	6	7	8	9	10
$fv_i(x)$	120.25	65	607.36	145	221	9.06	64.09	7.5044	4.06	82.34

Table 3 Calculated fitness value

SM No.	1	2	3	4	5	6	7	8	9	10
fitv(x)	0.0082	0.0151	0.0016	0.0068	0.0045	0.0994	0.0153	0.1175	0.1976	0.01119

$$\text{SM new}_{ij} = \text{SM}_{ij} + \text{rdno}(0, 1) \times \left(\text{LH}_{kj} - \text{SM}_{ij}\right) + \text{rdno}(-1, 1) \times \left(\text{SM}_{rj} - \text{SM}_{ij}\right)$$

Updating 1st SM ($i = 1$)
$J = 1 = x_1$ generate random no. rdno (0, 1) let rdno = 0.6
Since pr = $0.6 \leq 0.6$ (true)
So SM new = $10 + 0.6 \times (14.4 - 10) + (-0.7) \times (14 - 10)$
$x_1 = 9.84$
$J = 2 = x_2$ generate random no. rdno (0, 1) let rdno = 0.3
Since pr = $0.6 \leq 0.3$ (False).
Therefore SM new = SM_{ij} (previous value), $x_2 = 4.5$
So new solution = (9.84, 4.5)
Calculating function value and fitness value of SM new_1
$f_1(\text{SM new}_1) = 117.0755$
$\text{fit}(\text{SM new}_1) = 0.0084$
Applying greedy selection between SM new_1 and SM_1 based on fitness
0.0084<>0.0082
So there is no need to update, $\text{SM}_1 = (10, 4.5)$.
The updated solution is shown in Table 4.
Here local best and global best = 0.0016.
Now calculate the probability (P)

$$P = 0.9 * \left(\frac{fitv_i}{maxfit}\right) + 0.1$$

Table 4 Updated solutions

SM number	Updated dimension j	SM_{new}	f_i (x)	Fit
1	–	10, 4.5	120.25	0.0082
2	–	4, 7	65	0.0151
3	–	14.4, 20	607.36	0.0016
4	–	9, 8	145	0.0068
5	–	14, 5	221	0.0045
6	1	3.78, 2	18.28	0.0518
7	–	3.5, 7.2	64.09	0.0153
8	–	2.5, 1.12	7.5044	0.1175
9	2	1, 10.42	109.74	0.0090
10	–	3.6, 8.33	82.34	0.0119

The calculated probability list is shown in Table 5.

Global head stage

$$\text{SM new}_{ii} = \text{SM}_{ij} + \text{rdno}(0, 1) \times \left(\text{GH}_j - \text{SM}_{ij}\right) + \text{rdno}(-1, 1) \times (\text{SM}_{rj} - \text{SM}_{ij})$$

And here, select the j value randomly.
Updating SM 1
Random j = 1 = x_1
Rdno = 0.6 < probability (4.7125) yes
So here needs to update
SM new = 10 + 0.6 × (14.4–10) + (-0.6) × (14–10) = 10.24
So new solution is = (10.24, 4.5)
$fv_1(\text{SM new}_1) = 125.1076$
$\text{fitv}(\text{SM new}_1) = 0.0079$.
Applying greedy selection between SM new_1 and SM_1 based on fitness
0.0079 < 0.0082.

Here new fitness is better than previous fitness, so there is a need to update the position.

In Table 6, we can see at 1st round, a total of 10 SM updated their position, and modification should be equal to the total number of SM. So there is no need for a second round because 10 improvements have been done and a total of 10 SM.

In the above step gBest and lBest = 0.0015 with position (16.02, 20). Now, this gBest value forward to the next step, which is simulated annealing.

Table 5 Calculated probability list

SM no	1	2	3	4	5	6	7	8	9	10
probability	4.7125	8.5937	1	3.925	2.63	29.23	8.706	66.19	5.1625	6.793

Table 6 Updated location

SM number	Updated dimension j	SM_{new}	fv_i (x)	Fitv
1	1	10.24, 4.5	125.10	0.0079
2	2	4, 17.9	336.41	0.0029
3	1	16.02, 20	656.64	0.0015
4	2	9, 11.9	222.61	0.0044
5	2	14,7.4	250.76	0.0039
6	1	9.14, 2	87.61	0.0112
7	1	7.79, 7.2	112.52	0.0088
8	2	2.5, 2.48	12.4	0.0746
9	1	8.59, 10.42	182.36	0.0054
10	1	5.67, 8.33	101.53	0.0097

Implement SA algorithm, C = 0.8, max iteration = 2.

$$\begin{aligned}&\text{Calculate temperature, T}\\&= \frac{\text{fit}(1)+\text{fit}(2)+\text{fit}(3)+\text{fit}(4)+\text{fit}(5)+\text{fit}(6)+\text{fit}(7)+\text{fit}(8)+\text{fit}(9)+\text{fit}(10)}{10}\\&= 0.1303.\end{aligned}$$

gBest = 0.0015 = lBest and fit = 0.0015 (from SMO).

Iteration = 1, rdno no. $u_1 = 0.8$, $u_2 = 0.9$

With ±6 accuracy, generate new design point: $x_1 = (-8, 17)$, $x_2 = (-6, 11)$.

$r_1 = -8 + 0.8*(17 - (-8)) = 12$, $r_2 = -6 + 0.9 * (11 + 6) = 9.6$.

$X_{new} = (12, 9.6)$.

Now calculate $fv_x = 236.16$, $fitv_x = 0.0042$

(here we can see $fitv_{new} > fitv_{previous}$, we have to check other conditions to accept or reject this point)

$$\Delta c = fitv_{new} - fitv_{previous} \tag{7}$$

Now applying the Metropolis criterion is needed to accept or reject the current point.

Choose random no. in the range (0, 1) r = 0.83.

Metropolis criterion, $P[X_2] = e^{(-\Delta c/T)} = 1.0209$.

Here r < 1.0209, so we can accept the new point = (12, 9.6) and current temp is high.

Iteration = 2.

gBest = lBest = 0.0042, $u_1 = 0.92$, $u_2 = 0.82$.

generate new design point: $x_1 = (-12, 20)$, $x_2 = (-9, 29)$.

$r_1 = -12 + 0.92 * (20 - (-12)) = 17.44$, $r_2 = -9 + 0.82 * (29 + 9) = 22.16$.

$X_{new} = (17.44, 22.16)$.

Now calculate $fv_x = 795.21$, $fitv_x = 0.0012$

(Here we can see $fitv_{new} < fitv_{previous}$), $\Delta c = fitv_{new} - fitv_{previous} = -0.0030$

(Here Δf value is negative, so there is no need for a metropolis criterion, we can accept the current point and increase the iteration to j = 3, as given max j = 2)

Stop the process.

5 Results and Discussion

We mathematically analyzed the working of the proposed hybrid SMO-SA algorithm on the data set [25]. Here we can see that our proposed hybrid algorithm helps to minimize the fitness value and to give the best fitness value by giving more chances to improve their lBest and gBest value. This proposed hybrid algorithm gives better results for resource provisioning. Here we also calculate the speedup factor and MSE

Table 7 Comparative analysis between SMO and proposed hybrid SMO-SA algorithm

Algorithm	Fitness value	Speedup factor	MSE
SMO	0.0015	1.9969	0.01303
SMO-SA	0.0012	1.9974	0.01300

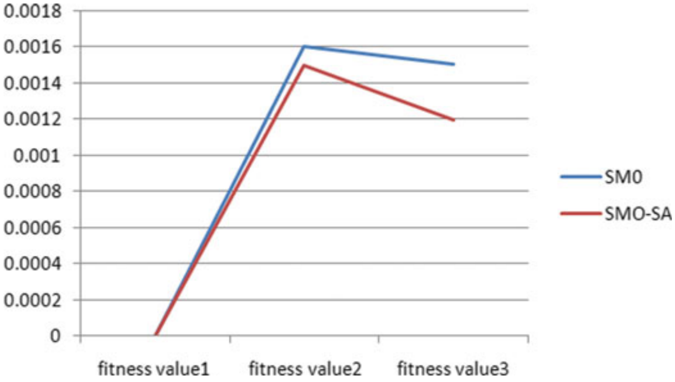


Fig. 2 Comparison between fitness value of SMO and proposed hybrid algorithm (SMO-SA)

value to show the better working impact of SMO-SA over SMO. Table 7 shows the comparative analysis between SMO and the proposed hybrid SMO-SA approach. Here 3 parameters have been calculated to compare the proposed approach with the existing one, i.e., fitness value, speedup factor, and mean square error (MSE). The calculated fitness value for the proposed SMO-SA algorithm is 0.0012, which optimizes the fitness value of SMO (0.0015). The speedup factor and MSE of the proposed SMO-SA algorithm are 1.9974 and 0.1300, respectively, while for SMO speedup factor is 1.9969, and MSE is 0.01303. The speedup factor of the proposed SMO-SA approach is maximized, and the MSE value is minimized, which shows the proposed approach's better working than SMO.

Figure 2 shows the graphical representation of the calculated fitness value of the proposed SMO-SA and also the existing one (SMO algorithm) as in Table 7. This figure analyzes the SMO-SA fitness value optimized compared to the SMO algorithm.

6 Conclusion

We suggested working in cloud computing to solve resource provisioning issues. A new hybrid method SMO-SA is proposed for resource provisioning based on the combination of spider monkey optimization with simulated annealing. In this paper, we mathematically show that our proposed method gives a better result. This proposed method shows that provisioning resources based on the SMO-SA algorithm help minimize the fitness value, which leads to keeping down the processing flow and resources cost. In this paper, mathematical proving is given to state the effective working of the proposed SMO-SA approach. Here three performance parameters are considered, i.e., fitness value, speedup factor, and MSE. There may provide the simulation results on this proposed algorithm for resource provisioning and consider execution time, price, and memory allowance as resources for future work.

References

1. Kumar TS (2019) Efficient resource allocation and QoS enhancements of IoT with fog network. J ISMAC 02:101–110
2. Chandy A (2019) Smart resource usage prediction using cloud computing for massive data processing systems. J Inf Technol Digit World 2:108–118
3. Srivastava P, Khan R (2018) A review paper on cloud computing. Int J Adv Res Comput Sci Softw Eng 8:17–20
4. Kumar N, Kumar S (2019) Resource management to virtual machine using branch and bound technique in cloud computing environment. Soft computing: theories and applications. Advances in intelligent systems and computing, vol 742. Springer, pp 365–373
5. Singh S, Chana I (2016) Cloud resource provisioning: survey, status and future research directions. Knowl Inf Syst 49(3):1005–1069
6. Sumalatha K, Anbarasi MS (2019) A review on various optimization techniques of resource provisioning in cloud computing. Int J Electr Comput Eng (IJECE) 9:629–634
7. Sharma H, Hazrati G, Bansal JC (2019) Spider monkey optimization algorithm. Evolutionary and swarm intelligence algorithms. Studies in computational intelligence, vol 779. Springer, pp 43–59
8. Dubey K, Sharma SC, Aida A (2020) A simulated annealing based energy-efficient VM placement policy in cloud computing. In: International conference on emerging trends in information technology and engineering (ic-ETITE). IEEE, pp 1–5
9. Addya KS, Kumar A, Sahoo B, Sarkar BKS (2017) Simulated annealing based VM placement strategy to maximize the profit for cloud service providers. Eng Sci Technol Int J 20:1249–1259
10. Leninfreda A, Dhanyab D, Kavithac S, Ashwini M (2019) Hybrid algorithm for resource provisioning with low cost and time using improved cost-based algorithm in hybrid cloud computing. J Intell Fuzzy Syst 1–10
11. Yasmeen A, Javaid N, Rehman O, Iftikhar H, Malik MF, Muhammad JF (2018) Efficient resource provisioning for smart buildings utilizing fog and cloud based environment. IEEE, pp 811–816
12. Mani K, Krishnan RM (2017) Flexible cost based cloud resource provisioning using enhanced PSO. Int J Comput Intell Res 13(6):1441–1453
13. Eawna MH, Hamdy S, EI-Horbaty EM (2015) New trends of resource provisioning in multi-tier cloud computing. In: Seventh international conference on intelligent computing and information systems (ICICIS'15). IEEE, pp 224–230
14. Gupta K, Deep K, Bansal JC (2017) Spider monkey optimization algorithm for constrained optimization problems. Soft Comput 21:6933–6962
15. Eawna MH, Mohammed SH (2015) Hybrid algorithm for resource provisioning of multi-tier cloud computing. In: International conference on communication, management and information technology (ICCMIT). Elsevier, pp 682–690
16. Sharma A, Sharma A, Panigrahi BK (2016) Ageist Spider Monkey Optimization algorithm. Swarm Evol Comput 1–23

17. Agarwal V, Rastogi R, Tiwari DC (2018) Spider Monkey Optimization: a survey. Int J Syst Assur Eng Manag 9:929–941
18. Swami V, Kumar S, Jain S (2018) An improved spider monkey optimization algorithm. Soft computing: theories and applications. Advances in intelligent systems and computing, vol 583. Springer, Singapore, pp 73–81
19. Hazratia G, Shannab H (2016) Modified spider monkey optimization. In: International workshop on computational intelligence (IWCI). IEEE, pp 209–214
20. Samriya JK, Kumar N (2022) Spider monkey optimization based energy-efficient resource allocation in cloud environment. Trends Sci 19(1):1–19
21. Kumar M, Kishor A, Abawajy J, Agarwal P (2022) ARPS: an autonomic resource provisioning and scheduling framework for cloud platforms. IEEE Trans Sustain Comput 7:386–399
22. Sharma Y, Taheri J (2020) Dynamic resource provisioning for sustainable cloud computing systems in the presence of correlated failures. IEEE Trans Sustain Comput (c) 1–13
23. Yashmeen A, Javaid N (2018) Resource provisioning for smart building utilizing fog and cloud based environment. In: 2018 14th international wireless communications & mobile computing conference (IWCMC). IEEE, pp 811–816
24. Gabi D, Ismail AB (2018) Hybrid Cat Swarm Optimization and simulated annealing for Dynamic task scheduling on cloud computing environment. J ICT 17(3):435–467
25. Kumar N, Kumar S (2018) Virtual machine placement using statistical mechanism in cloud computing environment. Int J Appl Evol Comput 9:23–31

A Comprehensive Study of Automation Using a WebApp Tool for Robot Framework

N. Alok Chakravarthy and Usha Padma

Abstract The procedure of manual testing takes a lot of time. Automation is particularly desired since testing process is also error-prone due to its repeated structure. Robot Framework is a flexible tool that makes use of the keyword driven testing methodology. It is straightforward to use when high-level keywords may be created from current keywords. Although a command line interface makes it easy to integrate new test libraries, it is also possible to create customised test libraries in Python or Java using a straightforward library API. All these capabilities guarantee that Robot Framework can be used to execute test cases in a timely manner. This paper explains how a WebApp tool could be used to quickly automate testing procedures by reducing expenses and raising the overall functionality of the software. Results comparison of manual testing with automation testing shows that automated tests run on average 80.46% faster than manual tests.

Keywords Software testing · Test cases · Test automation · Robot framework

1 Introduction

The typical approach used by businesses is manual testing. However, this is gradually disappearing due to the rise of quality tools and web-based apps. Automated testing has recently grown to be an essential component of the software development life cycle. The software product's functioning and all its components or use cases are tested using the test automation procedure. It offers an effective method for managing several product characteristics at once. There are several proprietary and open-source test automation resources available [1]. Many of the tools that are now present are best suited for unit tests carried either by programmers or quality analysts. It must also

N. Alok Chakravarthy (✉) · U. Padma
R. V. College of Engineering, Bengaluru, India
e-mail: alokcn.te18@rvce.edu.in

U. Padma
e-mail: kamakshimb@rvce.edu.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_43

offer several libraries for connecting to the back end of the device being tested, such as computer or mobile apps. Numerous testing tools offered by manufacturers are quite complex and employ open-source or closed-source programming language. Automating manual tests that already exist is comparable to a developer implementing a software in a coding language to automate any other manual operation [2].

2 Associated Works

Automation of robotic processes and acceptability testing are done using the Robot Framework. It is a platform-neutral programme with a developing network of libraries and outside tools. Selenium is a free and open-source automation tool that is quite well known right now. Selenium has several advantages for test automation. By permitting recording and playback, web apps may simply be evaluated. They can also execute different scripts across different browsers. There is a lot of difference between both. Robot is an amazing test framework, making use of test libraries both standard and external for executing tests. On the other hand, Selenium is just a WebDriver that would generally need some support from test automation runners for the test execution. Robot framework is used for performing all kinds of automated tests, whether they are related to UI Testing or API level testing. Selenium module facilitates the Selenium framework support in case of Robot.

Selenium is widely used for web app automation testing recently. Numerous programming languages, such as Java, Ruby, and Python, are supported by Selenium WebDriver. Both Selenium and Cucumber are powerful and comprehensive automation testing tools and are very useful for web application testing. But while Selenium is one of the best automation testing tools, using it can be a daunting task for a non-technical person. Therefore, more and more companies are using Selenium and Cucumber together for automation testing. There are several significant distinctions between cucumber and selenium. Cucumber is an automation tool for behaviour-driven programming, whereas selenium is an automation tool for web apps. Cucumber script development is also easier than selenium script generation. Selenium can function without cucumber also. For step-definition execution, cucumber relies on selenium.

Robot automation framework is best suited for web. The python test automation framework is the ideal option for test automation projects. Robot Framework can test MongoDB, FTP, Android, and Appium. It includes the Selenium WebDriver library among other helpful test libraries. To assist it would be as expandable as possible, it offers a variety of APIs. Integration testing is done in software testing while the software and hardware parts are integrated to look for errors in the test. All through the test procedure, it is a continual process. For a quick release, it is recommended to do integration testing. Network management includes elements like fault management, which controls how network element alarms are handled, as well as troubleshooting, log analysis, and responding to fault circumstances. Performance management, which assesses data collecting, reporting, data processing, and overall performance monitoring, is another component. Also, security management is for user and policy access control, along with logs authentication and authorization [3].

File handling refers to the method used by a software to save the relevant information and data in a file. To reuse this information in any software, it may be obtained or retrieved from these files. There are occasions when a program's outcome, after it has been compiled and executed, does not achieve the desired result. The output of the application is repeatedly examined in such circumstances. Now, it takes a lot of work for any coder to repeatedly compile and execute the same software. That is precisely why using file handling is beneficial. Few uses of file handling are reusability, saves time, commendable storage capacity, and portability.

To expand these tests and create less mistakes, a tool that was both simple to use and powerful was required. The tool would need to work on any system. User testing was performed on Windows and Linux, while server testing was done on Linux and Solaris. The tool was completely platform independent. The primary emphasis was on regression testing of effective and complete tests. Even though most of the tests had already been run at least previously, running regression analysis proved problematic. When several servers were deployed upon which tests were conducted, the problem became even more difficult due to their diverse configurations [4]. The basic method was for all test cases: setup environment, begin system tracking, execute test suite, terminate platform traceability, verify code traces, and review system traces. It was critical not to overlook the creation of a review at the end with survey results, which could take a significant time and resources because it is important to modify the list of test cases, label those that failed, and make some notes on why they failed, which can take a bit of time for more test cases [5].

The initial plan was to create a simple shell script that would run all the tests and evaluate the findings from log files. As a framework, keyword-driven testing, which allows test scripts to be executed at a higher level of abstraction, was examined. The concept of keyword driven evaluation is comparable to that of a service or subroutine in coding, in which the same code may be performed with varying parameters [6], making it an ideal candidate for the needed automation. Figure 1 depicts its incredibly modular architecture design.

After a thorough investigation, Robot Framework [7] was determined to suit all specifications. It was written in Python, which is available on all major resources. Therefore, all the conditions for 100% interoperability were satisfied. As indicated by the fact that it is referenced on [8], Robot Framework looks to be one of the rare free software tools that allows a multi-platform system and is consistently updated. Robot Framework is a multi-purpose framework that may be used with any software. The test results are given in a straightforward, customizable tabulated form. When initiated, Robot Framework analyses the findings, performs the test cases, and generates reports and outcomes. Test modules control the interaction with the topic under test. Modules can use programme interfaces directly or use drivers, which are subordinate test mechanisms [9]. What was required was a graphical user interface (GUI) that made generating and modifying test cases straightforward. Its primary aim is to provide a simple interface for the establishment and improvement of test

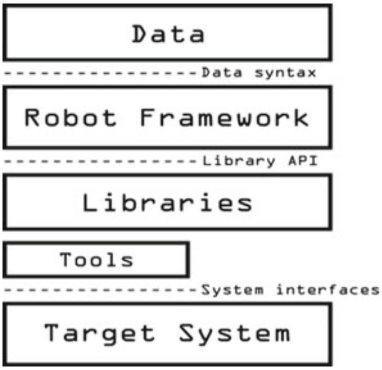


Fig. 1 Architecture of robot framework

data in the Robot Framework [10]. Test cases are produced from HTML pages and use keywords from test suites to control the product under test, whereas test suites are created from files to make it simple to store into any version of system. It can introduce additional, more precise keywords by combining and arranging the existing ones.

It is simple to automate web applications, which entails using numerous software automation testing techniques to gauge a website's performance. A procedure called automated web application testing employs several software tools to rate a website's functionality. The running of tests and comparison of actual test results with anticipated or projected outcomes are made possible using efficient automated testing tools for web application testing.

Web application automation testing is also lacking on websites. Therefore, companies of today should employ test automation for web apps to guarantee that the websites function properly and provide perfect performance to consumers. Benefits of Web Application Automated Testing are test coverage improves, ensures quicker debugging, effective test results, increased testing speed, rapid feedback, early bug identification. The test data for the Robot Framework is tabular and can be formatted in HTML, TSV or Text. When the Robot Framework is launched, it chooses a test data file converter based on the file extension and processes the test file. HTML files are typically utilised. Four different forms of file structure are gathered as test data files, as shown in Table 1.

Table 1 File structure of test data

Test name	Used for
Settings	(1) Adding test libraries, resource files, and variable files to the system (2) Specifying meta information for test cases and test suites
Variables	Defining an all-purpose variable for the test data
Test cases	Making test cases out of the provided keywords
Keywords	Generating user keywords from lower-level keywords already in use

3 Proposed Work

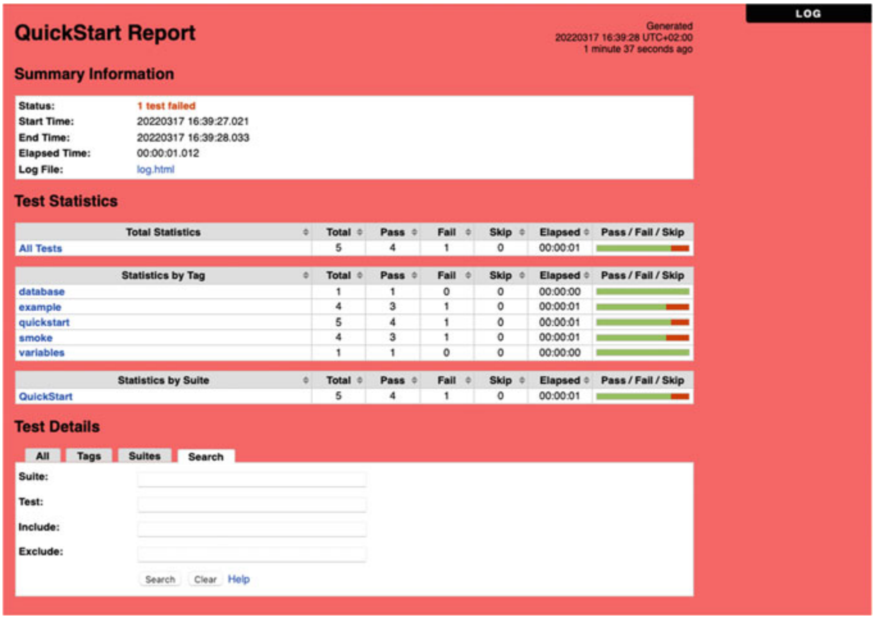
3.1 *Test Suite Creation*

Making programs with an instantaneous return is one technique to reduce errors that occur when new tool utilization is initiated [11]. Developing scripts that won't take too long to write but will undoubtedly reduce time spent on manual testing. More importantly, by writing the programs, users will have a deeper understanding of the tool's operation and be able to come up with even better codes. Since these programs have previously provided some benefit, little is lost if they are destroyed. Robot Framework may be thought of as a script because it is built on keywords, and a keyword conjunction can create a new user keyword. Robot Framework comes with a few predefined modules, but as it is a Python-based tool, it is simple to add libraries created in Python or Java. All users must do is create their individual function and return a value. A feature of RIDE called keyword completion displays the keywords that are identified in the test suite, the source that is being modified, the imported files, or modules. Additionally, parameters are checked continuously for all recognised keywords [12]. One of the characteristics of the Robot Framework is detailed keywords. Since RIDE can construct keywords, it is sufficient to characterize the test case first before creating the keywords and populating them with actions. Another scenario is learning that a particular sequence must be utilised frequently. In that scenario, the sequence can be grouped and designated as a new keyword. RIDE makes it simple to do this; all that is required is to mark the series, at which point RIDE will extract the lines and, if necessary, automatically construct a new keyword. RIDE will update the series and modify the test case after establishing a new keyword. It is possible to preserve keywords and variable definitions in source files so that they may be utilised across many suites. If the keyword is applicable to more tests, moving it to a common resource is a smart idea [13]. In this approach, redundant labour is prevented, and those keywords may be utilised in future testing.

3.2 *Test Case Execution*

Although it can run an entire test suite or just a few selected test cases from the RIDE GUI, it is essential to execute test cases from the command line so that their processing may be easily automated, for instance from a continuous integration server. This is often carried out in this manner as Robot Framework is a command-line programme. Different switches can be utilised in this manner. The specification of the key test cases is one of several things that may be stated. All important test cases must pass for the test suite to be successfully completed [14]. The background colour unquestionably indicates whether the entire test suite was successfully completed.

Specification of important test cases must be done with discretion. Irrespective of the outcomes of other test cases, if important test cases pass correctly, the report



QuickStart Report

Generated
20220317 16:39:28 UTC+02:00
1 minute 37 seconds ago

Summary Information

Status:	1 test failed
Start Time:	20220317 16:39:27.021
End Time:	20220317 16:39:28.033
Elapsed Time:	00:00:01.012
Log File:	log.html

Test Statistics

Total Statistics	Total	Pass	Fail	Skip	Elapsed	Pass / Fail / Skip
All Tests	5	4	1	0	00:00:01	🟩🟥

Statistics by Tag	Total	Pass	Fail	Skip	Elapsed	Pass / Fail / Skip
database	1	1	0	0	00:00:00	🟩
example	4	3	1	0	00:00:01	🟩🟥
quickstart	5	4	1	0	00:00:01	🟩🟥
smoke	4	3	1	0	00:00:01	🟩🟥
variables	1	1	0	0	00:00:00	🟩

Statistics by Suite	Total	Pass	Fail	Skip	Elapsed	Pass / Fail / Skip
QuickStart	5	4	1	0	00:00:01	🟩🟥

Test Details

Suite:

Test:

Include:

Exclude:

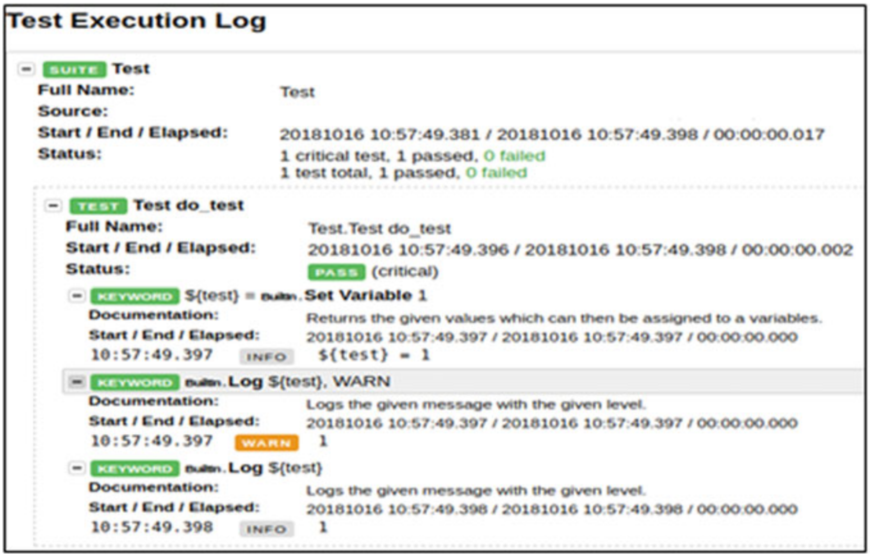
Fig. 2 Test case report file

will be deemed acceptable. Statistics, on the other hand, will list and display the set of test cases that failed, if any. There is also a thorough report file created for additional manual inspection, which includes all actions, a thorough explanation of the processing parameters, and keyword output with activities that were incorrectly tagged. Since the term "Log" has been established, it is also feasible to add whatever is necessary to the log file. Since all results are already beautifully structured in HTML, using it for reporting is extremely handy. Additionally, the Robot Framework creates an XML output file that may be utilised for additional study [15]. After running the test suite, an HTML report is produced, as seen in Fig. 2.

Figure 3 shows the test case log file which is generated as an output to analyse the results.

4 Result Analysis

Every time, a test suite that is automated may examine the entire application. It will take a bit longer to go over everything manually. Therefore, when automation does discover an issue, it usually does so quickly after the wrong modification was applied. When there have just been a day's number of modifications, testing is lot quicker, which also means inexpensive. Thus, the effectiveness of automation is increased. If automated tests are well-written, they may be executed in a certain



```
Test Execution Log

−  SUITE  Test
   Full Name:                Test
   Source:
   Start / End / Elapsed:    20181016 10:57:49.381 / 20181016 10:57:49.398 / 00:00:00.017
   Status:                   1 critical test, 1 passed, 0 failed
                             1 test total, 1 passed, 0 failed

   −  TEST  Test do_test
      Full Name:                Test.Test do_test
      Start / End / Elapsed:    20181016 10:57:49.396 / 20181016 10:57:49.398 / 00:00:00.002
      Status:                   PASS (critical)
      −  KEYWORD  ${test} = BuiltIn.Set Variable 1
         Documentation:            Returns the given values which can then be assigned to a variables.
         Start / End / Elapsed:    20181016 10:57:49.397 / 20181016 10:57:49.397 / 00:00:00.000
         10:57:49.397   INFO   ${test} = 1
      −  KEYWORD  BuiltIn.Log ${test}, WARN
         Documentation:            Logs the given message with the given level.
         Start / End / Elapsed:    20181016 10:57:49.397 / 20181016 10:57:49.397 / 00:00:00.000
         10:57:49.397   WARN   1
      −  KEYWORD  BuiltIn.Log ${test}
         Documentation:            Logs the given message with the given level.
         Start / End / Elapsed:    20181016 10:57:49.398 / 20181016 10:57:49.398 / 00:00:00.000
         10:57:49.398   INFO   1
```

Fig. 3 Test case log file

sequence that might change daily. This can be a quick and easy approach. Before Robot Framework, running the test suite required one candidate to run the test cases systematically, search for trails, and be occupied for couple of days [16]. With Robot Framework, the actual system only takes about an hour to complete, and since only one single operation is required, a user may focus on other things while the test suite is being executed.

The new automated testing platform, which is the WebApp tool, will be beneficial and useful in all areas of testing. As technology advances, different new features may be implemented into the tool to provide further functionality. It is an essential component in developing a successful testing plan. Various advanced machine learning and artificial intelligence methods can be used to develop a feature that speeds up request execution and placement. The benefits of automating this execution procedure can be leveraged to compensate for the disadvantages of the human approach. In the future, the WebApp may be made more reactive to handle severe server demands, and alternative testing frameworks and methods can be introduced [17]. Because there is a chance for human mistake, the manual testing procedure is wildly inaccurate, but the automated process is trustworthy because it is script- and code-based. Additionally, automated testing is done quickly, but manual testing takes a lot of time. Table 2 depicts the statistics for the comparison between Manual and Automated testing.

Results comparison of manual testing with automation testing using robot framework shows that automated tests run on average 80.46% faster than manual tests. Without the use of test scripts, a human executes the tests manually, step by step. Test automation frameworks, along with additional tools and software, are used to continuously perform tests during automated testing. The human tester performs it during

Table 2 Time comparison between manual and automated testing

Metrics	Manual (in hours)	Automatic (in hours)
Preparation of one test case	5:00	5:00
Execution of one test case	0:02	0:02
Report for one test case	0:04	0:01
Total time used for one test case	5:06	5:03
For 100 test cases—one suite run	10:24 Command line interface	2:00 Machine interface

manual testing. The tool performs it during automated testing. Manual testing is difficult and slow. But it excels at handling complicated situations, which is a strength. Coding and test management are necessary for automated testing. On the bright side, it covers a lot more combinations and is considerably faster. Users often prefer automated testing compared to manual testing since it saves time. The time-consuming nature of manual testing led to the development of automated testing. Testing that is automated moves forward much more quickly.

5 Conclusion

Working with the Robot Framework has the advantage of allowing users to write test cases that maintain an organic procedure, with prerequisites, execution, validation, and cleaning coming last. Real language is used for keyword definitions, making it easier for non-technical people to understand test cases. This, combined with its straightforward use and simple library extensions, make it an excellent tool for test case automation. Everything is reviewed periodically, and results are made and uploaded on the websites instantly as well. When the choice to incorporate continuous integration with a WebApp tool was taken, this also conserved a substantial amount of time. The easiest way to gauge the expense of automating a test is to look at how many manual tests it prevents from running and how many faults it creates [17]. This is undoubtedly the Robot Framework's largest asset. Robot framework is easy to set up and use, and it facilitates in the implementation and development of test cases. Due to automating, manual work can be reduced while still generating error-free results. By reducing the requirement for manual testing, the expense of automating is reduced, which gives the robot framework its greatest strength. It is a very dependable and viable framework for writing and validating test scenarios. The produced report and log files aid in the rapid investigation and troubleshooting of issues. Manual testing can also be used to calculate the cost of this. Integration with the WebApp tool via the robot framework significantly minimizes human intervention time and effort.

The drawbacks of automated testing are without a human element, it might be challenging to get understanding of the user interface's visual elements, such as colours, fonts, and sizes. The price of the testing project may rise due to the potential expense of the automated testing technologies. Testing automation software is not yet error-free. Each automation technology has constraints that restrict the breadth of automation. Another significant challenge with automated testing is script debugging. It costs money to maintain tests. Some common challenges faced during automation include email assertions and asserting with background colours, as part of GUI testing, it is often necessary to visually verify if the colour has been modified after activities. While the background colour isn't available at the tag level for certain web components, the style property and background colour are visible at the tag level for others. Additionally, file handling is required because the programme being tested includes capabilities like file upload and download. Values from external files must be read by the data-driven routines. A list of keywords for managing files in the operating system library is provided by the robot framework.

The future scope involves new automated testing platform that will be beneficial and useful in all areas of testing. As technology advances, different new features may be implemented into the tool to provide further functionality. It is an essential component in developing a successful testing plan. Various advanced machine learning and artificial intelligence methods can be used to develop a feature that speeds up request execution and placement. The benefits of automating this execution procedure can be leveraged to compensate for the disadvantages of the human approach. In the future, the WebApp can be made more reactive to handle severe server demands, and alternative testing frameworks and methods can be introduced.

References

1. Lei B, Li X, Liu Z, Morisset C, Stolz V (2010) Robustness testing for software components. Sci Comput Progr 75(10):879–897
2. P. Laukkanen (2006) Data-driven and keyword-driven test automation frameworks. Master Thesis, Helsinki University of Technology
3. Mishra A, Jaiswal A, Chaudhari L, Bodade V (2022) Health record management system—a web-based application. J ISMAC 4 (2021):301–313
4. Rice RW (2003) Surviving the top ten challenges of software test automation. In: Proceedings of the software testing, analysis & review conference (STAR) East 2003. Software Quality Engineering
5. Joby PP (2020) Expedient information retrieval system for web pages using the natural language modeling. J Artif Intell 2(02):100–110
6. Lewis WE (2005) Software testing and continuous quality improvement. Auerbach Publications
7. Liu J-P, Liu J-J, Wang D-L (2018) Application analysis of automated testing framework based on robot. In: Third international conference on networking and distributed computing. IEEE
8. Yuste P, de Andrés D, Lemus L, Serrano JJ, Gil P (2017) Integrated Nexus-based real-time fault injection tool for systems. In: IEEE international conference on dependable systems and networks

9. Batni NS, Shetty J (2020) A comprehensive study on automation using robot framework. Proc Int J Sci Res (IJSR) 9(7):1033–1036
10. Yadav V, Botchway RK, Senkerik R, Oplatkova ZK (2021) Robotic automation of software testing from a machine learning viewpoint. Soft Comput J 27(2):1–6
11. Yashaswini HR, Kulkarni P (2021) Robot automation framework for implementing and developing the requirements of network elements. Int J Adv Res Ideas Innov Technol 7(3):1–4
12. Poornachandra Tejasvi TM, Prakash KR (2019) Automation of NetAct integration using robot framework. IJERT Publications
13. Nagendra M, Chinnaswamy CN, Sreenivas TH (2018) Robot framework: a boon for automation. ISSN Publications
14. Hassan A, Hudec L (2018) Role based network security model: a forward step towards firewall management
15. Bhuvaneshwari, Hemanth Kumar AR (2017) Automation for configuration of mobile networks using robot framework
16. Monica, Shettar SN (2017) Survey on robot framework. IJERT Publications
17. Stresnjak S, Hocenski Z (2011) Usage of robot framework in automation of functional test regression. ICSEA Publications

Detection of Mirai and GAF-GYT Attack in Wireless Sensor Network

Hanjabam Saratchandra Sharma, Moirangthem Marjit Singh, and Arindam Sarkar

Abstract Wireless Sensor Network plays an important role in collecting data from different environments where human involvement is deemed fatal or unnecessary. Apart from its usefulness, security threats and vulnerabilities exist as a common problem. A robust IDS (intrusion detection system) in WSN will be helpful in detecting and classifying the types of attacks, so as to remove or nullify the security threats. In this paper, we proposed a method to detect Mirai and GAF-GYT attacks in WSN using CNN along with f_classif function and normalization. Further, implementation of the proposed method has been carried out considering the scenarios: CNN without normalization along with f_classif function and CNN without normalization and without f_classif function. It is seen that the method that uses CNN along with f_classif function and normalization exhibits better performance in terms of parameters such as TPR, PPV, TNR, NPV, FPR, FDR, and FNR.

Keywords Wireless sensor network · Convolution neural network · Mirai attack · GAF-GYT attack

1 Introduction

A Wireless sensor network is a network formed by a collection of a huge number of a distributed autonomous sensor nodes, each node is a sensor device which consists of a sensor that measures various evironmental surrounding conditions such as heat, humidity, sound, pollution level, vibrations, etc., a transceiver, that receives or transmits the sensed or collected data to other neighboring sensor nodes, a battery that powers up the sensor device, a processor unit for computational purposes and a

H. Saratchandra Sharma (✉) · M. Marjit Singh
Department of Computer Science and Engineering, North Eastern Regional Institute of Science & Technology, Nirjuli, Arunachal Pradesh, India
e-mail: sarat.hanjabam@gmail.com

A. Sarkar
Department of Computer Science and Electronics, Ramakrishna Mission Vidyamandira, Howrah, West Bengal, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_44

storage unit that may be used to store limited data. WSNs are widely used in various areas such as health care monitoring, industrial monitoring, Environmental/Earth sensing, battlefields, and traffic controls. The nodes are resource-constrained and hence they have limited processing power, storage, and limited energy. WSN can experience various problems during deployment because of hardware, software, and environmental issues. The reliability of the WSN is in constraint at all times with various basic objectives including long term deployment, highly reliable data transmission [1]. There are many functions of the WSN and data analysis is one of the key functions of the WSN. And intrusion detection is one of the main objectives of data analysis [2].

2 Background

There are several attacks that are encountered in the Wireless Sensor Networks. However, the paper focuses only on Mirai and GAF-GYT attacks which is briefly discussed in Sects. 2.1 and 2.2.

2.1 Mirai Attack

Mirai is a malware that converts normally functioning networks into bots that are remotely controlled that can be manipulated to be part of a botnet in large-scale network attacks [3]. It attacks devices like home routers, cctv cameras, making them into attack network of remote controlled bot. Mainly used by cybercriminals to attack computer system in massive DDoS.

2.2 GAF-GYT Attack

It is a malware that infect Linux System to launch DDoS. It is also known as Baslite. It attacks small routers in home and office to launch DDoS attacks. It targets various IoT and WSN and perform DDoS attacks at the same time. The GAFGYT mainly target the gaming industries [4].

3 Related Work

Sarma [5] proposed a model based on DCNN to detect Mirai and GAF-GYT attacks in IoT. DCNN is applied for classification. The proposed model used MAIG for optimization of filters and its size in the convolution layer. The result of the experiment shows that the proposed model achieved higher accuracy in the detection of Mirai and GAF-GYT attacks as compare to methods such as DCNN, FAE-GWO-DBN and AIG.

Liu et al. [6] formalized a multiple-mix-attack model. Then, PD (Perceptron Detection) was proposed that make use of perceptron and k-means to determine the trust values of the nodes of IoT and to detect malicious nodes. The route of the network is optimized and enhanced preceptron learning process named as PDE (Perceptron Detection with enhancement) is design to enhanced the detection accuracy. From the result it is shown that PD and PDE have higher rate of accuracy in detecting malicious nodes as compared to other similar methods.

Baig et al. [7] have proposed an ADE-based attack detection scheme suitable for IoT. The paper presents a smart DoS detection strategy that consists of modules for generation of data, ranking the features, training, testing and generation. The proposed method was tested in real life condition using actual IoT attacks and is found to be superior from the other existing traditional classification techniques in terms of detection accuracy.

Ismail et al. [8] proposed a multi-layer machine learning detection system to palliate different types of cyber-attacks that harms WSNs. Two models of machine learning detection system are deployed one at Base Station and another at the monitor nodes. For First-layer detection, Naive Bayes algorithm is used and for Second-layer detection, LightGBM algorithm is used for binary classification and multi-class classification respectively. Internal DoS (four network-layer) attacks are detected using the proposed system.

Umamaheshwari et al. [9] proposed a decision tree based attack detection system. To reduce the detection time, they experimented with different feature selection algorithms such as Correlation Score, Fisher Score, Relief and Minimum Redundancy maximum relevance (MRMR) algorithms. And from the results, MRMR feature selection performed superiorly from the other algorithms and yields 98.58% in accuracy, 92.81% in sensitivity, 98.46% in specificity, 93.86% in precision and 15.12 s training time.

Nguyen et al. [10] proposed a IOT botnet detection method based on PSI-rooted subgraph based feature. Their method shows superior performance as compared to Decision Tree, Bagging, KNN, Random Forest and SVM.

Jung et al. [11] proposed a CNN-based deep learning detection system that detect IoT botnets that are malicious.

Rathore et al. [12] proposed an attack detection system based on fog computing. NSL-KDD dataset is used in their method. Their method shows better detection accuracy as compared to centralized attack detection process.

Singh et al. [13] proposed a technique that detect wormhole attacks in wireless sensor network using artificial neural network.

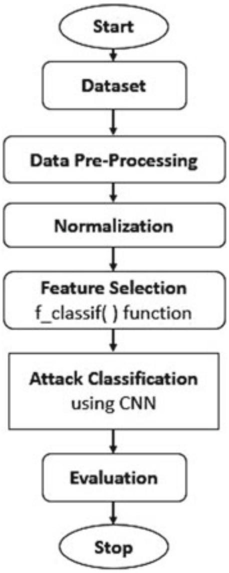
4 Proposed Method

In the proposed method, we used the dataset available from the UCI repository [14] to detect Mirai and GAF-GYT attacks in WSN. The proposed method consist of CNN for classifying the attack and f_classif function for selecting the best features from the dataset. After data preprocessing the normalization of data is carried out in order to: remove data redundancy and set a similar value of range of the features. After normalization phase, f_classif function is applied to the normalized dataset to select the best 20 features from the dataset. Thereafter, attack classification is performed using CNN and finally, performance is undertaken using the parameters such as TPR, PPV, TNR, NPV, FPR, FNR, and FDR. The flowchart of the proposed method is shown in Fig. 1.

4.1 Attack Detection System with F_classif Feature Selection

The principal motive of the proposed research work is based on the detection of different types of attack in WSN, here, an attack detection method is formed using two stages: selection of features using f_classif and Classification. The dataset that we are working on is obtained from the dataset [14] that have 9 applications. From among the 9 applications we worked on 7 applications namely Provision_PT_737E_Security_Camera, Ecobee_Thermostat, Simple-Home_XCS7_1002_WHT_Security_Camera, Samsung_SNH_1011_N_Webcam,

Fig. 1 Flowchart of the proposed method for detection of Mirai and GAF-GYT attack



Phillips_B120 N10_Baby_Monitor, Ennio_doorbell, Danmini_Doorbell and SimpleHome_X CS7_1003_WHT_Security _Camera. These datasets are prepared for pre-processing, here normalization is performed, that convert the data in the range of 0–1. The process of normalization is discussed in brief in the following.

4.2 Normalization

The process is mainly performed before the feature extraction operation. It is a scaling method that is mainly implemented in the data preparation stage to convert the values of numeric columns to a common scale in the dataset. The Normalization is explained in Eq. 1.

$$N = \frac{\widehat{e_{ij}}}{\max\left(\widehat{e_{ij}}\right)} \tag{1}$$

$$i = 1 \text{ to } M, \quad j = 1 \text{ to } N$$

After the normalization, features with best scores are selected using the f_classif method from the normalized data Z = {z_1, z_2, …, z_n}. And classification process is carried out on the selected or extracted features using CNN.

4.3 Feature Extraction Using f_classif Function

Features are selected using the f_classif function. The f_classif() is mainly utilized in extracting the best features (that have the highest value or score) through the SelectKBest class. İt is a method which is obtainable from the scikit-learn that uses a scoring function and and based on the scores of the features, the features are positioned accordingly [15].

5 Results and Discussion

5.1 Experimental Setup

The proposed detection technique was implemented in Jupyter Lab using scikit-learn. Seven applications were used that was retrieved from dataset given in [14]. For convinience the proposed method is abbrebriated as N and F_classif and the other two scenarios that is NC (method that uses CNN without normalization and without f_classif function) and FA method that uses CNN without normalization along with f_classif function). Here, two analyis was done under GAF-GYT attack and Mirai attack. The performance of the method N and F_classif was compared with a method NC and a method FA. Positive and negative measures were taken into account while carrying out the analysis. Positive measures includes sensitivity or TPR (True Positive Rate), specificity or TNR (True Negative Rate), precision or PPV (Positive Predictive Value), and NPV (Negative Predictive Value). And negative measures that includes FPR (False Positive Rate), FDR (False Discovery Rate), and FNR (False Negative Rate).

5.2 Performance Analysis with Positive Measure Under GAF-GYT Attack

The performance of the work is compared for 7 applications regarding the positive measures for the detection of the GAF-GYT attack (Fig. 2). The performance is found to be superior, if it maintained a higher score when juxtaposed to other models. For sensitivity or TPR measurement, the method N and F_classif achieves better performance over FA and NC by 0.81 and 33.21%, respectively. For specificity or TNR measurement, the method N and F_classif achieves better performance over FA and NC by 0.0048 and 10.64%, respectively. The method N and F_classif achieves better performance over FA and NC by 0.0097 and 6.375%, respectively for precision or PPV. And finally for NPV, the method N and F_classif achieves better performance over FA and NC by 0.93 and 16.58%, respectively. Thus, method (N and F_classif) performed superiorly over FA and NC for all positive measures.

5.3 Performance Analysis with Negative Measure Under GAF-GYT Attack

The method's performance is compared for 7 applications regarding the negative measures for the detection of the GAF-GYT attack (Fig. 3). The performance of a system under consideration is found to be better, if the result observed of negative measures is minimum. While considering the FPR measure the method N and F_classif is found to be −0.0035 and −10.64% lesser than FA and NC methods. For FNR measure the method (N and F_classif) is found to be −0.81 and −33.21% lower than that of FA and NC methods. And finally for FDR, the method N and F_classif achieves lower performance over FA and with NC by −0.01 and −6.37% respectively. Thus, from the above observations, the performance analysis of the method N and F_classif with respect to negative measures under GAF-GYT attack is low. Hence, we can deduced from the above observation that method (N and F_classif) performed much better that the other two methods.

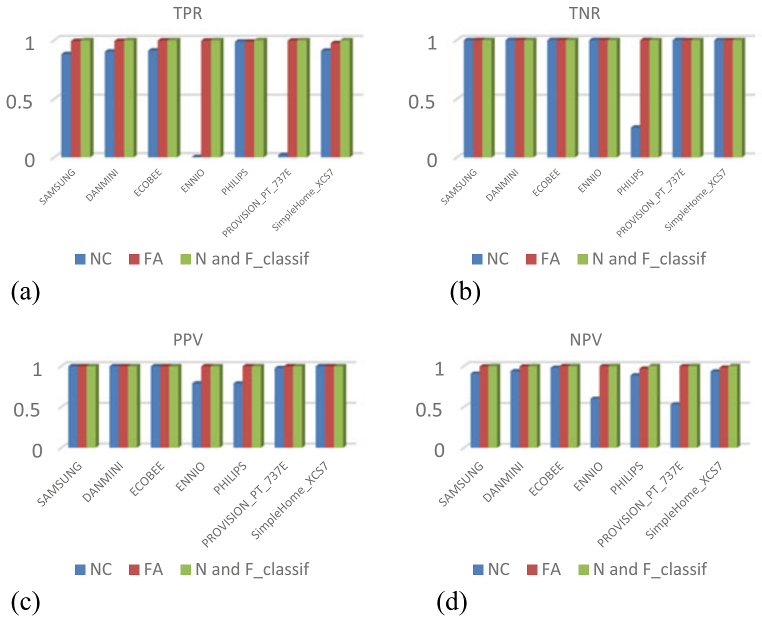


Fig. 2 Models' performance under GAF-GYT attack for positive measures **a** TPR **b** TNR **c** PPV **d** NPV

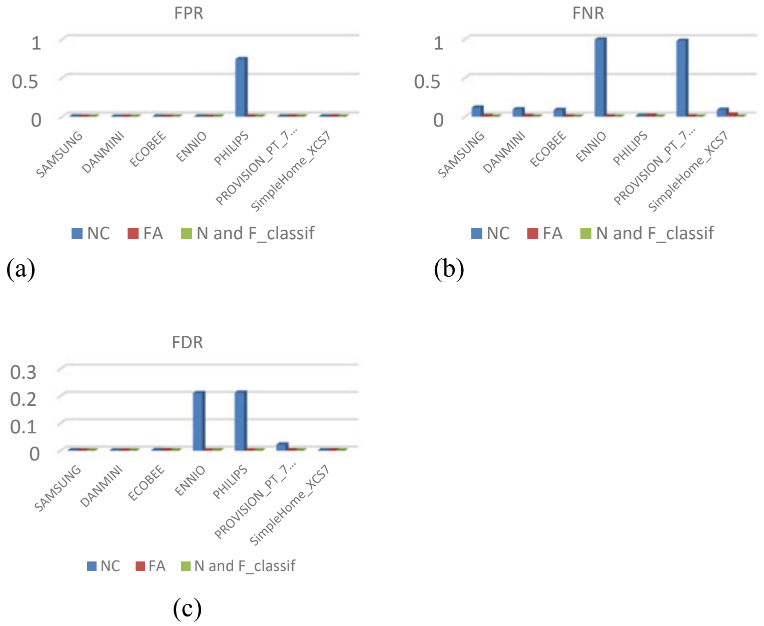


Fig. 3 Models' performance under GAF-GYT attack for negative measures **a** FPR **b** FNR and **c** FDR

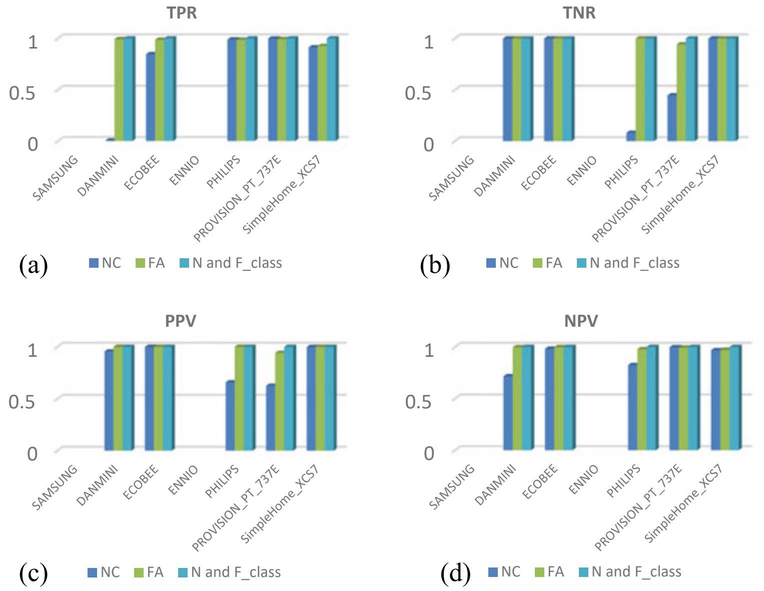


Fig. 4 Models' performance under Mirai attack for positive measures **a** TPR **b** TNR **c** PPV and **d** NPV

5.4 *Performance Analysis with Positive Measure Under Mirai Attack*

For sensitivity or TPR measurement, the method N and F_classif achieves better performance over FA and with NC by 2.25 and 22.65%, respectively. For specificity or TNR measurement, the methodN and F_classif achieves better performance over FA and NC by 1.14% and 28.31%, respectively. The method N and F_classif achieves better performance over FA and NC by 1.17% and 14.05%, respectively for precision or PPV. And finally for NPV, the method N and F_classif achieves better performance over FA and NC by 1.17% and 9.04%, respectively. Thus, method N and F_classif performed superiorly over FA and NC for all positive measures in the detection of Mirai attack (given in Fig. 4).

5.5 *Performance Analysis with Negative Measure Under Mirai Attack*

On observing the FPR measure the method N and F_classif is found to be −1.14% and −28.31% lesser than FA and NC methods. For FNR measure the method (N and F_classif) is found to be −2.25% and −22.65% lower than that of FA and NC

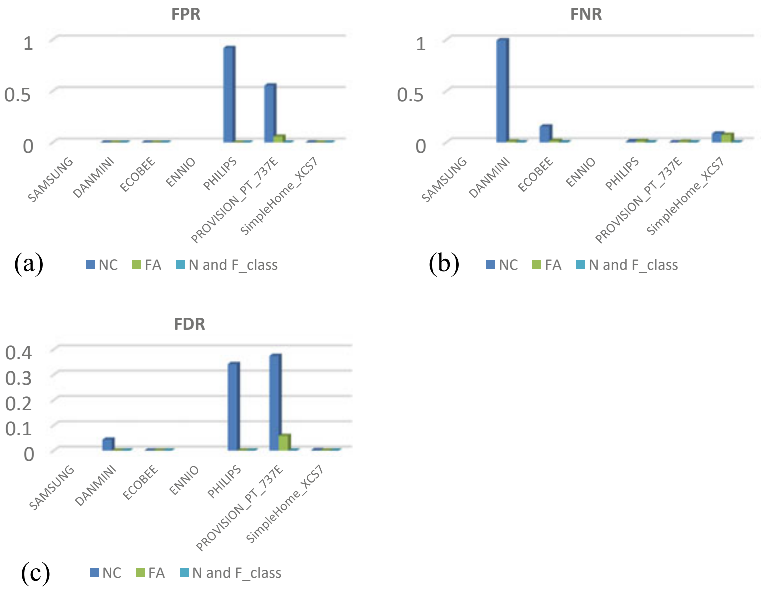


Fig. 5 Models' performance under Mirai attack for negative measures **a** FPR **b** FNR and **c** FDR

methods. And finally for FDR, the method N and F_classif achieves lower performance over FA and with NC by −1.17% and −14.05% respectively. Thus, from the above observations, the performance analysis of the method N and F_classif with respect to negative measures under Mirai attack is low. Hence, we can conclude that method N and F_classif performed better in detection of Mirai attacks as compare to the other methods (given in Fig. 5).

In the approached method N and F_classif we use CNN along with f_classif function to select the best features from the dataset. İn other scenarios FA we use CNN without normalization along with f_classif function and in NC method, we simply use CNN without normalizing the dataset and without f_classif function. And from the above analysis we can conclude that method N and F_classif performed much better as compare to the FA and NC.

6 Conclusion

An approach to detect Mirai and GAF-GYT attacks in WSN has been proposed using CNN and f_classif function. The attack detection technique is formed using two stages: feature selection using f_classif and classification. From among the 9 applications we worked on 7 applications namely Provision_PT_737E_ Security_Camera, Ecobee_Thermostat, Simple-Home_XCS7_1002_WHT_ Security_Camera, Samsung_SNH_1011_N_Webcam, Phillips_B120N10_Baby_ Monitor, Ennio_doorbell, Danmini_Doorbell and, SimpleHome_XCS7_1003 _WHT_Security_Camera. Further, implemetation of the proposed method has been carried out considering the scenarios: CNN without normalization along with f_classif function and CNN wihout normalization and without f_classif function. From the results it is seen that method N and F_classif exhibits better performance as compare to NA and NC in the detection of Mirai and GAF-GYT attacks. For TPR measurement, the method N and F_classif achieves better performance over FA and NC by 0.81% and 33.21%, respectively, for TNR, by 0.0048% and 10.64%, 0.0097% and 6.375%, respectively for PPV. And for NPV, by 0.93% and 16.58%, for FPR the method N and F_classif is found to be −0.0035% and −10.64% lower than FA and NC methods, −0.81% and −33.21% lower than FA and NC methods for FNR. And finally for FDR, −0.01 and −6.37% lesser than FA and NC respectively under GAF-GYT attack. İn case of Mirai attack detection, TPR for N and F_classif is greater than FA and NC by 2.25% and 22.65%, respectively, for TNR, 1.14% and 28.31%, respectively, 1.17% and 14.05%, respectively for PPV. And for NPV, the method N and F_classif achieves better performance over FA and NC by 1.17% and 9.04%, respectively, for FPR measure the method N and F_classif is found to be −1.14% and −28.31% lesser than FA and NC methods, −2.25% and −22.65% lower for FNR. And for FDR, the method N and F_classif achieves lower performance over FA and with NC by −1.17% and −14.05%, respectively. Although, it is noted that the performance of the approached method N and F_classif is better in comparison to the other two scenarios, the performance of the proposed method can be enhanced by using an optimizer to optimized filter size and filter count in the convolution layer and as well as the activation function selection. The selection of the use of perfect otimizer is kept for future work.

References

1. Jurdak R, Wang XR, Obst O, Valencia P (2011) Wireless sensor network anomalies: diagnosis and detection strategies. Intell-Based Syst Eng 309–325. https://doi.org/10.1007/978-3-642-17931-0_12
2. OReilly C, Gluhak A, Imran MA, Rajasegarar S (2014) Anomaly detection in wireless sensor networks in a non-stationary environment. IEEE Commun Surv Tutor 16(3):1413–1432. https://doi.org/10.1109/surv.2013.112813.00168
3. Das S, Amritha PP, Praveen K (2021) Detection and prevention of mirai attack. In: Reddy VS, Prasad VK, Wang J, Reddy KTV (eds) Soft computing and signal processing. Advances in ıntelligent systems and computing, vol 1325. Springer, Singapore. https://doi.org/10.1007/978-981-33-6912-2_8
4. Hu X, Sun R, Xu K, Zhang Y, Chang P (2020) Exploit ınternal structural ınformation for IoT malware detection based on hierarchical transformer model. In: 2020 IEEE 19th ınternational conference on trust, security and privacy in computing and communications (TrustCom), 2020, pp 927–934. https://doi.org/10.1109/TrustCom50675.2020.00124
5. Sarma SK (2021) Optimally configured deep convolutional neural network for attack detection in ınternet of things: ımpact of algorithm of the ınnovative gunner. Wirel Pers Commun 118(1):239–260. https://doi.org/10.1007/s11277-020-08011-9

6. Liu L, Ma Z, Meng W (2019) Detection of multiple-mix-attack malicious nodes using perceptron-based trust in IoT networks. Futur Gener Comput Syst. https://doi.org/10.1016/j.future.2019.07.021
7. Baig ZA, Sanguanpong S, Naeem Firdous S, Nhan Vo V, So-In C (2020) Averaged dependence estimators for DoS attack detection in IoT networks. Futur Gener Comput Syst 102:198–209
8. Ismail S, Dawoud D, Reza H (2022) A lightweight multilayer machine learning detection system for cyber-attacks in WSN. In: 2022 IEEE 12th annual computing and communication workshop and conference (CCWC), 2022, pp 0481–0486. https://doi.org/10.1109/CCWC54503.2022.9720891
9. Umamaheshwari S, Kumar SA, Sasikala S (2021) Towards building robust ıntrusion detection system in wireless sensor networks using machine learning and feature selection. In: 2021 international conference on advancements in electrical, electronics, communication, computing and automation (ICAECA), pp 1–6. https://doi.org/10.1109/ICAECA52838.2021.9675609
10. Nguyen H-T, Ngo Q-D, Nguyen D-H, Le V-H (2020) PSI-rooted subgraph: a novel feature for IoT botnet detection using classifier algorithms. ICT Express (in press, corrected proof, Available online 7)
11. Jung W, Zhao H, Sun M, Zhou G (2020) IoT botnet detection via power consumption modelling. Smart Health 15:100103
12. Shailendra Rathore J, Park H (2018) Semi-supervised learning based distributed attack detection framework for IoT. Appl Soft Comput 72:79–89
13. Singh MM, Dutta N, Singh TR, Nandi U (2020) A technique to detect wormhole attack in wireless sensor network using artificial neural network. In: Suma V et al (eds) Evolutionary computing and mobile sustainable networks. Lecture notes on data engineering and communications technologies, vol 53. Springer, Singapore, pp 297–307. https://doi.org/10.1007/978-981-15-5258-8_29
14. https://archive.ics.uci.edu/ml/datasets/detection_of_IoT_botnet_attacks_N_BaIoT
15. Pathan MS, Nag A, Pathan MM, Dev S (2022) Analyzing the impact of feature selection on the accuracy of heart disease prediction. Healthc Anal 2:100060. https://doi.org/10.1016/j.health.2022.100060,ISSN2772-4425

A Brief Review of Network Forensics Process Models and a Proposed Systematic Model for Investigation

Merly Thomas and Bandu Meshram

Abstract Network forensics is a branch of Digital Forensics concerned with analysing the network traffic to see if any anomalies are present that may indicate an attack or could lead to one. The goal is to figure out what kind of attack it is by capturing the details, store them in a forensically sound manner, analyse, and then present them in some visual form. A model based on traceability and scenarios, with proven literature and justification is desired. This study offers a professional digital framework in which the investigative process model enhances the systematic tracking of offenders. Cyber fraud and digital crimes are on the rise, and unfortunately less than two per cent is the conviction rate worldwide. Continuous and scientific research in this area is crucial to ensure safe and secure internet usage especially for money transfers and confidential personal communication. This paper examines the essential development phases of a Network forensics investigation model, and compares different network and digital forensic methods, and also offers a systematic model of a digital forensic model for cybercrime investigation. The survey also includes classifications based on infiltration detection systems, trace backs, distribution models, and attack maps. The aim of this study is to facilitate the digital forensic process and identify improvised practices. The Systematic Network Forensic Investigation model (SNFIM) aims to establish appropriate policies and procedures for practitioners and organizations.

Keywords Network forensics · Digital forensics · Infiltration detection systems · Traceback · Distribution · Attack maps

M. Thomas (✉) · B. Meshram
Veermata Jijabai Technological Institute, Mumbai, India
e-mail: mthomas_p18@ce.vjti.ac.in

B. Meshram
e-mail: bbmeshram@ce.vjti.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_45

1 Introduction

In today's world of technology, advances in computers, small communicating devices, and networks have made our society increasingly dependent on cutting-edge information and communication technology and mobile appliances for personal activities, business transactions, and government services [1]. Identity theft, cyber-bullying, data leakage or information denial, malicious software destroying the normal functions of the systems, distributed denial of service (DDoS) orchestrated through botnets, and malware targeting specific appliances such as smart vehicles, are all examples of new threats and cybersecurity issues widespread in the networked world. As a result of such incidents, the perpetrators are more intelligently and strategically identified, which makes it easier to perpetrate crimes and pursue illegal intentions [2]. Internet crimes are rampant, complicated and worth all the attention of academicians and researchers. Annual statistics paint a bleak picture of the number of people affected by cybercrime and the rate at which it is growing. It's no wonder that cyber-crime is on the rise, and as a result, many people worldwide are researching this field of Forensics, and enhanced methods has to constantly evolve to keep pace with the strategic moves and advances of the perpetrators [3]. When the nodes are decentralized a decentralized solution may be a more efficient solution in grey and black hole attack settings by solving the problem of message overhead by appropriate measures [4].

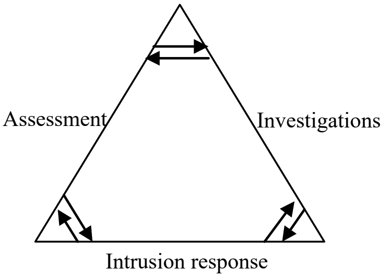
Digital forensic can be put into two broad categories as follows [5]:

- Computer forensics
- Network forensics.

As the Internet grows and the global spread of net-enabled devices increases, computer-related formal surveys are needed to get a grip of the situation. As a part of crime research, surveys describe the characteristics of a large population, map the entire networks in question, and acquire the knowledge necessary to administer such an extensive infrastructure. Crimes such as computer policy violations, fraud, email, and social media harassment, cyber-stalking, privacy, and terrorism need to be recorded on devices. Law enforcement bodies, network administrators, and criminal investigators rely on computer-assisted skills to investigate criminal and civil cases. Generally, digital forensics examines data that can be collected and analysed from device forensics or other residues, like when an archaeologist digs a site [6]. Information regarding a crime may already be recorded and available on a disk or device, but finding its location and accessing it is not easy. In its place of computer surveys, network expertise provides in sequence about how a user, or hacker gains the right of entry to the network. When a user or attacker logs in, criminal investigators use log files to explore, in sequence, what steps they took to commit the crime [7]. Finding the different URLs they visit and the login credentials used by a user or attacker may be found by different tools and multiple experts.

Digital forensic analysts work as a team with experts from various fields to secure computers and network assets for an enterprise [8]. Computer literacy is one of the

Fig. 1 Forensic investigation triangle



three triangular boundaries that ensure corporative computer defence and investigation. The triangle given in Fig. 1, the three sides depict the combined efforts of the team closely monitoring the assets of an organisation [9]:

- Vulnerability assessment and risk management
- Network intrusion detection and incident response
- Computer investigation.

Susceptibility appraisal and risk management experts have network monitoring, interruption discovery, and a team of digital investigators investigating forensic evidence in a suspicious context in order to gather digital confirmation connected to an occurrence or offence. Network expertise can be alienated into two distinct streams [10]. Network attacks can be divided into the following types:

- Probing: (Surveillance and Interception)—Probing is a type of attack in which an attacker explores a network for vulnerable information or to find known flaws. An attacker with a map of available services on a network can use this data to seek exploits. Examples for this type of attack include Mscan, Nmap, Pingsweep and Port sweep attacks.
- DOS: Denial of service (Interruption): An attack which involves flooding where attackers make computing or memory resource of the target machine too busy or too full to handle legitimate requests, thereby denying legitimate users access to the machine, e.g., syn flood.
- User to Root Attack (U2R): An exploit in which the attacker who has access as a normal user account of the system exploits some vulnerabilities to gain root or superuser access privileges of the system, e.g., Buffer overflow attacks.
- Remote to Local Attack (R2L): An attack instance where an attacker without having a user account gains access to a remote machine and sends packets. To get access as a user of that machine, the attacker exploits a vulnerability, such as guessing the password.

During the forensic analysis of any network, a team of investigators prepares and investigates the incidence, to obtain information on a public or confidential network and find the source of these attacks. In addition, the data collected by the Network Inquiry Committee, a number of facts and digital evidence are reconstructed that can be submitted to law enforcement agency throughout production [11]. This may include searching of digital devices to generate evidence related to an event that occurred during an investigation. As the quantity of digital data increases, analyzing data becomes a time-consuming process that may affect the legitimacy of the whole process. Existing search tools offer numerous search strategies for document analysis, but no document of interest addressing the magnitude of the types and number of cyber-attacks, could be found [12]. Thus, in the present scenario when a digital forensic tool is introduced into the system, investigators obtain a set of entry permit that are meaningfully related to the subject of the investigator's interest [13].

1. **Identification phase**: determine the data, mechanism and various entities related to the crime.
2. **Image Acquisition and Preservation**: preserving the crime scene by protecting it from any damage.
3. **Data Recovery**: collecting the relevant data or information interrelated to the crime.
4. **Analysis**: The researcher will perform a comprehensive analysis on the data by the theatre pattern matching; filter techniques by utilising a mixture of forensic tools.
5. **Reporting**: The canvassers summarize and present his hypothesis [14] (Fig. 2).

In recent years, with computer crime, even in ordinary crime, important evidence or information has been increasingly stored in various electronic media, such as computers or smartphones. Copying digital data can be achieved by a few key strokes and almost impossible to distinguish from the original. Additionally, electronic data can be easily created, modified, or deleted from the original. Therefore, criminal investigations require high-level as well as diverse technologies to obtain sustainable evidence from digital data on suspects' computers or smartphones [15].

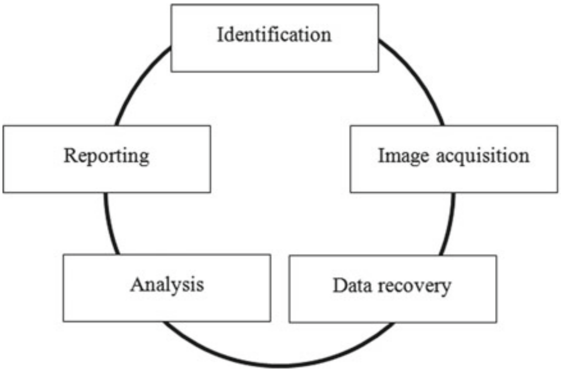


Fig. 2 Forensic process model

2 Related Works

Sathwara et al. [16] have studied a forensic mechanism to investigate IoT-related crimes. IoT poses a number of challenges for forensic analysts. For example, in differentiation between private networks through public networks, the blurred lines between networks are increasingly disappearing. In addition, the testing of IoT tools is further complicated by the integration of diverse tools of IoT forensics, as well as the compatibility of identified and stored devices. The reason to include this paper is the supply of an IoT expert structure. Their goal was to support the digital exploration of IoT devices and to overcome the growing challenges of digital forensics and connection development.

Shrivastava et al. [17] have studied Digital forensics is the systematic retrieval of evidence collected as a consequence of the investigation of critical data. After examining all the historical approaches used in the current models, their advantages and disadvantages are presented and an approach to the Forensic Model (EAF) is suggested; It connects all stages of the digital quest in detail.

Valjarevic et al. [18] shows how to implement a prototype to meet all needs. In their previous works, the authors have suggested a comprehensive and relevant digital forensic procedure, which is used as the basis of the prototype. There is a sample software with two main functions. The first function is the role of an expert organization that can be used to lead and train new operators. The second function is to reliably document all the steps engaged within the specified process in a complete and relevant model of the digital forensic process. The merits of the investigation model include better improvements in the efficiency and effectiveness of the investigation.

Omeleze et al. [19] have studied The Harmonised Digital Forensic Investigation (HDFI) progression representation is at present being developed as an international standard for digital forensics (ISO/IEC 27043), so it needs to be veteran. In this paper (HDFI) the process model is tested with an Android changeable cell phone. Kebande et al. [20] have introduced cost-effectiveness in companies by increasing the availability of digital devices in their day-to-day business operations. However, the development of this technology threatens many experts, because in the absence of an effective security model, digital forensic potential can be used to plan and prepare security events before an organization approved by the BYOD.

Burrows et al. [21] investigates forensic investigative tools used in various steganography techniques have been developed for steganography samples and steganography detection and extraction for smartphones/mobile devices. Charles et al. [22] examine the internet forensic examination process available at the university and determine its ability to conduct a comprehensive digital forensic examination. Kao et al. [23] have shown that the amount of data available for cybercrime investigations are growing at an unprecedented rate, creating problems and huge challenges for law enforcement personnel. In order to present digital data as evidence in court, all data sets in the crime scene or laboratory must be examined in detail.

Jain et al. [24] proposed the network forensic framework which combines the achievements of the past twenty-five forensic models to create a mechanism for creating a new digital forensic model. The planned model offers a consistent investigation technique, which can directly transform model theory into an instrument, an object of historical perspective, reducing the cost and time to investigate any digital crime. Mouhtaropoulos et al. [25] have studied Digital forensics primarily attempts to respond to an information security event.

3 Systematic Networkt Forensic Investigation Model (SNFIM)

The planned model will be based on the Network features and the common application vulnerable to the attacks. There are six main stages in this model, the structure specifies in Figs. 3 and 4.

Preparation Phase: The initial understanding of the problem, as well as the appropriate tools, are all part of the preparation phase. This step is used to secure authorization and approval, before developing a suitable plan. The planning period is all the work and activity that desires to be done until a real exploration is possible [26]. This includes reviewing forensic rules and guiding principle, obtains search warrants, supporting organization, preparation, and developing suitable strategies and equipment.

Acquisition and preservation phase: Prior to the actual forensic examination, the scope of investigating artefacts has to be understood. This phase will clarify which evidences are available and what do we need to acquire in addition. Acquiring and maintaining a clear life cycle and, identifying and collecting evidence of crime

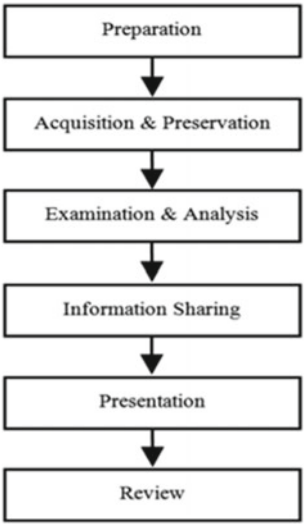


Fig. 3 Network forensic investigation phases

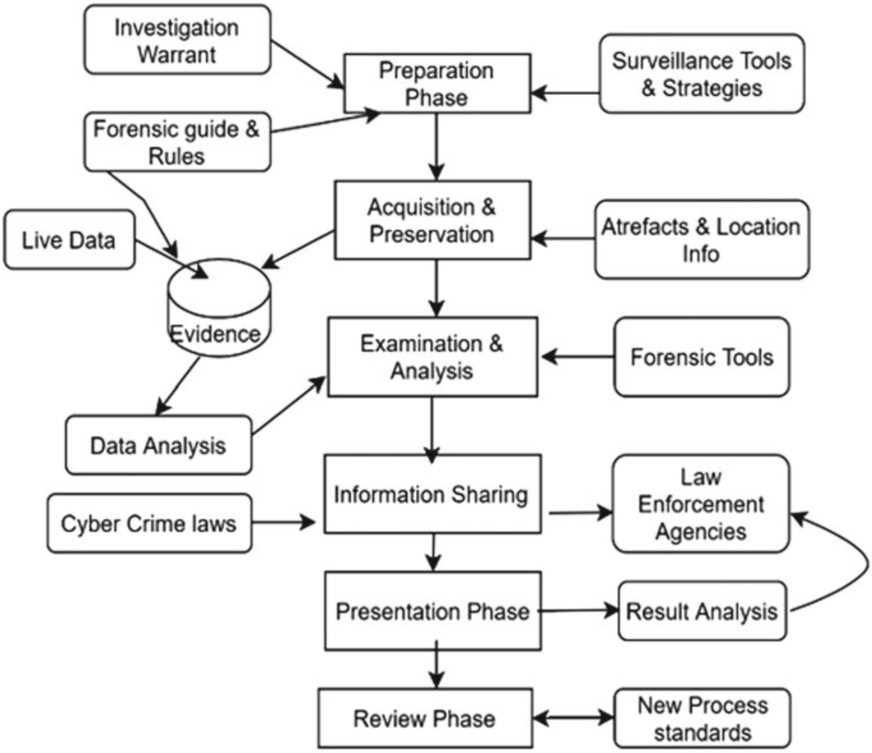


Fig. 4 Flow of the network forensic investigation process in the proposed model

includes labelling, packaging, transportation, image capture, and storage of evidence [27]. Therefore, all objects including access permissions, people of the organization, network configurations during the attack, must be lawfully obtained and properly documented in accordance with clear rules.

Examination and analysis phase: This is the phase where experts and professionals search for digital evidence by studying and analyzing various digital contents. Legally confiscated and properly protected equipments, its details, technical procedures, acceptable guidelines, and approved forensic tools are used to determine the origin of the crime and also to locate the culprit [28]. The evidence presented depends on the purpose of the appointment or it may contradict the initial theory.

Information sharing phase: The easiness with which Internet can be handled, are used effectively on social networks and by the hacker community. The information sharing phase has the ability to obtain a complete criminal profile of an individual, and to develop an effective interrogation strategy. This kind of collaboration and information sharing can effectively contribute to successful events [29].

Presentation Phase: The results of the examination and analysis are compiled and submitted to the concerned law enforcement authorities of multiple locations. For example. to make sure that the evidence is correctly identified and obtained, the evidence is objectively and adequately protected and the jurisdiction is appropriate, The language used in the presentation be concise and uncomplicated to understand by a judge. As an important step in this process, it is important to remember that the panel of defendants must submit the findings to the court for conciliation and approval before the trial judge or arbitral tribunal [30].

Review Phase: Evaluates the entire investigation and identifies areas for improvement. From the launch of the experiment to the test, the intermediate results are used for future improvements. The experience gained and the lessons learned are used to prepare new process steps and standards. Cases are categorized according to their status and understanding, the case is suspended, on going, and completed. This is done to resolve future complications, such as appealing to the court, appearing or referring to an acquittal [31].

4 Classification of Internet Forensic Investigation Technologies

Digital forensic technology on networks is alienated into the following [32, 33]:

- Digital evidence compilation techniques,
- Digital evidence analysis strategies,
- Digital evidence retrieval techniques, and
- Digital evidence documentation technique.

The technology examines and analyzes digital devices, to identify potential sources, transfer electronic data, analyze, search, and protect stored information [34].

All current and accepted network forensic investigations are classified as:

1. Intrusion detection systems
2. Traceback systems
3. Distributive systems
4. Attack graph systems.

4.1 Intrusion Detection Systems of Internet Forensic Investigation

Intrusion detection system (IDS) is software or hardware set-up that monitors the activities of the organization and network for criminal and malicious activities and generates reports for the central administration may be a policy. Figure 5 shows

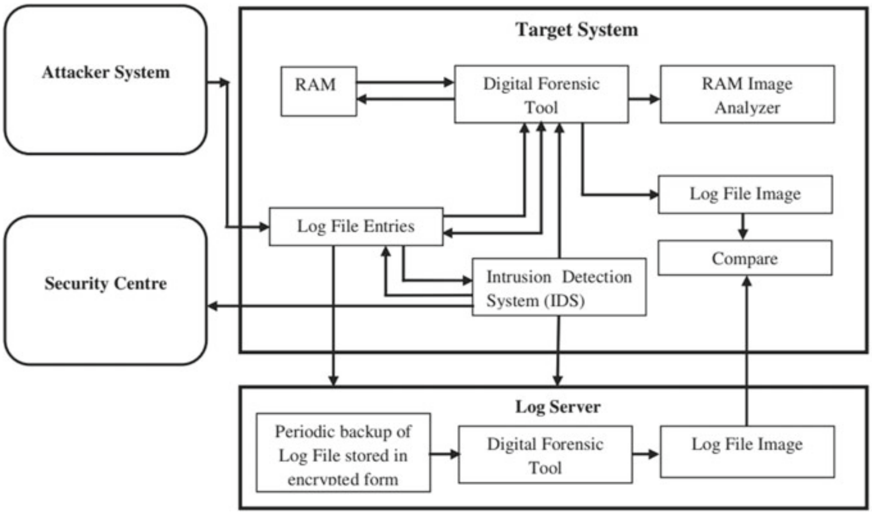


Fig. 5 Automated digital forensic technique with the intrusion detection

the digital forensic technique with intrusion detection system. The main purpose of infiltration detection systems is to detect events, record information about them, and report attempts [35].

4.2 Traceback in Internet Forensic Investigation

It is essential to offer a network expertise IP tracking model based on the assumption and needs listed in the formalities of the process model. The network architecture as shown in Fig. 6, is based on the deterministic packet marking (DPM) using the Autonomous System (AS) approach. In a dual tracking system where the first step includes the AS of each pocket defined by the first column router, and the second step indicates each packet that the AS edge router (ASER) indicates. In both cases, after the pocket is marked, the other router will not be able to mark it. Not every outgoing pocket is marked and incoming pockets are not marked [36].

4.3 Distributive Internet Forensic Investigation

The goal of digital forensics is to answer inquisitive and interesting questions: who, how, what, why, when, and where. The entire forensic examination process is implemented using analytical methods such as the Integrated All-in-Encase Selectors Guide, the U.S. Digital Judicial Forensic Analysis Method, and/or the U.S. National

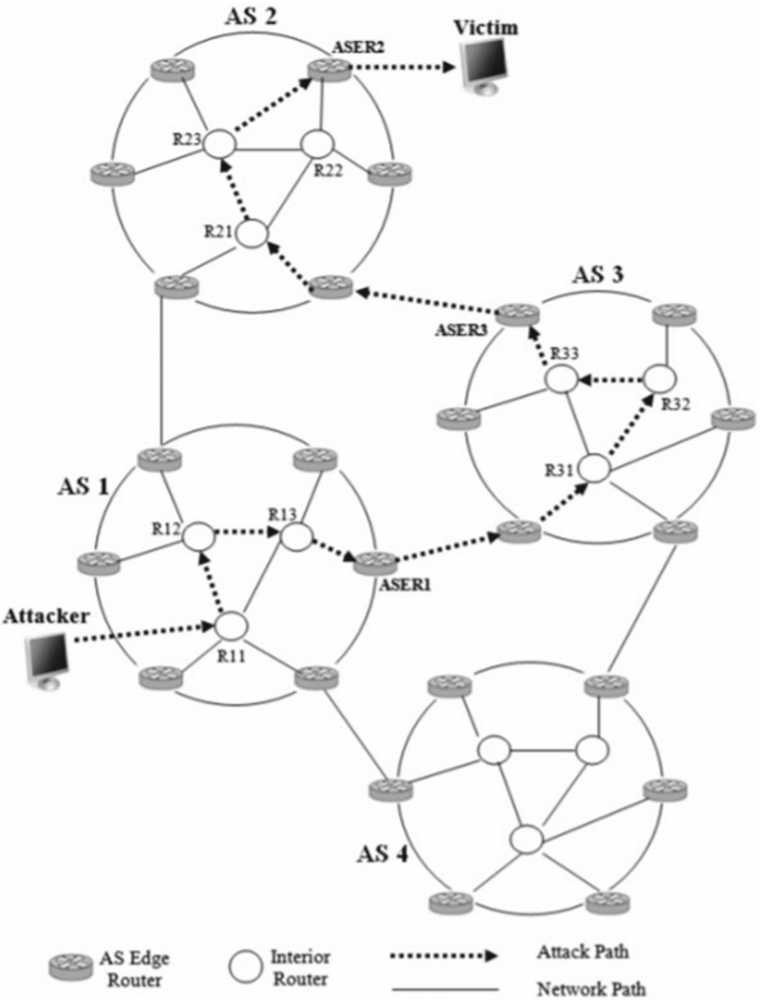


Fig. 6 Architecture for IP traceback

Justice Analysis Guidelines. The use of an analytics framework in conjunction with a subset of data has the potential to quickly identify information from electronic sources and provide a full backup of the media for analysis [37]. In this study, they suggest a method of rapid forensic analysis of digital forensics, which allows digital forensic data subgroups to periodically review and analyze information contained throughout the forensics.

4.4 Attack Graph of Internet Forensic Investigation

To investigate integrated nature and anti-crime attacks, innovative and innovative approach such as bother maps should be used. Harass maps are distinct as a tool for evaluating the performance of bother scenes with the help of acknowledged damages and configurations. Attack graphs are generally used by mainframe administrators and investigators to analyze the type of attack, diverse types of attack methods, detection, and anticipatory channel. Used against these attacks. The different tools in use for generating attack graphs are:

- **TVA** (Topological analysis of network attack vulnerability)—It creates attack maps using a map investigate algorithm. It uses pro maps to produce before and after circumstances.
- **NETSPA** (Network security planning architecture)—Framework uses known vulnerabilities and network security rules to create network models. It serves as a foundation for generational attack mapping to recognize possible attack and root directions.
- **MULVAL** (Multi-host, multistage vulnerability analysis)—Data recording is a structure for integrating damage and system configurations used as a language. It contains a scanner and an analyzer. A logical machine uses data recording rules to record computer behaviour [38].

5 Methodology

Digital forensics has straight requirements in the field of Internet and cyber activities, i.e., to prevent further adverse events, successfully identify the events that lead to the crime and identify the culprits. The methodology should improve the existing preventive measures to prevent recurrences as corporate security professionals improve their standards to protect their corporate networks. The rest of the study pattern is summarized in the following sections.

5.1 Research Questions

This section describes the background of this research and identifies key research questions and methods for clarifying and defining research focus based on the results of previous attempts and new expectations.

- What is the present measurement methods for Netork Forensics tool quality?
- What are the changing technique and directions (data related)?
- Whether the data remain unaffected (to a level acceptable in a court of law)?

The addition and subtraction criteria of the research papers are as follows.

5.2 Quality Assessment

When evaluating the resources, quality assessment (QA) is first used to identify data obtained in meta-analysis, which is very relevant for the resource collection strategy. In addition, quality evaluation results are provided as additional selection criteria.

5.3 Data Extraction and Data Synthesis

Data on research questions and activities were collected before the formation of relevant research questions. Once the data collection is complete, the analysis and review process will be conducted and the results will be presented for review The collected data are listed based on the key components of their solution.

5.4 State-of-the-Art Internet Forensic Models

5.4.1 Intrusion Detection System

Achille et al. [39] have studied Abstract layer theory is applied to describe the purpose and objectives of digital forensic equipment. With compression layers, we determine where errors can occur on the equipment and offer tools to protect them. Forensic types are also defined in terms of abstract layers. Abstract layers are not a new concept, but their use in digital forensics has not been properly documented. Barhate et al. [40] studied that the Intrusion Detection System (IDS) is used to determine infiltration. Its main meaning is to attack and react in a judicious approach. In earlier language, IDS activity is for recognition and rejoinder only. IDS were not capable to achieve system status during the infiltration. Therefore, he could not argue the evidence from the attack in its original form. Rani et al. [41] studied that the Cloud computing has become the flagship service model of computer platforms, providing the necessary resources to all types of users. Singhal et al. [42] Data processing and storage technology were explored to better understand the presentation and use of IDS. Current IDS does not support historical data analysis and data summary. This data model was used to analyze network security and detect attacks on services. Analysts can use their data model to manage multiple data sources (e.g. firewall logs, computer calls, network flow data) and to run the required response time for two orders faster than modern technology. Inayat et al. [43] A special feature for accessing cloud resources is the connection link allowing malicious individuals to perform malicious activities at a loss. This protection is a major challenge in the MCC environment [44]. Occupancy processes are studied through the recording and analysis of computer calls. These methods do not provide sufficient detection accuracy because they do not explain internal infiltration events. Sy et al. [45] the aim of this study is to demonstrate the analytical intrusion detection framework (AIDF) that combines (i) the Probability Model Detection Approach and (ii) the Probability Hypothesis Mechanism that monitors forensic interpretation. Infiltration detection warnings, but unregistered signature rules not expressed in the probability model. AITF compatibility for data integration, (i) understanding of the distribution IDS environment with multiple sensors, and (ii) the mechanism for selecting and integrating multiple model hypotheses. Definition of the most probable forensic examination Zhang et al. [46] studied the Network Intrusion Detection System (NIDS) as an important security tool for detecting malicious activity on modern networks. To speed up the detection of minority classes and ensure performance, proposing a new class of unbalanced processing technology called SGM, a large-scale database that synthetic minority over-sampling technique (SMOTE) and Cassian clustering models. SGM-CNN integrates with a neural network, modifies the unbalanced class process, and investigates the influence of different numbers of nuclei and different learning speeds on model performance. Rahouma et al. [47] discusses the future role of machine learning technology in identifying security threats, the automated response to address security challenges in the Egyptian fiber optic network, and the proposed new threat and response detection model for protecting customer data. Aloqaily et al. [48] ıntroduce an automated secure continuous cloud service access framework for smart connected vehicles that is protected from security attacks and meets service quality (QoE) requirements. Continuous service is available in clusters based on smart car service. Trusted Third Party Companies (TTPs) select cluster leaders who act as intermediaries between service complaints and providers. Customers get better service from selected service providers.

Houmansadr et al. [49] reading a mobile device called a smart phone is becoming more and more difficult and powerful to effectively enhance more features, which raises security concerns for smartphone users. Yampolskiy et al. [50] studied human computer communication based on the skills, style, preferences, knowledge or strategy that people use when working with a computer. It analyzes the features described to describe HCI behaviour. Tthe experimental results are presented to determine behavioural interactions in online gaming networks based on the strategies used by players. Zhang et al. [51] proposed lightweight semantic-based knowledge fusion model (SKFM) model to provide natural language processing (NLP) and credit resources using Schools educational documents to generate information. SKFM relies on NLP based on automated components, and initiates fan searches when unsolicited cases occur.

Vasudevan et al. [52] network infiltration detection system is an important task to detect the attack. At the time of signature or rule based diagnosis, the previous attack is modelled and signatures/rules are adopted. These rules will be used to detect such attacks in the future, but if there is a malfunction or external detection system, normal network traffic is the standard. Data mining and machine learning techniques are widely used in offline NDS. The characteristics of supervised and supervised training include identification of data templates, identification of publishers, definition of learned activity for input characteristics, and generalization of data events. Ulltveit-Moe et al. [53] provides a mobile intrusion prevention system (mIPS) framework that provides enhanced privacy integrated with the Managed Security Service (MS).

5.4.2 Trace Back System

Armoogum et al. [54] studied Network forensic technology is a new approach with active and real-time response features that can be used to investigate network attack such as denial of service (DoS) attacks. After a compromise, there is always a check and investigation after the attack, so useful immediate evidence is lost. Adjustment is required without the loss of serious traffic and live inquiry. IP tracking schemes can be used to identify the source of the attack. We offer the use of mobile agents for live IP trading. Cheng et al. [55] studied the Cloud services offer great options for the practical use of the IP tracking system. An innovative cloud based tracking framework with many positive features that enable ISPs to use tracking services on their networks. While this will give access to advanced training services, restricting access will become a major issue. Selamat et al. [56] in general, the purpose of all methods is to establish the origin of the incident and to maintain the network of custody so that the course of the legal process can be adopted. The trace-qualification process has become an important or integral part of the digital quest process because it has the ability to generate a map from a variety of sources to obtain event evidence to use event events for other related query characteristics. The purpose of this integration is to provide a link between the evidence, the companies, and the evidence involved in this process, as well as the evidence-based collection and evidence, especially during the digital forensic framework collection phase. Alex et al. [57] due to the popularity and constant availability of the cloud, the issue of security is still a major issue [58]. The user does not know where the data is stored and in which data centre. The distributed nature of the cloud is again the protection point of the cloud, which allows harmful actions to be performed very easily. Although much has been done to ensure the required data protection in Cloud, much work needs to be done to maintain technological advancement, and data production from various computing environments will increase rapidly [59]. Digital Forensics process depends on the continuous TF steps, each step depends on the continuous TF procedures, and each process depends on the assignments and sub-assignments, respectively. There are several TF process models in the literature that define DF stages, but there are no DF models that define step-by-step sequencing procedures for identified crimes [60]. An analytical crime scene procedure model (ACSPM) addresses this shortcoming. In addition to the entire digital investigation process and the completion stages of the trial, crime is the focus of digital case law. Helix, Blackrock Network Security, GNU/Linux A unique solution is to create a unique live CD GNU/Linux distribution after the digital forensic distribution. The resulting software enables effective network monitoring and analysis, further possibilities for future needs, and greater portability compared to standard digital forensic software [61].

Kwon et al. [62] the leader of the global joint investigation was arrested in March 2018, but in the event of a Bank of Bangladesh robbery, the damages have not yet been resolved due to a lack of digital forensic evidence. This study offers the best IP forensic tracking and visualization techniques for digital forensics. Various process models describing digital court procedures and procedures have been developed with issues associated with these models [63]. Security incidents targeting information systems as complex and intricate items may be investigated and incoming individuals may be responsible for the lack of evidence of guilt. They have set up a digital forensic system (DigFornet) [64] on the net that can be used to analyze security incidents and explain the activities of attackers. The Digfornet Infiltration Response Team uses knowledge and system tools to reconstruct attack potential and show how a computer works at each stage of the scene and introduced a hypothetical concept to address the lack of data related to attack scenes or automated and lost evidence of investigative knowledge. Industrial Revolution 4.0 [65] explores the automatic communication of digital devices around the world, including cyber physics devices, IoT devices, mobile devices, storage devices, and many other digital devices, including PCs and PCs; The rate of cyber crime will increase. It questions the need for an additional digital forensics investigation framework (DFIF) to effectively investigate digital crimes in court; the integrity of the resources must be maintained when working on the structure. Mapping leads to a better investigation process. Jain et al. [66] the use of technology is important. Digital forensics is a specialized field of computer forensics in which scientific procedures and tools are used to obtain digital evidence in court. The main goal is to introduce a digital forensic tool based on the digital forensic framework. Wazid et al. [67] clarified the biggest challenge in the cyber world has emerged as realism. Many digital forensic devices are being developed to meet this challenge, but hackers are developing counter-technologies at the same pace. It covers the basics of digital expertise along with the latest trends in social networking, cloud computing, websites and phishing.

5.4.3 Distributed System

Trcek et al. [68] it turned out those global networks were moving towards service-oriented structures and touch networks. On the one hand, Spectrum creates a mix of new services with multiple services, and on the other hand, it represents the influence of the Internet. These two types are different and can be used in many different ways, which can lead to serious complications in case of abuse. As a result, both trends increase the need for new approaches to digital skills, as evidenced. Since technology alone is not enough, it should provide adequate support for relevant investigative procedures that have become the subject of international consensus. Singh et al. [69] the computer system reads forensic tracking and these files contain unstructured text, so their analysis is difficult, leading to the introduction of automated technology that provides new and important information on how to display the clustering of displayed documents. This can be achieved by changing the algorithm differently from the content used to create the cluster. They work here in the expert process of document summary in expert investigation. Here we associate it with the K-algorithm, which improves the performance of computers and actively converts this work using MATLAB.

Roussev et al. [70] distributed digital forensic (DDF) devices and traditional intelligence tools operating in a single station offer numerous real-world examples of their range, which is a serious barrier to the timely processing of digital evidence. Based on the observations of routine intelligence work, we will highlight the DDF software about computer requirements. Describe another modern forensic technology implemented by switching to DDF equipment.

Ellison et al. [71] it has been explored that there are different models in the areas of digital expertise and security due to the unique function of each knowledge. A knowledge base that operates on a set of assumptions, facts, and laws is called a knowledge base, and is built on the structure of dynamics. Ontology has its philosophical origins; it deals with existence and what is true. In this case, gynaecology, which is presented as the basis of knowledge, requires the suggestions (and universes) of truth based on knowledge and logic. Such clear ideas are expressed as explanatory logic, so they are logically useful. However, existing digital forensic logic and knowledge management models do not contain clear principles and concepts that can be used to build a general knowledge base for digital forensics and security. To solve this study, important theories and hypotheses have been developed that include explanatory logic that can be used reliably. Satpathy et al. [72] investigate whether security has become a major issue in cyberspace. Digital analysis of forensics at a relatively young age has become an unstoppable victim of the rapid advancement of computer technology, so this is possible thanks to new and innovative computer approaches. Digital forensic analysis is unique to forensics because it contains more data from the investigation than mathematical and other types of forensic problems. The digital investigation process can be performed using several forensic models. The Domain Operations Panel of Data Integration reflects on balancing the investigative process, creating high quality data for analysis, and generating legal digital evidence as a court certification expert.

Beebe et al. [73] digital text string search tools use appropriate and/or indexing methods to identify physical sources at the physical level to search for the given text strings. They are designed to provide 100% on-demand memory (e.g. find all examples of text strings). Depending on the nature of the data set, this leads to very high success rates and may not meet demand targets. When websites are similar to search engines, they use a ranking algorithm to get the most efficient and effective search results. This is the reason why digital forensic text search tools fail to determine team and/or search success, which significantly improves (or accelerates) the ability of searchers to succeed.

Beebe et al. [74] investigators and researchers have found that large and terabit-sized data sets are increasingly being used to conduct digital surveys. Sophisticated digital search tools and processes control work from a computer and human perspective because they constantly rely on very simple algorithms to reduce and recover data. Expanding data mining research in the field of digital forensics would have the following or several advantages: (i) reduce data processing and human processing time; (ii) improved quality of data related to data analysis; (iii) Reduce the cost associated with digital claims. It offers recommendations for using digital forensics for data forensics.

Bermad et al. [75] the effectiveness of the mobile investigation process (smartphone forensics) is related to its evidence analysis phase. Depending on this collection and location of all resources, their temporary, functional and associated combinations. The abundance of evidence, their complexity, and the number of correlations between different types of data complicate the analysis phase of the evidence and the reconstruction of the crime. They provide a method of temporal and functional analysis based on data processing (supervisory classification). Introduce new clustering techniques based on dynamic causes and event reconstruction (SMS and calls). In this case, we can help the investigator to identify inconsistencies and information about the crime and to inform everyone of the global outlook on all events. Evidence was collected.

Verma et al. [76] rapid technological changes, the use of digital tools (diversity and numbers), and the large amount of data available on these devices have been found to continue to challenge the current state of digital expertise. Because the protection of confidentiality and the completeness of the investigation are mutually exclusive, digital forensic testing focuses on the effectiveness of the investigative process or the protection of data confidentiality to address these challenges. However, the general approach to data confidentiality remains an open question that does not affect the investigator's capabilities or the overall performance of the intelligence process. Duce et al. [77] examined the rapid emergence of technological development and the challenges of digital forensics in the face of attacks on criminal thinkers that are naturally distributed in natural development. Collect evidence and make assumptions at the Digital Forensic Centre. Marangos et al. [78] CC studied the impact of the model on digital forensic systems. Digital forensics (DF) is a field of computer science that generates relevant evidence in the digital form acceptable and scientifically proven for use in civil or criminal proceedings. Their analysis significantly affects the right to share digital resources inherited from the CC model.

Homem et al. [79] the mobile devices and virtual devices in the Internet of Things will gradually become the endpoints of different networks owned by different parties, while engaging in unlicensed licensing or illegal activities. Digital Forensics and Incident Response (DFIR) tools today are struggling to conduct digital research in a freely controlled online environment because they have a number of issues: lack of resources, availability, reliability, confidentiality, data size, speed, and diversity. They integrate the structural and structural P2B overlays of our architecture to enable efficient capture, analysis. Singh et al. [80] security vulnerabilities in Windows operating systems that have been studied for years have increased significantly, and the most serious threat is malware (malware). Theoretically, the recovery cache stores the unstable memory of the attack process and the affected machine in the Windows registry. Digital Forensic Product Process (DFR) offers suggestions for collecting active digital tracks.

Tewelde et al. [81] the concept of "scientific theory" is even more important in the new field of digital forensics, because not all empiricism is hypothetical. Although the term “concept” is widely used in digital expert narrative, its use is not adequately reflected from a scientific point of analysis. This issue is discussed with special reference to the proper functioning of the carrier where concept plays an important role. Bashir et al. [82] with the regular use of the Internet and the advancement of technology, cyber and malware applications are growing on digital devices. Activities performed electronically can be analyzed using digital forensic methods.

5.4.4 Attack Graph System

Wang et al. [83] proposed a network prototype a skill that integrates the automatic logic of presentation, fabrication, and infiltration evidence. We offer a new map of Evidence Map to help you navigate and process navigation resources. For automated analysis of evidence, we develop a hierarchical basic framework that integrates local logic with global logic. In local logic, we use Rule-based Fuzzy Cognitive Maps (RBFCM) to model the evolution of the status of suspicious hosts. In order to achieve a global goal, we aim to identify the host group that is strongly associated with the attack and gain their connection in the context of the attack. Zhu et al. [84] it has been learned that remote hackers rarely seize network attacks on computers; When an attack is detected successfully or simply, it is often advisable to reconstruct the background of this security breach: all violations and related incidents. They record the latest network traffic data to identify these instances of attack. Their simulations show that the algorithm achieves accuracy in detecting attack patterns. The added benefit of not tracking their attack mode detection, for example, is that it does not require a user-defined gateway. Barik et al. [85] As computer networks evolve day by day, network security has become a major issue. At the same time, attacks become more complex and computer networks become more secure. Attack Map is a modelling tool used to evaluate the security of organizational charts. Since its introduction, significant research efforts have been made to develop a theory and application around the concept of an attack map. Liu et al. [86] it explores how difficult it is to model an effective and accurate security event to determine the attack situation for an enterprise network. In this study we discuss how to create an attack situation and source map using the evidence of security studies. To achieve the accuracy and completeness of the source map, we use introductory induction and hijacking logic to verify the evidence. In addition, as the structural circumstances and the evidence presented may stand in court, federal evidence laws will be taken into account to determine the evidence in advance. Barrère et al. [87] the network has been explored as an effective mechanism for guiding investigators during forensics. To this end, they present the concept of a basic offense map, a brief overview of key pathways that can attack targets on a particular network. Such a summary allow forensic investigators to spotlight on important nodes that may be part of an attack path, thereby minimizing common nodes (devices, network privileges). Conlan et al. [88] Expert anti-expert tools, techniques, and methods were studied, which became a strong barrier to the digital forensic community. Strategies need to be developed to address this growing problem. Al-Mahrouqi et al. [89] achieving high security on low-end and wired communication networks is consistently difficult to achieve. The growing rate of secure network owners and attackers is relatively close.

Ha et al. [90] studied that the capability acquisition graphs (CAGs) that provide a robust framework for modelling internal threats and attacks, and the system vulnerabilities. The CAG-based security modelling systems have not yet been used in practice. In particular, it describes an information-based graphical database tool known as ICMP. Imran et al. [91] the growing nature and increasing demand for the open source cloud is making the infrastructure an ideal target for malicious attacks, unauthorized access to data storage, and a serious threat to cloud software security. In the event of any malicious action, the cloud resource information used by digital forensics to diagnose the problem may fall prey to malicious companies and cloud security software. Liu et al. [92]. Problems with reconstruction of attack scenes include large amounts of data is damaged or destroyed. This model solves these problems using a variety of methods, including evidence-based displays, including computer damage, inductive logic, and reconstruction of attack scenes. In the introduction the system uses known vulnerability databases and expert anti-databases, which extend to standard databases such as the NSD National Vulnerability Database. Created for network criminal analysis, this system reduces the time and effort required to reach final conclusions about network attacks. Peisert et al. [93] Judgments in the field of computer forensics were often considered conditional. It discusses the need for rigorous expertise and outlines the features that should exist in this model. This requires far less carefully selected and effective data than the forensic model. Poolsapassit et al. [94] the computer log files contained vital evidence related to the computer attack. However, journal files are often so large that most of the information in them is irrelevant to the query. Rekhis et al. [95] security measures and the logic of digital inquiry were tested in its high-level specification language. Rekhis et al. [96] studied the analysis and reconstruction of attack processes, and motivates attackers to resist counter-attacks, whether it is difficult, difficult or impossible to defeat the investigative process. The literature has suggested several methods for systematically reconstructing the development of actions that occur during an event use theoretically and scientifically proven methods. However, these methods are not intended to address forensic attacks because they believe that the evidence collected is reliable and that experts have not modelled the operations and have developed samples of attacks, security solutions and resources. Liu et al. [96] learned that attack maps are used to estimate attack paths and known computer damage from a computer configuration. Attack maps can be used to determine the order of damage used to carry out attacks, and forensics can help identify potential attacks.

6 Comparative Analysis of State-of-Art Models

6.1 Intrusion Detection System in Network Forensic Analysis

In this subsection, we analyze the different IDS for digital forensic analysis. It is very irrational to find infiltration as a discipline. Table 1 shows the brief review on IDS for digital forensic analysis. Form state-of-art papers, most of the studies discussed the intrusion detection through attack types which are DoS, DDoS, botnet, portnet and web attack.

Table 1 IDS in forensic analysis

IDS	IP spoofing attacks	SNORT	Accurate authentic	[39]
IDS	DOS: port scan: DDoS	SNORT	Nfinimize false alarm rate	[40]
NIDS	Cloud anti forensic	SNORT	RBNN+, Kmeans+, Correlation	
IDS	DOS Attack	SNORT	NetSTAT- real time network: smurfand SYN flood	[42]
DRS	Dos and DDoS	SNORT	Multi-tenancy and distributed neÄvork feature	[43]
PIDAS	Botnet attack	SMOTE	Potential system vulnerabilities	[44]
AIDE and IDS	DOS attack	SNORT	Augnlented feature and speculative fast	[45]
NIDS	Dos, DDoS, Bomet: Port Scan, Web attack	SMOTE	K-means homogeneity metri feature GBT, DNN, binary and multiclassification	[46]
IDS	Dos, DDoS	SNORT	Artificial intelligence (Al) prediction feature	[47]
IDS	Port Scan	SNORT	NS-3 collected	[48]
IDS	Dos, DDoS	SNORT	Kmeans, ANN	[49]
GUI	Web attack	SMOTE	Time between windows and new windows	[50]
	Dos, DDoS	SMOTE	Multi-tenancy	[51]
NIDS	DOS attack	SNORT	SSENet	[52]
mlRS, IDS and IPS	DOS and DDoS attack	SNORT	No. of TCP/IP connective CPU usage, battery cllarge	[53]

Table 2 Trace back techniques in forensic analysis

Trace-back scheme	Advantages	Disadvantages	Sources
Input debugging	Support Incremental Implementation Can be used against both DOS or DDoS Functions are not necessary to perform	Utilization of time is high DOS or DDoS attack at the same time For a successful trace: the attack must last a long time	[54, 55]
Controlled flooding	Easy to implement Can be used against both DOS or DDoS Suitable for network infrastructure	Utilization of time is high Cannot distinguish DDoS and genuine flash crowed It can only be used in attacks	[56–58]
Logging	Easy to implement Using single packet to reconstruct attack path Allowing post packet analysis	Have potential hash collision Substantial storage required Vulnerable to compromised router	[59–61]
Hop count filtering	Easy to implementation Bandwidth above is very low Less infrastructure support	Long detection cycle Lack of economic incentive Vulnerable to compromised router	[62–64]
Packet marking	It is scalable Low processing Suitable for a variety of attacks	Applying the program is very expensive Costing data fragmentation High bandwidth overhead	[65–67]

6.2 Tracebacks in Network Forensic Analysis

In this section, we discuss the development of effective forensic surveillance technologies for dealing with cybercrime through anonymous communication networks. Associations around the world today rely heavily on the Internet for business, military communications, and information dissemination. However, this has led to cyber security issues. Network expertise plays an important role in supporting legal oversight as the number of cybercrimes increases in the integrated and fast-growing Internet world. Table 2 describes the trackbacks usage in Network Forensic analysis.

6.3 Distributive Systems in Forensic Analysis

Digital forensics is a bough of forensics that deals with the examination of digital crime, which is primarily an attack on digital plans. There are several subdivisions of digital forensics based on objectives, counting computer forensics, network forensics, and mobile forensics. The aptitude to conduct digital research is straight connected to the accessibility of confirmation, chiefly in journal records. It includes databases where researchers can find information about the causes of crime and their source (computer, laptop, and tablet). As a result, it causes the attacker to be found and/or confiscated. In the age of the Internet of Things (IoT), digital forensics will have fun an important role in investigate crime alongside IoT devices (Table 3).

Table 3 Distributive techniques in forensic analysis

Source	Category	Approach	Limitations
[68]	Cognitive mapping	Top down	MACs and digital signature to provide more reliability
[69]	Fuzzy c-means	Hierarchical	Clustering generates subset of clusters or partition
[70]	Gigabit Ethernet	DDF toolkit	DDF toolkit is similar to commercial forensic tools
[71]	Description logics	Ontology based	Language defined for computer reasoning must have rules
[72]	DFRWS framework	Cyberspace based	Comprehension or the need to reduce information overload
[73]	EF database	CPC database	Rules for conducting forensic analysis can be stored in CPC
[74]	Mining algorithm	Social network analysis	Uses of data discrimination like data characterization
[75]	WWW Consortium	Combichem approach	W3C RDF vocabulary description language used
[76]	Cloud computing	RCG based	CSP has no privilege on the upper layer
[77]	Hashing algorithm	Defense in depth	Unstructured P2P overlays in various capacities
[78]	RSA encryption	Trigger-based	RaaS provides a ransomware service
[79]	Notorious Duhem-Quine dilemma	Hypothesis based approach	Carrier’s framework is normative rather than descriptive
[80]	DES algorithm	Live Forensic Analysis	FAT do not support bitlock drive encryption drive
[81]	Clustering algorithm	Kohonen’s self organising map (SOP)	Feasibility at low risk
[82]	CATFA	CATFA approacgh	Scalable with many events because it is parallel

6.4 Attack Graphs in Forensic Analysis

Regardless of momentous efforts to look after network from cyber-attacks (Gartner, Inc. 2014), computer administrators have not been able to cope with modern threat technology, as evidenced by the recent history of corporate infringement (Lord 2015). One of the largely ordinary network protection strategies is to detect and insert risks. Such an approach requires network risk assessment, prioritization of the most important threats, and subsequent evaluation of risk performance (Wheeler 2011). Finally, administrators use these values to select the appropriate commands. But most of the time this analysis is implemented individually for each network component, i.e., ignoring the relationship between the vulnerabilities, i.e. how one vulnerability can be successfully exploited so that the attacker can exploit vulnerability and thus gain exclusive rights across the networkas given in Table 4.

7 Conclusion

In this study, we presented an undemanding framework for digital inquiry based on the causes and consequences of events. The benefits of this work contribute to the development of a common solution that meets the needs of a swiftly altering and greatly impulsive digital technology environment. Gain the honesty and acceptance of digital resources and internetworking devices. The solution to the research questions is as follows. This paper describes in detail the various techniques used in basic Forensic Methodologies with comparisons on all relevant strategies. Network forensics can be classified is based on four technologies: infiltration detection, tracking, distribution, and attack maps.

Table 4 Attack graphs in forensic analysis

Source	Category	Tool	Technique	Vulnerability	Effect
[83]	Attack tool		Clustering technique	Unix	Attack instances are propagated
[84]	Attack tool		RBFCM	Buffer overflow	Analytics system effectively integrates feedback into an automated reasoning process
[85]	Attack tool		Obfuscate signature	SNORT rule	Mislead forensic analysis by crafting image
[86]	Destroy data	BCWipe	Remove log file	CVE-2009-2446	Delete data permanently
[87]	Attack tool	Wireshark	SQL injection attack	GNS3	Was designed to increase the filter-ing process of the output event logs
[88]	Attack tool	Nu SMV tool	ICMAP	CERT	Visualization used as an effective method for integrating and analysing diverse information
[89]	Destroy data	BCWipe	Delete content	CVE-2009-1918	Access control modules
[90]	Attack tool		SQL injection	Buffer overflow	Eliminates info from log file
[91]	Attack tool	CVSS	Steganography	SNORT rule	Provenance detection
[92]	Attack tool		Steganography	Buffer overflow	Bypass being detected by rules
[93]	Attack tool	Appraiser tool	Postgres SQL	Linux	
[94]	Attack tool		AI techniques	Unix	System dependent
[95]	Attack tool	Model checker	I-TLA	SNORT rule	Bypass being detected by rules
[96]	Attack tool		SQL injection	SNORT rule	Attacker could alter the evidence
[96]	Attack tool		Obfuscate signature	SNORT rule	Bypass being detected by rules

References

1. Pilli ES, Joshi RC, Niyogi R (2010) Network forensic frameworks: survey and research challenges. Digit Investig 7(1–2):14–27
2. Mohamed IA, Manaf ABA (2014) An enhancement of traceability model based-on scenario for digital forensic investigation process. In: 2014 third international conference on cyber security, cyber warfare and digital forensic (CyberSec). IEEE, pp 12–15
3. Palmer GL (2001) A road map for digital forensic research. Technical Report DTR-T0010-01, DFRWS. Report for the First Digital Forensic Research Workshop (DFRWS)
4. Sivaganesan DD (2021) A data driven trust mechanism based on blockchain in IoT sensor networks for detection and mitigation of attacks. J Trends Comput Sci Smart Technol 3(1):59–69. https://doi.org/10.36548/jtcsst.2021.1.006
5. Cleland-Huang J, Hayes JH, Domel JM (2009) Model-based traceability. In: Proceedings of the 2009 ICSE workshop on traceability in emerging forms of software engineering. IEEE Computer Society, pp 6–10
6. Carrier B, Spafford E (2004) An event-based digital forensic investigation framework. Digit Investig
7. Agarwal A, Gupta M, Gupta S, Gupta SC (2011) Systematic digital forensic investigation model. Int J Comput Sci Secur (IJCSS) 5(1):118–131
8. Chung H, Park J, Lee S, Kang C (2012) Digital forensic investigation of cloud storage services. Digit Investig 9(2):81–95
9. Köhn M, Olivier MS, Eloff JH (2006) Framework for a digital forensic ınvestigation. In: ISSA, pp 1–7
10. Perumal S (2009) Digital forensic model based on Malaysian investigation process. Int J Comput Sci Netw Secur 9(8):38–44
11. Ademu IO, Imafidon CO, Preston DS (2011) A new approach of digital forensic model for digital forensic investigation. Int J Adv Comput Sci Appl 2(12):175–178
12. Delport W, Köhn M, Olivier MS (2011) Isolating a cloud instance for a digital forensic investigation. In: ISSA
13. Perumal S, Norwawi NM, Raman V (2015) Internet of Things (IoT) digital forensic investigation model: top-down forensic approach methodology. In: 2015 fifth ınternational conference on digital ınformation processing and communications (ICDIPC). IEEE, pp 19–23
14. Valjarevic A, Venter HS (2012) Harmonised digital forensic investigation process model. In: 2012 ınformation security for South Africa. IEEE, pp. 1–10
15. Kanellis P, Kiountouzis E, Kolokotronis N, Martakos D (2006) Digital crime and forensic science in cyberspace. Idea Group Publishing, London, UK
16. Kebande VR, Ray I (2016) A generic digital forensic investigation framework for internet of things (IoT). In: 2016 IEEE 4th ınternational conference on future ınternet of things and cloud (FiCloud). IEEE, pp 356–362
17. Sathwara S, Dutta N, Pricop E (2018) IoT forensic a digital investigation framework for IoT systems. In: 2018 10th ınternational conference on electronics, computers and artificial ıntelligence (ECAI). IEEE, pp 1–4
18. Shrivastava G, Gupta BB (2014) An encapsulated approach of forensic model for digital investigation. In: 2014 IEEE 3rd global conference on consumer electronics (GCCE). IEEE, pp 280–284
19. Valjarevic A, Venter HS, Ingles M (2014) Towards a prototype for guidance and implementation of a standardized digital forensic investigation process. In: 2014 ınformation security for South Africa. IEEE, pp 1–8
20. Omeleze S, Venter HS (2013) Testing the harmonised digital forensic investigation process model-using an Android mobile phone. In: 2013 ınformation security for South Africa. IEEE, pp 1–8
21. Kebande VR, Karie NM, Venter HS (2016) A generic digital forensic readiness model for BYOD using honeypot technology. In: 2016 IST-Africa week conference. IEEE, pp 1–12

22. Burrows C, Zadeh PB (2016) A mobile forensic investigation into steganography. In: 2016 international conference on cyber security and protection of digital services (Cyber Security). IEEE, pp 1–2
23. Charles T, Pollock M (2015) Digital forensic investigations at universities in South Africa. In: 2015 second ınternational conference on ınformation security and cyber forensics (InfoSec). IEEE, pp 53–58
24. Kao DY, Wu NC, Tsai F (2019) The governance of digital forensic investigation in law enforcement agencies. In: 2019 21st ınternational conference on advanced communication technology (ICACT). IEEE, pp 61–65
25. Jain N, Kalbande DR (2015) Digital forensic framework using feedback and case history keeper. In: 2015 ınternational conference on communication, ınformation & computing technology (ICCICT). IEEE, pp 1–6
26. Mouhtaropoulos A, Grobler M, Li CT (2011) Digital forensic readiness: an insight into governmental and academic initiatives. In: 2011 European ıntelligence and security ınformatics conference. IEEE, pp 191–196
27. Kyei K, Zavarsky P, Lindskog D, Ruhl R (2012) A review and comparative study of digital forensic investigation models. In: International conference on digital forensics and cyber crime. Springer, Berlin, Heidelberg, pp 314–327
28. Boateng R et al (2010) Cyber crime and criminality in Ghana: its forms and implications. In: Proceedings of the 16th Americas conference on ınformation systems
29. Smith RG, Grabosky PN, Urbas G (2004) Cybercriminals on trial. Cambridge University Press. ISBN: 9780521840477
30. Kent K, Chevalier S, Grance T, Dang H (2006) NIST SP 800-86 guide to ıntegrating forensic techniques into ıncident response
31. Bunting S (2007) Mastering Windows network forensic and investigation, 1st edn. Sybex. ISBN-13: 978-0470097625
32. Volonino L, Anzaldua R, Godwin J (2007) Computer forensics: principles and practices. Pearson Prentice Hall, NJ, USA
33. Jiao F, Gao W, Duan L, Cui G (2001) Detecting adult ımage using multiple features. In: Proceedings of the ICII 2001, vol 3, pp 378–383
34. Phung S, Chai D, Bouzerdoum A (2003) Adaptive skin segmentation in color ımages. In: IEEE international conference on acoustics, speech, and signal processing (ICASSP 2003)
35. Cosic J, Baca M (2010) A framework to (im)prove "chain of custody" in digital ınvestigation process. In: Proceedings of the CECIIS, Varazdin, Croatia
36. Hemdan EED, Manjaiah DH (2018) Cybercrimes investigation and intrusion detection in internet of things based on data science methods. In: Cognitive computing for big data systems over IoT. Springer, Cham, pp 39–62
37. Pilli ES, Joshi RC, Niyogi R (2010) An IP traceback model for network forensics. In: International conference on digital forensics and cyber crime. Springer, Berlin, Heidelberg, pp 129–136
38. Quick D, Choo KKR (2017) Big forensic data management in heterogeneous distributed systems: quick analysis of multimedia forensic data. Softw: Pract Exp 47(8):1095–1109
39. Chhabra GS, Singh P (2015) Distributed network forensics framework: a systematic review. Int J Comput Appl 119(19)
40. Achille MM, Roger AE (2014) Obtaining digital evidence from intrusion detection systems. Int J Comput Appl 95(12)
41. Barhate K, Jaidhar CD (2013) Automated digital forensic technique with intrusion detection systems. In: 2013 3rd IEEE ınternational advance computing conference (IACC). IEEE, pp 185–189
42. Rani DR, Geethakumari G (2020) A framework for the identification of suspicious packets to detect anti-forensic attacks in the cloud environment. Peer-to-Peer Netw Appl 1–14
43. Singhal A, Jajodia S (2006) Data warehousing and data mining techniques for intrusion detection systems. Distrib Parallel Databases 20(2):149–166

44. Inayat Z, Gani A, Anuar NB, Anwar S, Khan MK (2017) Cloud-based intrusion detection and response system: open research issues, and solutions. Arab J Sci Eng 42(2):399–423
45. Xie Y, Feng D, Tan Z, Zhou J (2016) Unifying intrusion detection and forensic analysis via provenance awareness. Future Gener Comput Syst 61:26–36
46. Sy BK (2009) Integrating intrusion alert information to aid forensic explanation: an analytical intrusion detection framework for distributive IDS. Inf Fusion 10(4):325–341
47. Zhang H, Huang L, Wu CQ, Li Z (2020) An effective convolutional neural network based on SMOTE and Gaussian mixture model for ıntrusion detection in ımbalanced dataset. Comput Netw 107315
48. Rahouma K, Ali A (2019) Applying intrusion detection and response systems for securing the client data signals in the Egyptian optical network. Procedia Comput Sci 163:538–549
49. Aloqaily M, Otoum S, Al Ridhawi I, Jararweh Y (2019) An intrusion detection system for connected vehicles in smart cities. Ad Hoc Netw 90:101842
50. Houmansadr A, Zonouz SA, Berthier R (2011) A cloud-based intrusion detection and response system for mobile phones. In: 2011 IEEE/IFIP 41st international conference on dependable systems and networks workshops (DSN-W). IEEE, pp 31–32
51. Yampolskiy RV (2007) Human computer interaction based intrusion detection. In: Fourth ınternational conference on ınformation technology (ITNG'07). IEEE, pp 837–842
52. Zhang Y, Saberi M, Chang E (2018) A semantic-based knowledge fusion model for solution-oriented information network development: a case study in intrusion detection field. Scientometrics 117(2):857–886
53. Vasudevan AR, Selvakumar S (2016) Local outlier factor and stronger one class classifier based hierarchical model for detection of attacks in network intrusion detection dataset. Front Comp Sci 10(4):755–766
54. Ulltveit-Moe N, Oleshchuk VA, Køien GM (2011) Location-aware mobile intrusion detection with enhanced privacy in a 5G context. Wirel Pers Commun 57(3):317–338
55. Armoogum S, Mohamudally N (2008) Mobile agent based real time IP traceback. In 2008 third ınternational conference on digital ınformation management. IEEE, pp 69–74
56. Cusack B, Tian Z, Kyaw AK (2016) Identifying DOS and DDOS attack origin: IP traceback methods comparison and evaluation for IoT. In: Interoperability, safety and security in IoT. Springer, Cham, pp 127–138
57. Selamat SR, Yusof R, Sahib S, Hassan NH, Abdollah MF, Abidin ZZ (2011) Traceability in digital forensic investigation process. In: 2011 IEEE conference on open systems. IEEE, pp 101–106
58. Alex ME, Kishore R (2017) Forensics framework for cloud computing. Comput Electr Eng 60:193–205
59. Shah JJ, Malik LG (2014) An approach towards digital forensic framework for cloud. In: 2014 IEEE ınternational advance computing conference (IACC). IEEE, pp 798–801
60. Koleoso RA, Ajayi OO, Oladeji FA, Uwadia CO (2018) A digital forensics investigatıon model that incorporates digital chain of custody for its integrıty, and authenticity. Computer Networks, Infrastructure Management and Securıty (CoNIMS), p 55
61. Bulbul HI, Yavuzcan HG, Ozel M (2013) Digital forensics: an analytical crime scene procedure model (ACSPM). Forens Sci Int 233(1–3):244–256
62. Chlapoutakis G (2022) Use of network monitoring and analysis tools and methodologies in digital forensic ınvestigations
63. Kwon S, Jeong J, Shon T (2019) Digital forensic readiness for financial network. In: 2019 ınternational conference on platform technology and service (PlatCon). IEEE, pp 1–4
64. Montasari R (2016) An ad hoc detailed review of digital forensic investigation process models. Int J Electron Secur Digit Forens 8(3):205–223
65. Rekhis S, Krichene J, Boudriga N (2008) DigForNet: digital forensic in networking. In: IFIP ınternational ınformation security conference. Springer, Boston, MA, pp 637–651
66. Singh KS, Irfan A, Dayal N (2019) Cyber forensics and comparative analysis of digital forensic ınvestigation frameworks. In: 2019 4th ınternational conference on ınformation systems and computer networks (ISCON). IEEE, pp 584–590

67. Jain N, Kalbande DR (2014) A comparative study based digital forensic tool: complete automated tool. Int J Forens Comput Sci
68. Wazid M, Katal A, Goudar RH, Rao S (2013) Hacktivism trends, digital forensic tools and challenges: a survey. In: 2013 IEEE conference on ınformation & communication technologies. IEEE, pp 138–144
69. Trček D, Abie H, Skomedal Å, Starc I (2010) Advanced framework for digital forensic technologies and procedures. J Forens Sci 55(6):1471–1480
70. Singh AP, Jain RC (2015) Group identification based classification technique for aggregated common data used in digital forensic ınvestigation (DFI)
71. Roussev V, Richard III G (2004) Breaking the performance wall—the case for distributed digital forensics. Digit Investig
72. Ellison D, Ikuesan AR, Venter H (2019) Description logics and axiom formation for a digital forensics ontology. In: European conference on cyber warfare and security. Academic Conferences International Limited, pp 742-XIII
73. Satpathy S, Pradhan SK, Ray BB (2010) A digital investigation tool based on data fusion in management of cyber security systems. Int J Inf Technol Knowl Manag 2(2):561–565
74. Beebe NL, Clark JG (2007) Digital forensic text string searching: improving information retrieval effectiveness by thematically clustering search results. Digit Investig 4:49–54
75. Beebe N, Clark J (2005) Dealing with terabyte data sets in digital investigations. In: IFIP ınternational conference on digital forensics. Springer, Boston, MA, pp 3–16
76. Bermad N, Kechadi MT (2016) Evidence analysis to basis of clustering: approach based on mobile forensic investigation. In: 2016 7th ınternational conference on sciences of electronics, technologies of ınformation and telecommunications (SETIT). IEEE, pp 300–307
77. Verma R, Govindaraj J, Gupta G (2018) DF 2.0: designing an automated, privacy preserving, and efficient digital forensic framework
78. Duce D, Mitchell F, Turner P (2007) Digital forensics: challenges and opportunities. In: 2nd conference on advances in computer security and forensics (ACSF), LJMU, Liverpool, UK
79. Marangos N, Rizomiliotis P, Mitrou L (2012) Digital forensics in the cloud computing era. In: 2012 IEEE Globecom workshops. IEEE, pp 775–780
80. Homem I, Kanter T, Rahmani R (2016) Improving distributed forensics and incident response in loosely controlled networked environments. Int J Secur Appl 10(1):385–414
81. Singh A, Ikuesan AR, Venter HS (2018) Digital forensic readiness framework for ransomware investigation. In: International conference on digital forensics and cyber crime. Springer, Cham, pp 91–105
82. Tewelde S, Gruner S, Olivier M (2015) Notions of hypothesis in digital forensics. IFIP international conference on digital forensics. Springer, Cham, pp 29–43
83. Bashir MS, Khan MNA (2013) Triage in live digital forensic analysis. Int J Forens Comput Sci 1:35–44
84. Wang W, Daniels TE (2005) Network forensics analysis with evidence graphs (demo proposal). In: Proceedings of the digital forensic research workshop
85. Zhu Y (2011) Attack pattern discovery in forensic investigation of network attacks. IEEE J Sel Areas Commun 29(7):1349–1357
86. Barik MS, Sengupta A, Mazumdar C (2016) Attack graph generation and analysis techniques. Def Sci J 66(6):559
87. Liu C, Singhal A, Wijesekera D (2014) A model towards using evidence from security events for network attack analysis. In: WOSIS, pp 83–95
88. Barrère M, Steiner RV, Mohsen R, Lupu EC (2017) Tracking the bad guys: An efficient forensic methodology to trace multi-step attacks using core attack graphs. In: 2017 13th international conference on network and service management (CNSM). IEEE, pp 1–7
89. Conlan K, Baggili I, Breitinger F (2016) Anti-forensics: furthering digital forensic science through a new extended, granular taxonomy. Digit Investig 18:S66–S75
90. Al-Mahrouqi A, Abdalla S, Kechadi T (2015) Cyberspace forensics readiness and security awareness model. Int J Adv Comput Sci Appl 6:123–127

91. Ha D, Upadhyaya S, Ngo H, Pramanik S, Chinchani R, Mathew S (2007) Insider threat analysis using information-centric modeling. IFIP international conference on digital forensics. Springer, New York, NY, pp 55–73
92. Imran A, Aljawarneh S, Sakib K (2016) Web data amalgamation for security engineering: digital forensic investigation of open source cloud. J UCS 22(4):494–520
93. Liu C, Singhal A, Wijesekera D (2015) A logic-based network forensic model for evidence analysis. IFIP international conference on digital forensics. Springer, Cham, pp 129–145
94. Peisert S, Bishop M, Karin S, Marzullo K (2007) Toward models for forensic analysis. In: Second international workshop on systematic approaches to digital forensic engineering (SADFE'07). IEEE, pp 3–15
95. Poolsapassit N, Ray I (2007) Investigating computer attacks using attack trees. IFIP ınternational conference on digital forensics. Springer, New York, NY, pp 331–343
96. Rekhis S, Boudriga N (2011) Logic-based approach for digital forensic investigation in communication Networks. Comput Secur 30(6–7):376–396

IOT Based Solution for Effective Social Distancing and Contact Tracing for COVID-19 Prevention

S. Kanakaprabha, P. Arulprakash, V. Priyanka, Vineetha Varghese, and A. Sureshkumar

Abstract Coronavirus has infected billions of individuals worldwide, with the number of persons infected continuing to rise. Humans contract the virus through direct, indirect, or close contact with infected individuals. This proposed work introduces a new feature, an intelligent community distance system, that allows people to maintain community distances among others in the indoor and outdoor areas, to avoid exposure to COVID-19 and to delay its spread locally and internationally, to help prevent the spread of COVID-19. The proposed research intends to monitor an IoT-based portable monitoring device that is designed to measure COVID-19 signals. Furthermore, by monitoring real-time GPS data, the system automatically notifies medical authorities concerned about any confinement violations of patients who may be infected. Also, figure out what new tool will be beneficial for tracking and predicting COVID-19 collections. To support in the analysis of COVID-19, the solution incorporates a mobile system coupled with a portable device that is equipped with clever IoT capabilities (complex data analysis and intelligent data detection) embedded within the system. A comparison of various machine learning classifier algorithms such as SVM, Random Forest, KNN, and Decision Tree is presented as

S. Kanakaprabha (✉) · S. Kanakaprabha (✉) · P. Arulprakash · V. Priyanka · V. Varghese · A. Sureshkumar
Department of Computer Science and Engineering, Rathinam Technical Campus, Anna University, Coimbatore, India
e-mail: kanakaprabha.cse@rathinam.in

P. Arulprakash
e-mail: rtchod.cse@rathinam.in

V. Priyanka
e-mail: priyanka.cse@rathinam.in

V. Varghese
e-mail: vineetha.cse@rathinam.in

A. Sureshkumar
e-mail: suresh.cse@rathinam.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*, Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_46

the best model for making predictions and determining accuracy. We observed that KNN performs better, with a 95% accuracy rate. COVID-19 will be used to prevent the spread of diseases in future global medical problems using an automatic social distance monitoring and contact tracking system.

Keywords Coronavirus · Decision tree · Global positioning system · Internet of things · Support vector machine

1 Introduction

COVID-19 is a respiratory disease that is very infectious (SARS-CoV-2), which is very dangerous infectious virus. The respiratory system is attacked by SARS-CoV-2, and the flu virus, which causes symptoms such as poisoning, fever, nausea and shortness of breath. To prevent the epidemic, many countries have it did Lockdown when the government forced it for residents to stay home during this critical time. Community health facilities, such as the Centers for Disease Control and Prevention (CDC), had to make it clear that avoiding close contact with other persons is the most efficient strategy to prevent the transmission of Covid-19. Surrounding dwellers of the earth have become accustomed to physical work in order to flatten the curves to the Covid-19 outbreak. Implementing community activities, team activities as well congregations like travel, meetings, gatherings, Workshops, prayers were banned during incarceration time. People are encouraged to use telephone and email manage and run events as much as possible to minimize personal contact. Continuing to contain the spread of virus, Adults are also recommended to follow good hygiene habits such as hand washing, wearing masks, and avoiding close contact with sick people. There is, however, a distinction between knowing what to do to minimize virus infection and truly doing it [1]. Using public broadcasts among persons at least one meter apart is one well-known method of avoiding the spread of Covid-19. The coronavirus is spread from person to person through tiny droplets from the nose and mouth, according to WHO. To put it another way, social dispersal is the greatest strategy for limiting physical contact with probable corona-virus carriers by keeping at least a meter between people. The proposed work is to support Covid-19 antiretroviral measures. It provides a solution for finding people who gather in public places such as shopping malls, banks, market, temples, mosques, public transport and government offices etc. [2]. Monitoring and support of social isolation has evolved important in reducing the risk of internal transmission of disease crowded indoor areas [3].

Figure 1 shows in public areas, social precautions such as contrasting signs on the seats deter individuals from sitting next to each other. Social isolation and face mask is the best form of protection used in modern times. The rules set by the World Health Organization (WHO) provide a detailed review of external activities that one must strictly follow: WHO follow-up rules: (1) The face mask/cover must be worn before exposure. (2) A person leaving his or her private area must ensure that public

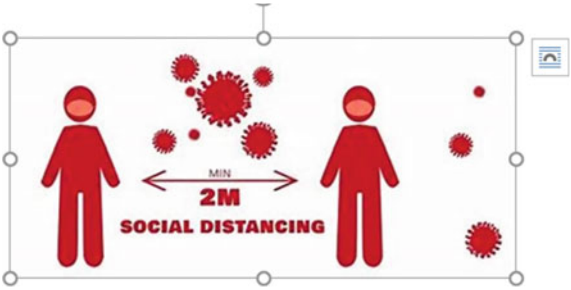


Fig. 1 Social distancing [4]

distance is maintained during his or her time in public places. As a result, there is no permanent solution in the absence of medical assistance. Few efforts are being made around the world to use IOT statistics as a preventative or predictable method against Covid-19, or as models in aspects of epidemiology. With the use of a food security camera that can analyses live or offline to detect social divisions and face mask coverage in the official, public, or public area, a vision tool for controlling social distance and face mask at work has been developed, will aid in the monitoring of health policies The Practicing Health protocol helps living longer and control rashes [5].

The spread of the COVID-19 epidemic has had a devastating effect on human mobility patterns as social behavior related to daily mobility. There is a need to understand the patterns of disease spread and its mechanisms among neighboring people in order to implement corrective measures at this time. To enhance the effectiveness of social media tracking, countries around the world are advancing advances in mobile technology and Internet of Things (IoT) to help keep track of common contacts to track people close to identified patients with COVID-19. Even with the advent of vaccine treatment the COVID-19 management approach will be continued until 2021, look to the foreseeable future in terms of digital communication that is an integral part of the response and use of preventive measures such as social isolation, and the use of masks. After several months of use of digital contact tracking technology, in-depth details of the suitability of the various methods and the usability, privacy, and trade-offs of ethical principles involved. In this propose, we provide a comprehensive analysis of digital communication tracking solutions in terms of their processes and technologies due to new emerging data on international information for the distribution of digital contact tracking technology. Contact tracking apps should establish data collection and data interpretation. Figure 2 shows contact tracing areas in public places.

2 Related Work

Sengupta et al. [6] propose the need to develop a plan to respond to environmental outbreaks to assist in tracking and tracking safety-related concerns in industrial

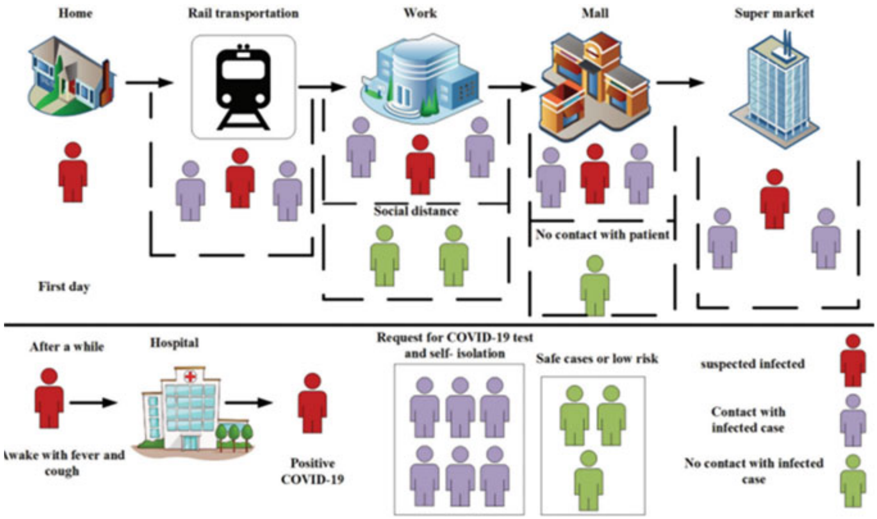


Fig. 2 Contact tracing areas in public places

and community setting. Controlling infectious diseases and their spread necessitates complete communication. To track human activity, the framework will incorporate video feeds from surveillance cameras and IoT edge devices located in industrial or public spaces. The architecture suggested here is a hybrid method to integrating feeds from existing cameras and IoT devices with cloud-based computer-based edges.

S. Srinivasan et al. suggests a complete and effective solution detection of a person, detection of a violation of social distance, detection of a face-to-face mask separation using object detection, integration and Convolution Neural Network (CNN) based on the binary category. In this case, YOLOv3, a local collection based on congestion audio applications (DBSCAN), Dual Shot Face Detector (DSFD) and MobileNetV2 based Binary classifier used for monitoring video databases. They done by comparative research for different facial expressions and facial mask classifications models. Finally, a video database labelling and labelling method is suggested Video data set to compensate for public data shortages and is used system testing. System performance is accurately assessed, Fl score as and predictive timing, which should be low enough to be applied in a practical way [7].

Khanfor et al. [8] to provide a comprehensive framework for improving pedestrian safety while roaming the real-world map of the smart city using the notion of Social IoT (SIoT).The purpose is reducing the risk of infection in very densely populated areas where social distance may not be properly addressed. Proposed the walkway recommends a pedestrian route in real time method while processing the movement of other devices. First, IoT devices are grouped into communities based on two SIoT relationships that took into consideration device locations as well as friendship norms among their owners. As a result, the weights on the city map roadways represented their safety standards. Afterwards, they utilize the Dijkstra algorithm, which is a navigation algorithm, to recommend the safest route to go. Imitation effects were applied in the real-world IoT data set demonstrates this capability a proposed approach to achieving a trade between the two most secure and shorter routes depending on pedestrian preferences.

This paper develops [9] a signal processing framework that enables for integrated topic movement analysis as well as automatic temperature testing. The system includes infrared-based sensors that use temperature data to track subject mobility and health. The sensors are connected to the network by existing IoT wireless devices planted according to different structures. The goal of this project is to link the local action of the headers by tracing their equal distance and route of arrival, as well as the remarkable discovery of body temperature in subjects near the IR sensors. Focus by the Bayesian methods, this paper also discusses good practices as well appropriate application implementation using field standards. For privacy neutrality, the proposed framework can be applied to it public and private services for health care, intelligent living and sharing space conditions without concern for privacy.

W. Lv et al. propose an extended and unauthorized blockchain protocol, called by chain. Specifically, (1) the SRC confidentiality of the SRC protocol and the corresponding block structure are upgraded, by linking the anonymous evidence-based protocol with the key security mechanism. As a result, the connection between personal identity and on-chain location information is severed. At that point, the owner of the on-chain property may still claim ownership without disclosing the private key to anyone else. (2) Proposed field-based practice of incentives to propose to encourage IoT witnesses to advance the oversight of the monitoring industry. The suggested communication tracking and location verification technique works effectively in the actual world, according to several results. The suggested contract tracking protocol's power consumption, time delay for each procedure, and BLE performance have all been studied to ensure the availability of tracking of digital communications in the actual world [10].

S. Arun Kumar et al. study proposes the concept of a smart wrist band with a heat sensor and IoT technology as a preventative approach. Blood pressure measures are also taken from time to time with the use of a blood pressure sensor. As a result, once the temperature or blood pressure are determined to be abnormal, this device helps to generate an alarm. With the use of IoT technology, faster information is transmitted on to the basic level user and second level relatives. As a result, by monitoring and notifying victims, this joint and active wrist band plays a critical role in saving lives. Because bacterial infections are linked to a rise in body temperature, our device will be extremely useful in detecting them early. Traditional measurement methods frequently necessitate human participation and are not of combined size. These issues are addressed in this model, which uses minimal control and sensor settings to handle temperature and blood pressure measurement, data processing, and storage [11].

Chloros and Ringas [12] the goal of this article is to look at the constraints and opportunities that applications that track transmission face, as well as how IoT systems might be used to record symptoms. The benefits and need for these applications' development will be highlighted by evaluating their potential. The Fluspot application was created specifically for this study. By boosting public awareness and providing timely information, Fluspot hopes to help prevent the spread of infections this season. Fluspot uses a wearable IoT device to carefully monitor flu flows and collect user inputs for viral propagation to the site. This anonymised and aggregated data is displayed on a map to provide a more accurate picture of the situation in each location. Another key element is that the artefact's ability to monitor wearable signals is critical for users when it comes to using it in their daily lives.

Waheed and Shafi [13] the study examines a number of technologies that are employed in a range of situations, including social isolation and prevention, isolation and isolation, COVID detection and evaluation, therapy and patient care, and hospital management. This study discusses transparent planning, technological techniques, and digital procedures, as well as the most up-to-date intelligent technologies in a range of disciplines that can aid in overcoming coronavirus intensity. IoT, AI, and machine learning play a significant role in the fight against COVID-19.AI has contributed significantly to the epidemic of resource management, public awareness, management of security measures, and assisting professionals in enforcing strict rules.

Shubina et al. tracking of wearable contact is gaining more attention in the COVID-19 era in order to effectively prevent disease. Therefore, it is timely to identify available solutions for tracking wireless communication and their wearability. Existing contact monitoring app trading necessitates a detailed examination of technical skills such as accuracy, power consumption, availability, error sources when dealing with wireless channels, privacy concerns, and hurdles to larger apparel market access. We find, based on considerable literature study, that demarcated buildings, when compared to intermediate techniques, provide a better place to trade in terms of precision and user desire to use them, taking privacy considerations into account. This paper provides a brief overview of the technical solutions available for human tracking services, outlines key principles that affect the effectiveness of digital communication tracking, and presents a discussion of the potential impact of wear on coping with the spread of viral infections [14].

Luo et al. [15] proposes a model for the spread of infectious, contagious, infectious, asymptomatic disease, Diagnosis, Death (SEINRHD). The model was created using epidemiological data from COVID-19 in China and the estimation of social network heterogeneity. The original Wuhan public epidemic was recreated and updated with accurate data. We used this model to look into ways to manage the outbreak in instances when three-dimensional signals were not visible. The impact of undetectable cases on the spread of the epidemic was assessed based on effective replication rates, incidence of unusually high infections, and the type and structure of transmission. Management of asymptomatic cases can help reduce the curve of infection. Tracking of 75% of non-symptomatic cases is associated with a total reduction of 32.5% in new cases compared to asymptomatic and non-symptomatic tracking). The control and prevention of disease in families should be emphasized during the epidemic.

Sungheetha [16] the ongoing COVID-19 pandemic, there is a lot of interest in using smartphone-based Digital Proximity Tracing Technology (DPTT) for mitigation, containment, and monitoring due to the technology's effectiveness and population acceptance. This research compares the Data-Driven Epidemic Intelligence Strategies (DDEIS) and DPTTs to highlight their differences and provide a fresh approach to address them. The essential components of DDEIS may be included to offer a social as well as technological solution for reducing the risk of epidemic revival, ensuring public health safety, and hastening the return of cities to normalcy. While evaluating its drawbacks and advantages for both societal decision-making and private decision-making, human behaviour is taken into account.

3 Proposed Method

The proposed system consists of the following sub-systems: (1) Arduino Uno-based temperature measurement system (2) IoT system that measures the Arduino board via social distancing (3) on the server (4) A security guard smartphone application. To begin, everyone attempting to enter the residence must pass an unmodified temperature check. We're utilising an Arduino Uno with an infrared thermometer (e.g., MLX906148) or a hot camera sensor for this (AMG88339 for example). It also employs the ESP8266 Wi-Fi module for MQTT protocol connection with Edge servers. If that person's body temperature is abnormal, the door is locked, and a MQTT message with the temperature and the place where it was recorded is sent to the server. This message is received by the server, which then transfers it and carries on. with a security guard's smartphone app, so they can arrive and make sure the person isn't attempting to enter a work zone. In specific regions, Arduino board devices verify whether public distances are being utilised properly or not. Similarly, when public distances do not work well in particular rooms, a MQTT message will be sent to alert security personnel. The MQTT broker and the triple semantic store are used on the server side to provide message processing, event logging, reflection, and message transfer. Edge servers receive communications, do semantic annotations, and make assumptions to determine the appropriate security guard to notify. Security personnel use a basic Android smartphone app. Figure 3 is representing social distancing using smart device for Covid-19 is a programme that strives to guarantee that COVID-19 safety rules are followed correctly indoors.

The proposed calculation will successfully give an answer for physical separating utilizing the ultrasonic sensor. The pseudo-code of the proposed calculation is referenced as follows (Fig. 4).

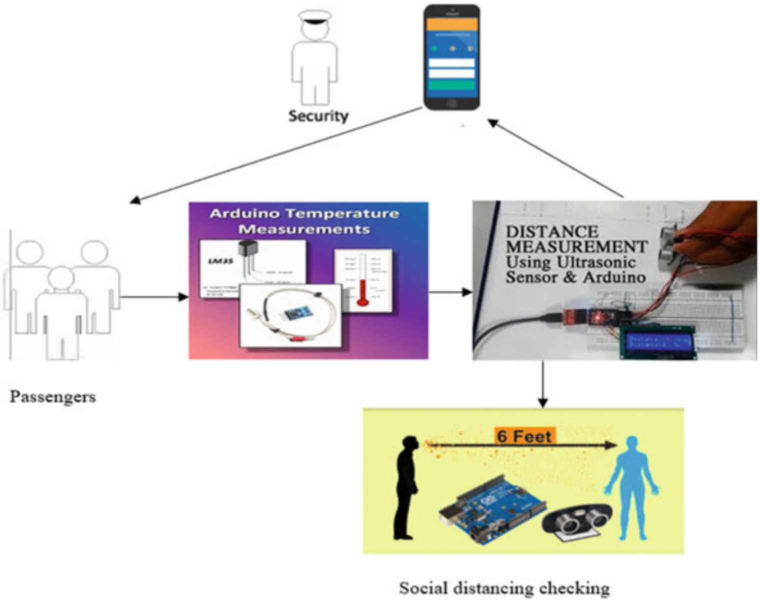

Security
Arduino Temperature Measurements
LM35
DISTANCE MEASUREMENT Using Ultrasonic Sensor & Arduino
Passengers
6 Feet
Social distancing checking

Fig. 3 Overall architecture of social distancing

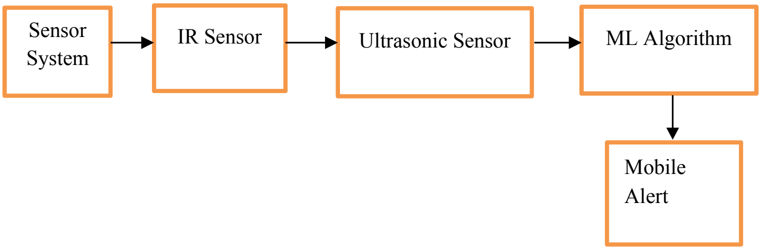


Fig. 4 Social distancing using smart device for COVID-19

3.1 IR Sensor

There are two sorts of temperature measurement tools: touch and non-touch. Thermocouples, heat-resistant heat exchangers (RTDs), thermistors, and semiconductor temperature sensors are examples of infrared temperature sensors used in communication equipment. Because contact lenses measure temperature, they require physical contact with the object being measured to bring the sensor body up to temperature. When a relatively large sensor meets a small object and functions as a heat sink, the temperature of the object can be altered. Figure 5 IR sensor respectively.

Fig. 5 IR sensor

3.2 Ultrasonic Sensor

Ultrasonic sensors use soundwaves to measure distance. The sensor head sends out an ultrasonic wave that is reflected back to it from the direction. Ultrasonic sensors use the time between output and reception to calculate the distance to the target. The ultrasonic sensor can help us to identify that are far away from the robot. The ultrasonic sensor, unlike the touch sensor, is not affected by physical contact. The range gives you plenty of room to react. For distances of 10 in. or more, an ultrasonic sensor is typically used, while for shorter distances, a light sensor is typically employed. ultrasonic sensors detect moving objects, and measure the relative position and movement of each object. The movement in the measuring area of each ultrasonic sensor is measured using a modified distance data conversion, and the vertical movement is measured using a measurement of the measurement range. Figure 6 showing ultrasonic sensor.

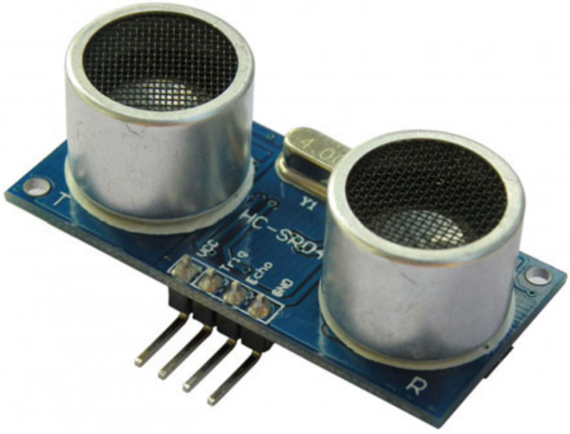
4 Social Distancing Algorithm

It is the second step of our framework proposal. The suggested algorithm for measuring social distance serves two purposes. Function 1 aids in the identification of things in an image. It uses a detection method and provides human locations in the form of aggregate values such as XA (left), YA (top), XB (right), and YB (bottom).

$$X = (X_A + X_B)/2 \tag{1}$$

$$Y = (Y_A + Y_B)/2 \tag{2}$$

Fig. 6 Ultrasonic sensor



where XA, YA, XB, and YB are compound numbers (left, top, right, bottom) of an item. X and Y are coin or centroid values. In addition, these parameters are transferred to the next function to measure social distance. Equation 3 shows the distance between the two items using the Euclidean distance, which establishes how close they are. When comparing this distance vector to the previously indicated threshold value, the decision was taken. If the Euclidean range is below a certain limit, it would be assumed that the two elements did not comply with the terms of social reduction or did not create sufficient distance between them.

$$\mathrm{D} = \sqrt{(X2 - X1)2} + (Y2 - Y1)2 \tag{3}$$

where (X1, X2) and (Y1, Y2) are Centroid values of two objects.

Pseudocode

To identify human/object in the scope of physical removing.

```
1. Input: Human/object in a scope of ultrasonic sensors.
2. output: Alerting a sound sign to the client.
3. Faculties the presence of people utilizing the ULTRASONIC sensor module.
4. On the off chance that a human/moving item is identified, check for the distance.
5. On the off chance that the distance is under 2 m, alert the sound directive for the
   distance in particular.
6. In the event that the distance is equivalent to 1.5 m, ready message and caution
   for safe separating and disinfection.
7. The message will be gone on till the individual isn't cleaning. After the
   disinfection, individual needs to press the reset button.
```

4.1 Algorithm: Social Distancing Measurement

```
Input: $I_N$: Image I containing N Number of frames of size 225x225x3
Output :D: Distance between two objects
Initialize Parameter:
Threshold = 90.0,        //Distance between two objects
Human_Count = 0;     // Number of persons present within input scene default
value = 0
Cvo = [],                    // Coordinate value of object
Lcvo = [],                  // List of Coordinate value of objects
Cen = [],                    // Centroid value of object
Center = [],               // List of Centroid values of objects
Function Object_Coordinates(V)
Picks = Human_Detection_Framework (VN) // provides the number of objects
with their
                                                              coordinate values.
For ($X_A$,$Y_A$,$X_B$,$Y_B$) in Pick :
Cvo = [$X_A$,$Y_A$,$X_B$,$Y_B$]
C1 = (($X_A$ + $X_B$) * 0.5)                    //Centroid Value for Coordinate X
C2 = (($Y_A$+$Y_B$)*0.5)                        // Centroid Value for Coordinate Y
Cen = [C1,C2]
Center.append(Cen)
Lcvo.append(Cvo)
Human_Count +=1
End For
Return Human_Count, Lcvo,center,Image
End Function
```

4.2 Temperature Checking

Using a contactless IR sensor, the temperature-checking device based on Arduino Uno detects the passenger's temperature. One by one, the passengers go. If the temperature of the passenger is higher than the average human body temperature (37 °C), the Arduino Uno generates a signal to lock the door, preventing the person from entering the building, and sends a MQTT message indicating that a person with a high body temperature has been detected at a specific location. Otherwise, the door is opened to welcome the visitor inside.

5 Result and Analysis

The IoT-based portable monitoring device is meant to measure the signals associated with COVID-19 and uses a machine learning model to anticipate the various machine learning techniques. Machine Learning techniques to increase the model's accuracy and impact, as well as to avoid disease transmission in future global health issues.

Figure 7 showing the various machine Learning algorithms so, With a 95% accuracy, KNN offers the best combination of performance values. The performance of two persons was examined in the remote sensing test, with the number of people within the distance view expected to decrease as the number of people increases. The performance of the distance monitor varies with the distance of objects from the camera, as it changes with respect to the measurement originally calculated between pixels and meters. The above graph shows an SVM accuracy is 60%, the Random forest accuracy is 70%, KNN accuracy is 98% and the decision tree accuracy is 50 respectively.

Figure 8 shows the K-Means algorithms are used to calculate social distance. It is utilised to execute two points on the folks who have been detected. Because social distancing is tested between a minimum of two people, the cluster's minimum necessary points are set to two, and the two-person distance parameter is set to two metres. The orange colour represents the safe people and the blue colour reprsents the unsafe peoples. If the space between two people is very small, it is considered risky; if the distance is greater, it is considered safe.

Table 1 shows all of the models' performance was evaluated using measures such as accuracy, specificity, precision, recall/sensitivity, and F1 score. The SVM is shows the accuracy rate is 60%, the precision rate is 63%, the specificity is 70%, the recall rate is 90%, and the F1-Score is 75%. Random Forest is shows the accuracy rate is 88%, the precision rate is 73%, the specificity is 75%, the recall rate is 75%, and the F1-Score is 90%. And the KNN is accuracy rate is 95%, the precision rate is 90%,

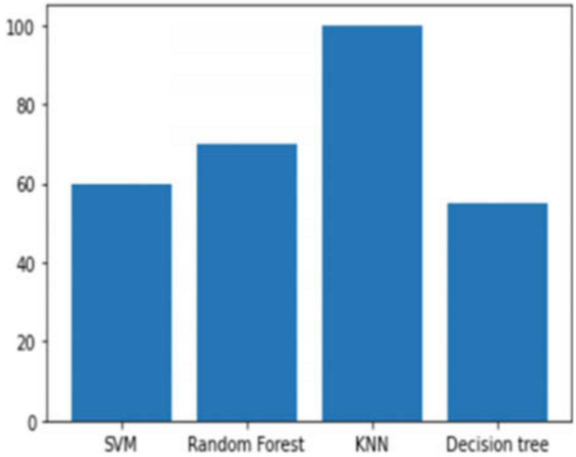


Fig. 7 Machine learnings algorithms

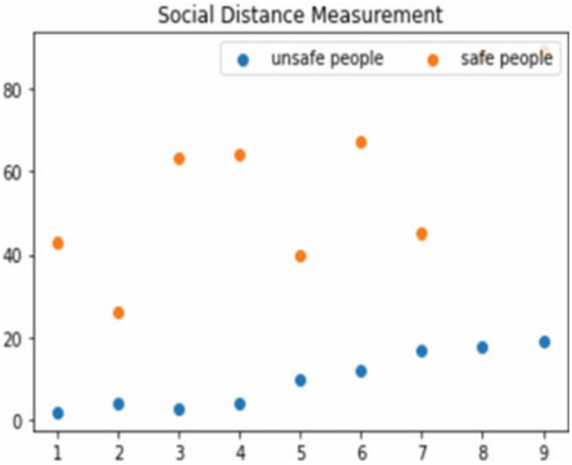


Fig. 8 Social distance measurement

Table 1 Classification result for four models

Models	Evaluation metrices								
	TP	TN	FP	FN	Accuracy	Specificity	Precision	Recall	F1 score
SVM	28	6	2	3	0.60	0.70	0.63	0.90	0.75
Random forest	29	6	2	2	0.88	0.75	0.73	0.75	0.90
KNN	29	5	3	2	0.95	0.89	0.90	0.93	0.92
Decision tree	26	6	2	5	0.65	0.75	0.69	0.73	0.72

the specificity is 89%, the recall rate is 93% and the F1-Score is 92%. Decision tree is shows the accuracy rate is 65%, the precision rate is 69%, the specificity is 75%, the recall rate is 73% and the F1-Score is 72%. The accuracy value of four models reveals that KNN is reliable for monitoring IoT-based portable monitoring device is developed to measure COVID-19 signals. With the highest F1 score of the four models, KNN emerges as the best. Although the SVM model is relatively similar to the decision tree, it cannot be regarded a robust model because to its low recall of 0.83.

6 Conclusions

The proposed tool is based on real-time sensors such as infrared (IR) and ultrasonic sensors for effective social distancing and contact tracing for COVID-19 prevention. Temperature, heart rate, SpO_2, and cough rate are all measured using the wearable sensor layer. It also sends real-time patient GPS position data to medical administrators and notifies family members to alleviate stress. The peripheral interface of the app is responsible for storing, collecting and analysing data in order to monitor and control public life and administration during the epidemic. The Android mobile app is extremely helpful in informing family members about patient status and lowering transmission rates. The wearable device has been fully designed to monitor patient health during and after infection. In order to control, monitor, and control patients who may be infected with COVID-19 in the spread of the disease, this system was tested and validated in real time at a hospital. A wearable device might be used as a model, allowing airport passengers to sit alone as they enter and exit. This work has undergone significant investigation in order to deliver the greatest device performance by comparing existing domains. The project's new capabilities are useful for evaluating health symptoms, tracking and monitoring a patient during detention, storing data to predict the scenario, and contacting authorities in a timely manner so that they may be properly monitored and use the Android platform to stay informed. The patient status of family respondents. Our proposed tool can also be used to prevent the spread of disease in future global health problems. And also take the advantage of this proposed technology that can help to identifying the early symptoms and treatment.

References

1. Hou YC, Baharuddin MZ, Yussof S, Dzulkifly S (2020) Social distancing detection with deep learning model. In: 2020 8th international conference on information technology and multimedia (ICIMU), 2020, pp 334–338. https://doi.org/10.1109/ICIMU49871.2020.9243478
2. Ahamad AH, Zaini N, Latip MFA (2020) Person detection for social distancing and safety violation alert based on segmented ROI. In: 2020 10th IEEE international conference on control system, computing and engineering (ICCSCE), 2020, pp 113–118. https://doi.org/10.1109/ICCSCE50387.2020.9204934
3. Motlagh NH et al (2021) Monitoring social distancing in smart spaces using infrastructure-based sensors. In: 2021 IEEE 7th world forum on internet of things (WF-IoT), 2021, pp 124–129. https://doi.org/10.1109/WF-IoT51360.2021.9595897
4. https://binged.it/3rD1btm
5. Sathyabama B, Devpura A, Maroti M, Rajput RS (2020) Monitoring pandemic precautionary protocols using real-time surveillance and artificial intelligence. In: 2020 3rd international conference on ıntelligent sustainable systems (ICISS), 2020, pp 1036–1041. https://doi.org/10.1109/ICISS49785.2020.9315934
6. Sengupta K, Srivastava PR (2022) HRNET: Ai-on-edge for mask detection and social distancing calculation. SN Comput Sci 3:157. https://doi.org/10.1007/s42979-022-01023-1
7. Srinivasan S, Rujula Singh R, Biradar RR, Revathi S (2021) COVID-19 monitoring system using social distancing and face mask detection on surveillance video datasets. In: 2021 international conference on emerging smart computing and ınformatics (ESCI), 2021, pp 449–455. https://doi.org/10.1109/ESCI50559.2021.9396783
8. Khanfor A, Friji H, Ghazzai H, Massoud Y (2020) A social IoT-driven pedestrian routing approach during epidemic time. In: 2020 IEEE global conference on artificial intelligence and ınternet of things (GCAIoT), 2020, pp 1–6. https://doi.org/10.1109/GCAIoT51063.2020.9345900

9. Savazzi S, Rampa V, Costa L, Kianoush S, Tolochenko D (2021) Processing of body-induced thermal signatures for physical distancing and temperature screening. IEEE Sens J 21(13):14168–14179. https://doi.org/10.1109/JSEN.2020.3047143
10. Lv W, Wu S, Jiang C, Cui Y, Qiu X, Zhang Y (2022) Towards large-scale and privacy-preserving contact tracing in COVID-19 pandemic: a blockchain perspective. IEEE Trans Netw Sci Eng 9(1):282–298. https://doi.org/10.1109/TNSE.2020.3030925
11. Arunkumar S, Mohana Sundaram N, Ishvarya D (2021) Temperature sensing wrist band for Covid-19 crisis. In: 2021 international conference on advancements in electrical, electronics, communication, computing and automation (ICAECA), 2021, pp 1–5, https://doi.org/10.1109/ICAECA52838.2021.9675689
12. Chloros D, Ringas D (2020) Fluspot: seasonal flu tracking app exploiting wearable IoT device for symptoms monitoring. In: 2020 5th south-east Europe design automation, computer engineering, computer networks and social media conference (SEEDA-CECNSM), 2020, pp 1–7. https://doi.org/10.1109/SEEDA-CECNSM49515.2020.9221843
13. Waheed A, Shafi J (2020) Successful role of smart technology to combat COVID-19. In: 2020 fourth international conference on I-SMAC (IoT in social, mobile, analytics and cloud) (I-SMAC), 2020, pp 772–777. https://doi.org/10.1109/I-SMAC49090.2020.9243444
14. Shubina V, Ometov A, Simona Lohan E (2020) Technical perspectives of contact-tracing applications on wearables for COVID-19 control. In: 2020 12th ınternational congress on ultra modern telecommunications and control systems and workshops (ICUMT), 2020, pp 229–235. https://doi.org/10.1109/ICUMT51630.2020.9222246
15. Luo T, Cao Z, Wang Y, Zeng D, Zhang Q (2021) Role of asymptomatic COVID-19 cases in viral transmission: findings from a hierarchical community contact network model. IEEE Trans Autom Sci Eng. https://doi.org/10.1109/TASE.2021.3106782
16. Sungheetha A (2021) COVID-19 risk minimization decision making strategy using data-driven model. J Inf Technol 3(01):57–66

Design and Implementation of Highly Secured Nano AES Cryptographic Algorithm for Internet of Things

E. Roopa and Yasha Jyothi M. Shirur

Abstract Advancement in the internet of things to meet the requirement of human beings and society makes integration of multiple devices into a single system. The integration of hardware and software needs to be provided with security to avoid the stealing of the data. Otherwise, the hacker may gain control over the devices and change the functioning of the system which may lead to malfunction. In order to provide security for the data transfer in IoT, the security algorithm need to be embedded with the IoT. The algorithm should provide high security and at the same time, it should be efficient. In this paper, an attempt is made to design a synthesizable Deoxyribonucleic acid (DNA) based Nano Advanced Encryption Standard (AES) Intellectual Property (IP) Core which can be used as a crypto engine in an IoT system. The crypto engine developed is optimized in terms of power, area and delay. The developed design when compared with the conventional design has given an area advantage of 81.6%, power of 21.17%, gate delay of 88.44% and path delay of 99.64%.

Keywords Internet of things (IoT) · Hacker · Crypto engine · Deoxyribonucleic acid (DNA) · Nano · Advanced encryption standard (AES) · Intellectual property (IP) core

E. Roopa (✉)
Department of ECE, VLSI Design and Embedded Systems, BNM Institute of Technology, Bengaluru, Karnataka, India
e-mail: roopa.elavarthi@gmail.com

Y. J. M. Shirur
Department of ECE, BNM Institute of Technology, Bengaluru, Karnataka, India
e-mail: yashajyothimshirur@bnmit.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_47

1 Introduction

1.1 Internet-of-Things

IoT is a physical object or network that is inbuilt with software, sensors and smart techniques and algorithm for data exchange. The data exchange will happen between two devices or multiple devices through the internet and it's termed as machine-to-machine (M2M) Communication. Every year the no of devices connecting to IOT is increasing rapidly, based on the statistics given by IoT Analytics, by 2025 the global number of connected IoT devices will be around 21.5 billion which is a 17% growth increased when compared to 2018 [1]. The devices include only active devices/nodes or gateways that interact with the end point sensors. It is because of smart techniques and algorithms embedded with it which supports wired and wireless technologies. The main motto behind integration is proper connectivity, low power consumption, low-cost sensor technology and to provide high security for data transfer. It can be accomplished by wireless communication and sensor networks, VLSI Design, artificial intelligence, machine learning, and cloud computing analytics. Figure 1 shows sudden growth in IoT devices with the technology supporting [1].

Nowadays, the whole world is full of IoT devices and providing security is the major concern while developing the IoT systems for applications like military, agriculture, educational system and even in health care where the data transfer is confidential and to be transferred securely. There may be multiple attacks which need to be addressed. Some of the major attacks usually dealt with are stealing information, taking control over the system and disrupting services [2].

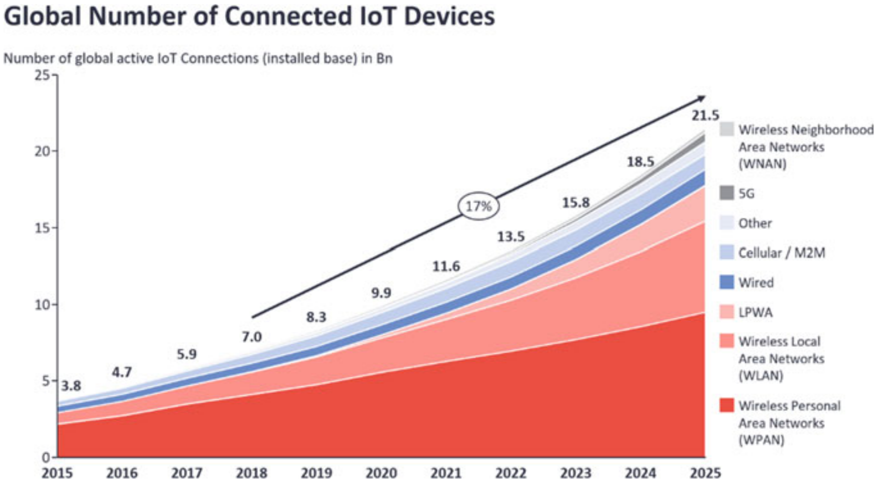


Fig. 1 Statistics to show rapid growth of IoT devices [1]

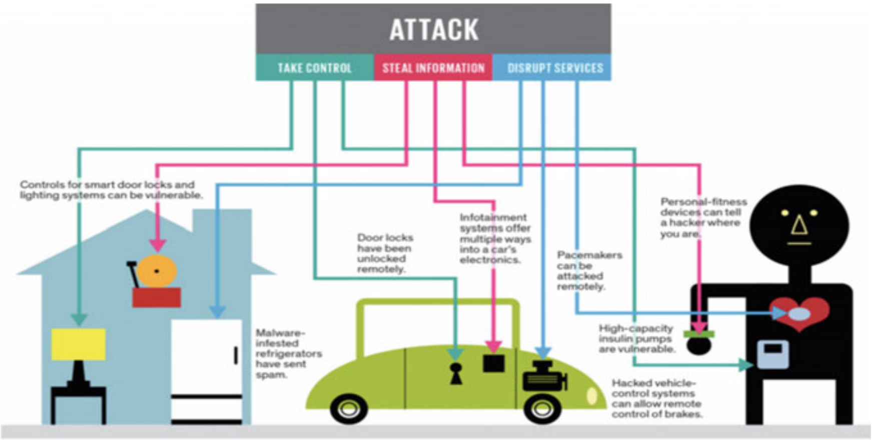

ATTACK
TAKE CONTROL
STEAL INFORMATION
DISRUPT SERVICES
Controls for smart door locks and lighting systems can be vulnerable.
Malware-infested refrigerators have sent spam.
Door locks have been unlocked remotely.
Infotainment systems offer multiple ways into a car's electronics.
Pacemakers can be attacked remotely.
Personal-fitness devices can tell a hacker where you are.
High-capacity insulin pumps are vulnerable.
Hacked vehicle-control systems can allow remote control of brakes.

Fig. 2 Attacks addressed in IoT system to provide security [2]

The solution for the above attack is secured data transfer, which is done by embedding the crypto engine into the IoT system. Figure 2 shows the major attacks which are considered to be addressed before the IoT systems are rolled out commercially for usage by the mankind.

1.2 Crypto System

Crypto System involves encryption and decryption. In the encryption process the plain text will be changed into an unreadable format, called cipher text. The cipher text will be changed into plain text in the decryption process. Figure 3 shows the block diagram of traditional crypto system.

The classification of cryptography algorithms is shown in Fig. 4. It is basically classified as symmetric and asymmetric key cryptography, further the classification

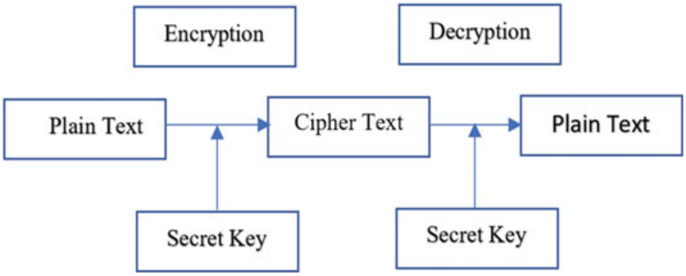


Fig. 3 Traditional crypto system

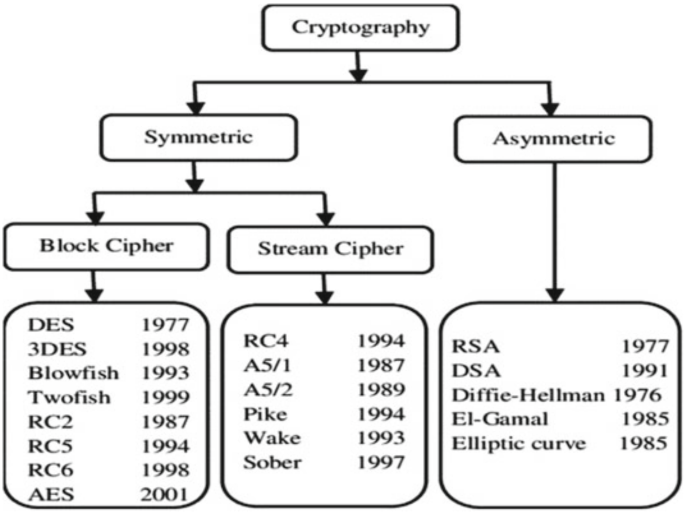


Fig. 4 Classification of cryptography algorithms [3]

is made based on the cipher text. In order to suit the requirement of the applications, there are wide ranges of cryptography algorithms are there in the literature [3].

DES—Data Encryption Standard
RC—Rivest Cipher
RSA—Rivest, Shamir, and Adelman
DSA—Digital Signature Algorithm.

They use different approaches to transfer the data securely. Some algorithms have drawbacks to transfer the data through a secured network. Since 2001 Advanced Encryption Standard (AES) is the highly secured algorithm used for data transfer, it over comes all the draw backs of other algorithms. It performs 10 rounds of operation for 128-bit input along with a 128-bit key and provides 128-bit output. This is the standard AES algorithm by the inventor Rijndael. This process is called as Rijndael encryption. The highlight of this algorithm is providing different keys in each round with the key generation technique, so that the secured transfer of data is possible. It is impossible to retrieve the data by third parties. In this work, the AES is implemented based on the concept of complex DNA structure to provide efficient and high security for data transfer in IoT connected devices.

1.3 Advanced Encryption Standard (AES)

Advanced Encryption Standard (AES) is a type of block cipher, which uses an encryption (secret key) key and many rounds of operation. This block cipher at a time uses a single block of data [4]. AES algorithm is a mathematical description, that reveals the procedure to encrypt data. It is not a source code or computer program to do the process on its own.

1.4 Deoxyribonucleic Acid (DNA)

DNA is an essential component in living organisms. The functioning of all living organisms depends on DNA [5]. This is basically used to store genetic information.

This information is unique, it can't be copied or duplicated. The idea behind biotechnology is that, it is not possible to change or steal the information while transmitting from sender to receiver. That is why it is called as DNA cryptography. Table 1 gives the details of DNA encoder symbols. It uses 40 characters out of which 26 (A–Z) alphabets, 10 (0–9) numbers and 4 special characters (space, comma, full stop and semicolon).

DNA Coding Technology

Table 2 clearly explains how the conversion will be performed. In this technique, each alphabet is given by a set of three letters. Every alphabet has its own notation. This three-group set is converted into decimal code. This decimal code is changed and is represented into an 8-bit binary code. It can be observed the input data is a character which is finally converted into a triplet code after subjecting it to a processing unit.

Table 1 DNA encoder symbols [6]

Character	DNA symbol	Character	DNA symbol	Character	DNA symbol	Character	DNA symbol
A	CGA	K	AAG	U	CTG	0	ACT
B	CCA	L	TGC	V	CCT	1	ACC
C	GTT	M	TCC	W	CCG	2	TAG
D	TTG	N	TCT	X	CTA	3	GCA
E	GGC	O	GGA	Y	AAA	4	GAG
F	GGT	P	GTG	Z	CTT	5	AGA
G	TTT	Q	AAC	Space	ATA	6	TTA
H	CGC	R	TCA	Comma (,)	GAT	7	ACA
I	ATG	S	ACG	Full stop (.)	GAT	8	AGG
J	AGT	T	TTC	Semicolon (;)	GCT	9	GCG

Table 2 ASCII codes

Alphabet	DNA conversion	Decimal code	Binary code
A	CGA	67 71 65	01,000,011 01,000,111 01,000,001
B	CCA	67 67 65	01,000,011 01,000,011 01,000,001
C	GTT	71 84 84	01,000,111 01,010,100 01,010,100
D	TTG	84 84 71	01,010,100 01,010,100 01,000,111
E	GGC	71 71 67	01,000,111 01,000,111 01,000,011
F	GGT	71 71 84	01,000,111 01,000,111 01,010,100
0	ACT	65 67 84	01,000,001 01,000,011 01,010,100
1	ACG	65 67 71	01,000,001 01,000,011 01,000,111
2	TAG	84 65 71	01,010,100 01,000,001 01,000,111
3	GCA	71 67 65	01,000,111 01,000,011 01,000,001
4	GAG	71 65 71	01,000,111 01,000,001 01,000,111

2 Literature Survey

The highlights of the literature survey are show in Table 3.

Some of the Differential Power Analysis (DPA) attacks are vulnerable to Advanced Encryption Standard algorithms. It is giving an approximate value rather than exact values. For different applications, different approaches and methods are used. There is no exactly single methodology for all applications. Power consumption is reduced in some approaches but there is a trade-off with respect to resource usage. The available approaches make use of both combinational logic and sequential logic, and switching between these two consumes more power. Therefore, in the developed algorithm an attempt is made to replace some part of sequential logic to avoid the switching between the logic and in turn reduce noticeable power.

3 Design Implementation

AES implementation consists of two main blocks, one is state-register and other one is key-register. In the implemented technique the state-register itself acts as a shiftRow. The implementation is performing 128-bit encryption [7]. The total process is performed in 10 cycles, in first compute—1 round per cycle, here hardware of each round is being reused to save the area against a fully unrolled implementation. The proposed AES design is shown in Fig. 5. The inputs plain text and key are given to the state-register and key-register, both are sharing the same s-box, and they are merged to reduce the area [8]. This is the main idea of the proposed method. The flow is as follows; SubBytes, ShiftRows, MixColumn and AddRoundKey [9]. It repeats until 9 rounds, then the final round operation will be performed. In the final round there is no MixColumn. The output is given to the DNA encoder, and it encodes the

Table 3 Highlights of the literature survey

SI. No.	Title of the paper	Authors	Publication details	Description
1	Differential power analysis: a serious threat for FPGA security	M. Masoumi	International Journal of Internet Technology and Secured Transactions 4(1):12–25	Using differential power analysis (DPA) is a good practice for attack measurement
2	The research of DPA attacks against AES implementations	Yu HAN, Xuecheng Zou Liu Zhenglin, and Yi-cheng CHEN	The Journal of China Universities of Posts and Telecommunications 15(4):101–106	Hardware implementation has less data-dependent power leakages to resist power attacks
3	AES against first and second-order differential power analysis applied cryptography and network security	J. Zhou and M. Yung, Eds	Vol. 6123, Springer-Verlag, pp. 168–185. Berlin, Germany	AES is secure for software implementation to be resistant against first and second order DPA in practice
4	High-speed VLSI architectures for the AES algorithm	X. Zhang and K. K. Parhi	IEEE Transaction Very Large Scale Integrated (VLSI) System, vol. 12, no. 9, pp. 957–967	The speed will be increased using combinational logic
5	Post-quantum crypto processors optimized for edge and resource-constrained devices in IoT	Shahriar Ebrahimi, Siavash Bayat-Sarmadi, Hatameh Mosanaei-Boorani	IEEE Internet of Things Journal PP (99):1–1	Optimized variant for binary learning with errors over the ring (Ring-LWE) is used against quantum attacks
6	A high data rate pipelined architecture of AES encryption/decryption in storage area networks	Hossein Kouzehgar, Meisam Nesary Moghadam and Pooya Torkzadeh	26th Iranian Conference on Electrical Engineering (ICEE2018)	High throughput can be achieved through pipeline structure
7	A high throughput and secure authentication-encryption AES-CCM algorithm on asynchronous multicore processor	Ali Akbar Pammu, Weng- Geng Ho, Ne Kyaw Zwa Lwin, Kwen- Siong Chong and Bah-Hwee Gwee	IEEE Transactions on Information Forensics and Security PP (99)	High through can be achieved using parallel-encryption implemented on an asynchronous multicore processor (AMP-MP)

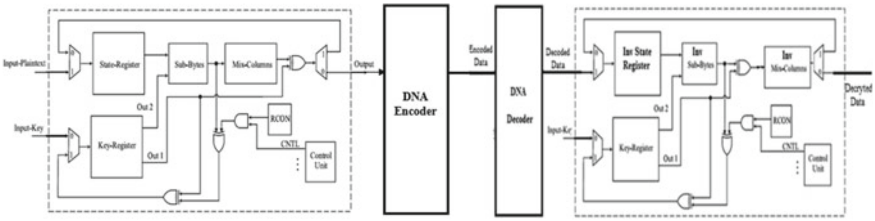


Fig. 5 Block diagram of the proposed nano-AES crypto system

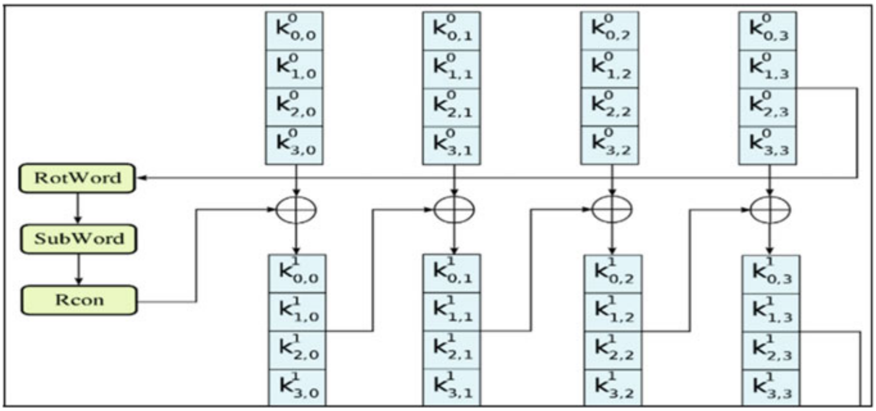


Fig. 6 Key generation process

data. The encoded data is transmitted over the allocated network. At the receiver side the encoded data is decoded using the DNA decoder, then it would be given to the decrypted to perform the decryption. In the key generation process, first data is Rot Word (Rotation Word), it takes out the last column of the data and rotates the column one time, then change the data with s-box. That data is XOR-ed with the first column of the actual data. This data is XOR-ed with an RCON, called Round Constant [8], which is shown in Fig. 6. So new keys will be generated for each and every round. This total process is explained in the flow chart, which is given in the Fig. 7.

Flow Chart of AES Design

See Fig. 7.

4 Result and Discussions

Nano-AES crypto algorithm is coded in Verilog and simulated in ModelSim and synthesized in Xilinx9.1. Figure 8 depicts the simulation results obtained; the details

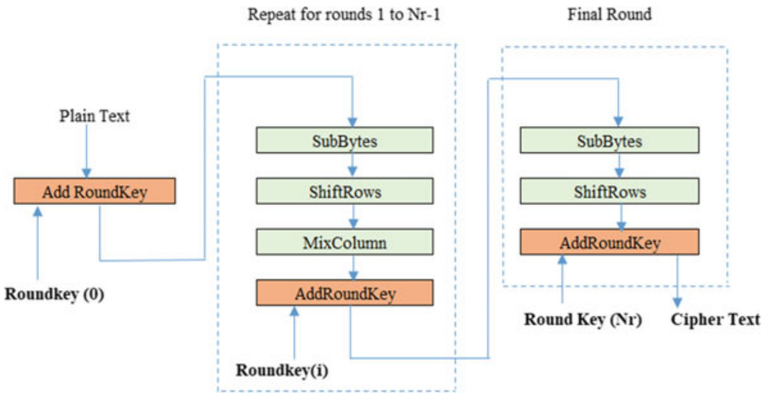
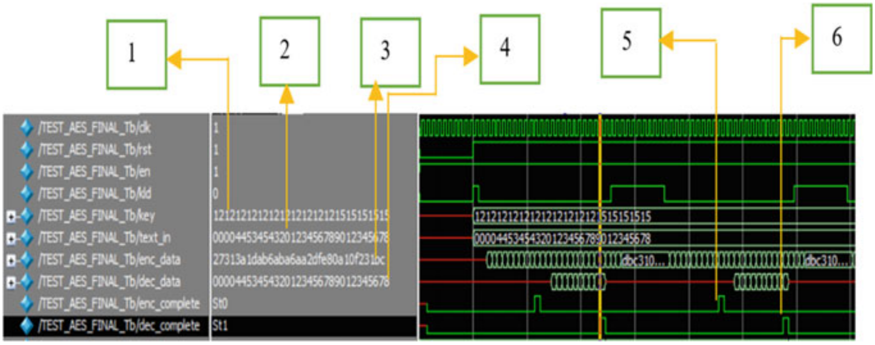


Fig. 7 Execution process of AES design

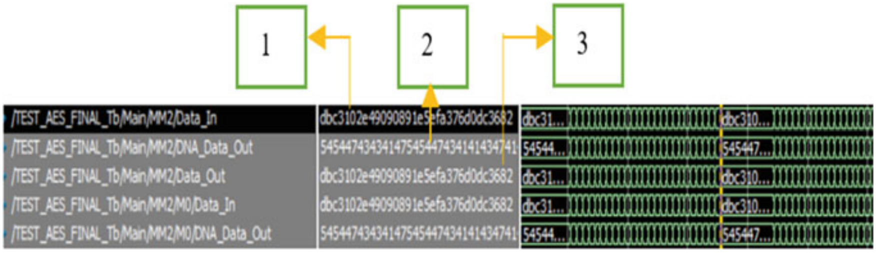
of the marker representation are as follows: 1. Input Key 2. Plain Text 3. Encrypted Data 4. Decrypted Data 5. Encryption Acknowledgement 6. Decrypted Acknowledgement. It is to be observed that the plain text and the decrypted data are the same. The obtained results are cross verified using the online AES calculator.

The Nano DNA AES encoder and decoder output simulation results are shown in the Fig. 9, the details of the marker representation are as follows: 1 Encrypted Data 2. DNA Encoder output 3. DNA Decoder output. Figure 10 shows the device utilization chart. The time taken by the encryption process is 8.801 ns, which is depicted in the synthesis report shown in Fig. 11. By using the new approaches in the proposed algorithm, the area, delay and power have reduced. This can be observed from the bar graphs shown in Figs. 12, 13 and 14. Table 4 shows the comparison between existing and proposed design with clock gating.



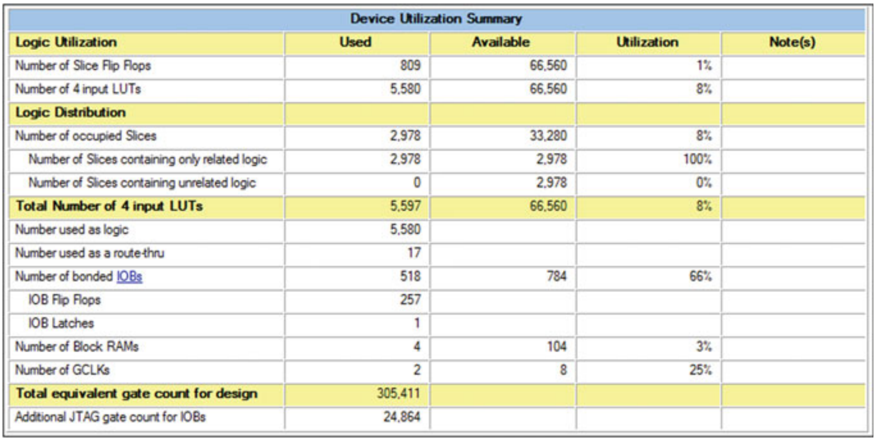
```
/TEST_AES_FINAL_Tb/clk            1
/TEST_AES_FINAL_Tb/rst            1
/TEST_AES_FINAL_Tb/en             1
/TEST_AES_FINAL_Tb/kld            0
/TEST_AES_FINAL_Tb/key            12121212121212121212121215151515
/TEST_AES_FINAL_Tb/text_in        00004453454320123456789012345678
/TEST_AES_FINAL_Tb/enc_data       27313a1dab6aba6aa2dfe80a10f231bc
/TEST_AES_FINAL_Tb/dec_data       00004453454320123456789012345678
/TEST_AES_FINAL_Tb/enc_complete   St0
/TEST_AES_FINAL_Tb/dec_complete   St1
```

Fig. 8 Nano-DNA AES encryption and decryption with clock gating



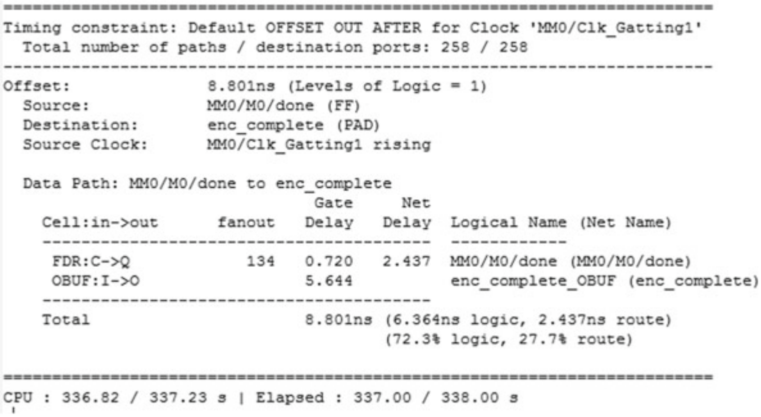
/TEST_AES_FINAL_Tb/Main/MM2/Data_In	dbc3102e49090891e5efa376d0dc3682	dbc31...	dbc310...
/TEST_AES_FINAL_Tb/Main/MM2/DNA_Data_Out	54544743434147545447434141434741	54544...	545447...
/TEST_AES_FINAL_Tb/Main/MM2/Data_Out	dbc3102e49090891e5efa376d0dc3682	dbc31...	dbc310...
/TEST_AES_FINAL_Tb/Main/MM2/M0/Data_In	dbc3102e49090891e5efa376d0dc3682	dbc31...	dbc310...
/TEST_AES_FINAL_Tb/Main/MM2/M0/DNA_Data_Out	54544743434147545447434141434741	54544...	545447...

Fig. 9 Simulated results obtained for Nano DNA AES encoder and decoder



Device Utilization Summary				
Logic Utilization	**Used**	**Available**	**Utilization**	**Note(s)**
Number of Slice Flip Flops	809	66,560	1%	
Number of 4 input LUTs	5,580	66,560	8%	
Logic Distribution				
Number of occupied Slices	2,978	33,280	8%	
Number of Slices containing only related logic	2,978	2,978	100%	
Number of Slices containing unrelated logic	0	2,978	0%	
Total Number of 4 input LUTs	5,597	66,560	8%	
Number used as logic	5,580			
Number used as a route-thru	17			
Number of bonded IOBs	518	784	66%	
IOB Flip Flops	257			
IOB Latches	1			
Number of Block RAMs	4	104	3%	
Number of GCLKs	2	8	25%	
Total equivalent gate count for design	305,411			
Additional JTAG gate count for IOBs	24,864			

Fig. 10 Device utilization chart



```
========================================================================
Timing constraint: Default OFFSET OUT AFTER for Clock 'MM0/Clk_Gatting1'
  Total number of paths / destination ports: 258 / 258
------------------------------------------------------------------------
Offset:              8.801ns (Levels of Logic = 1)
  Source:            MM0/M0/done (FF)
  Destination:       enc_complete (PAD)
  Source Clock:      MM0/Clk_Gatting1 rising

  Data Path: MM0/M0/done to enc_complete
                                Gate     Net
    Cell:in->out      fanout   Delay   Delay  Logical Name (Net Name)
    ----------------------------------------  ------------
     FDR:C->Q            134   0.720   2.437  MM0/M0/done (MM0/M0/done)
     OBUF:I->O                 5.644          enc_complete_OBUF (enc_complete)
    ----------------------------------------
    Total                      8.801ns (6.364ns logic, 2.437ns route)
                                       (72.3% logic, 27.7% route)

========================================================================
CPU : 336.82 / 337.23 s | Elapsed : 337.00 / 338.00 s
```

Fig. 11 Synthesis timing report for encryption

Fig. 12 Comparison of area utilization

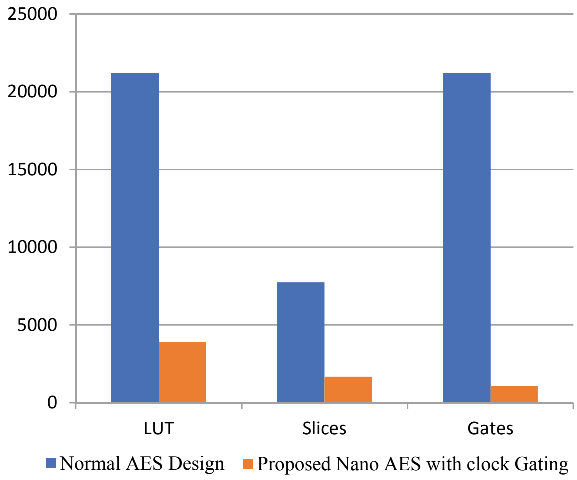


Fig. 13 Comparison of delay

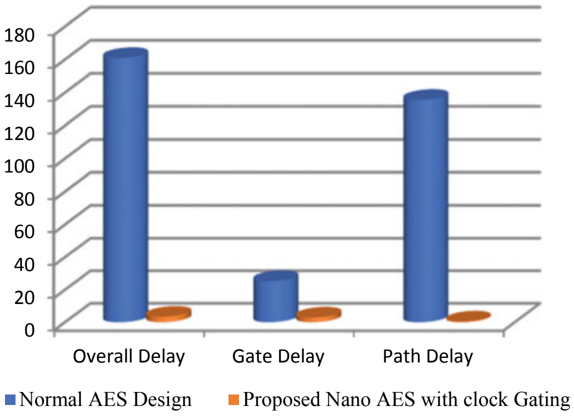


Fig. 14 Comparison of power consumption

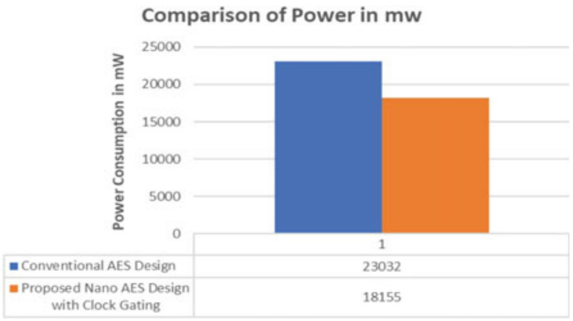


Table 4 Comparison between conventional AES design with proposed nano AES with clock gating

Sl. No.	Method name	Area			Delay			Total power (mw)
		Slice	Flip flops	LUT	Max delay (ns)	Gate delay (ns)	Path delay (ns)	
1	Conventional AES design	7734	21,207	21,207	160.860	25.302	135.558	23,032
2	Proposed nano AES with clock gating	1670	1066	3900	3.405	2.923	0.482	18,155
3	Percentage reduction (%)	78.4	94.9	81.6	97.88	88.44	99.64	21.17

5 Conclusion

The light-weight AES architecture is the best choice for resource-constrained IoT devices. It uses a symmetric cryptography algorithm and provides high security. This algorithm is used by many applications and networks. The design contains two main registers each is of 8 bits, one is State-register and the next one is a key-register. It is designed in a way that the ShiftRows are running inside the state-register to reduce the logic. SubBytes are optimized for these two registers as they both uses the same S-Box. Next is a mix-Column with 8-bit input and output, with this low-area design is achieved. The area and power consumption are reduced by using the clock gating technique. The crypto engine developed is optimized in terms of power, area and delay. The developed design when compared with the conventional design has given an area advantage of 81.6%, power of 21.17%, gate delay of 88.44% and path delay of 99.64%.

Acknowledgements Authors wish to thank the management of BNM Institute of Technology and Visvesvaraya Technological University for their support and encouragement rendered to in this work.

References

1. Lueth KL (2018) State of the IoT 2018: number of IoT devices now at 7B—market accelerating. IoT Analytics
2. Joe (2015) The 10 challenges of securing IoT communications
3. Elgeldawi E, Mahrous M, Sayed A (2019) A comparative analysis of symmetric algorithms in cloud computing. Int J Comput Appl
4. Biryukov A (2015) Block ciphers and stream ciphers: the state of the art
5. Condon A (2006) Designed DNA molecules: principles and applications of molecular nanotechnology. British Columbia V6T 1Z4, Canada
6. Miles I (2020) An in-depth guide to AES encryption with angular service implementation

7. Arrag S, Hamdoun A, Tragha A, Khamlich S (2012) Design and implementation a different architectures of mixcolumn in FPGA. Research Gate
8. Waqas U, Afzal S, Mir MA, Yousaf M (2015) Generation of AES like S-boxes by replacing affine matrix. Research Gate
9. Oh J-Y, Yang D-l, Chon K-H (2010) A selective encryption algorithm based on aes for medical information. Healthc Inf Res

Convergence of Communication Technologies with Internet of Things

V. Dankan Gowda, Suma Sira Jacob, Naziya Hussain, R. Chennappan, and D. T. Sakhare

Abstract Internet of Things (IoT) is the term used to describe a network of physical things such as mobile devices and household appliances that are embedded with electronics, software, sensors, and network connection that enables these objects to gather and exchange data. Sensors, recognition and remote control of items are all made possible by the Internet of Things (IoT). Once this property is combined with sensors and actuators, it becomes an example of a cyber-physical system, which includes technologies like intelligent power grids (grids), intelligent homes (smart homes), smart cities (smart cities), and intelligent transportation systems (ITS). Integrating MANET and WSN with IoT is covered in this study. Technology and protocols needed to deploy the Internet of Things (IoT) are explored in this article.

Keywords Internet of things · Wireless sensor network · Protocol · Network · Sensor · Node

V. Dankan Gowda (✉)
Department of Electronics and Communication Engineering, BMS Institute of Technology and Mangement, Bangalore, Karnataka, India
e-mail: researchr08@gmail.com

S. S. Jacob
Department of Information Technology, Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India

N. Hussain
School of Computers, IPS Academy, Indore, Madhya Pradesh, India

R. Chennappan
Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

D. T. Sakhare
Department of Chemistry, U.G., P.G. and Research Centre, Shivaji Arts Commerce and Science College, Kannad Dist. Aurangabad, Maharashtra, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_48

1 Introduction

Kevin Ashton was the first to use the phrase "Internet of Things" in relation to supply chain management. As a result, the concept has evolved over the last decade to include a wider range of applications, including healthcare, utilities, transportation, and more [1]. Computers can now discern information without any human intervention, even if the concept of 'Things' has evolved throughout time with advancements in technology. One of the most significant developments in the recent decade has been the advent of the Internet of Things concept. It is now possible to turn an isolated item into a communicative device because to advancements in the electronics industry, the expansion of communication protocols, and the shrinking size of devices and transceivers. As a result, compact processing and sensor devices have become more powerful, more energy efficient, and more capable of storing data. The exponential growth in the number of Internet-connected sensing and computing devices known as "smart gadgets" has been sparked by these advancements in electronics and computer science. These devices may deliver massive services only limited by human imagination. The Internet of Things (IoT) is a network of interconnected physical devices that communicate and exchange data with one another. Sensors, radio frequency identification devices (RFID), infrared sensors (IR), laser scanners, Global Positioning System (GPS), wireless LANs (WLAN), and even Local Area Networks (LANs) interfaces [2] are required for these devices. By using IPv6, these devices may be linked to the Internet and managed from a distance. Sensors may convey data to other devices for processing, allowing these devices to communicate with one other. An M2M connection is one in which a machine communicates with another machine under the IoT paradigm. The Internet of Things (IoT) relies heavily on WSN and RFID technology [3]. In recent years, a number of studies have been conducted to examine the increasing use of IoT applications. The integration of heterogeneous devices, including mobile phones, laptop systems, PDAs and smart phones, wireless sensors and network-enabled physical items, such as RFID and smart visual tags, is also taking place. Once they've been integrated, these gadgets will be able to communicate easily with the Internet [4]. When it comes to building smart cities, IoT will play a critical role in ensuring its residents have access to high-quality amenities and services. To reach this purpose, it is necessary to gather data from a variety of sources, analyse it, and display it in a variety of ways. This action requires a great deal of standardisation work from several angles. It is critical that new, self-governing and adaptable smart city services be developed that can be used in a variety of application areas, from environmental monitoring to security control and sophisticated applications. Monitoring of natural resources such as air, water, noise, and light pollution, as well as pollution caused by human activity such as automobiles, industries, and traffic, are all included in environmental monitoring. Monitoring natural resources, military operations and others are included in this category as well. Structural monitoring to avoid bridge and ancient building collapses, support for people living and travelling, assistance for elderly and handicapped persons, emergency response are all examples of security control. Smart TV, uninterrupted multimedia streaming, online gaming through mobile devices, and a host of other high-tech applications are just a few examples. In smart cities, WSNs and MANETs are key technology for a variety of IoT applications [5]. These two technologies are more suited for the installation of IoT applications because of their localised and self-configuration capabilities. The implementation of WSNs and MANETs in cities for public safety, location-aware computation, and environmental monitoring is predicted by several studies to occur shortly [6]. Mobile phones equipped with several wireless interfaces (IEEE 802.11, Bluetooth, and 3G) have already begun delivering low-power connection alternatives, such as IEEE 802.15.4, onboard. It is projected that in the near future, most consumer gadgets will have access to low-power connection [7].

2 Literature Survey

The Internet of Things (IoT) is a network of things that are linked through the Internet. Internet of Things (IoT) is an amalgamation of WSNs and the old Internet. Resource-constrained sensors are the norm in WSNs, but powerful devices predominate over them on the internet. Consequently, IoT may be described as a network of diverse gadgets. There are many different types of Internet of Things devices, from sensors to light bulbs to energy metres to vehicle parts to smartphones to PCs and tablets to powerful servers to cloud computing systems. As a result, the Internet of Things has the potential to link billions of IP-enabled devices. In [8], authors conducted a survey on WSN assaults, and in Raymond and, authors reported their findings. WSN DoS attacks and responses were examined in [9] by the authors. As outlined in [10], IP-based WSN security concerns and responses were discussed in depth in this paper. Many studies have shown that there are still many unsolved problems. An in-depth look at the hurdles and issues linked with IoT is provided in [11] author's work. To name just a few, there's interoperability, an IoT-based business model that allows for the networking of billions of devices, security and privacy issues such as trustworthiness and complete end-to-end encryption, and more. Security models for IoT technologies must be efficient, according to the study authors. A different point of view was expressed in [12] by the authors, who pointed out that there are two primary concerns: security and privacy. They've gone through the obstacles and spoken about them from various angles. Data storage and secure processing have all been taken into consideration while addressing security concerns, including authentication and authorisation as well as Denial-of-Service (DoS) attack mitigation. Additionally, while discussing privacy issues, they explored the privacy choices, identity control, and commercial demands of passive users. Among the issues discussed in [13] are those relating to the security of the Internet of Things. The authors made it apparent that large-scale implementation of IoT would be fraught with difficulty and would need careful consideration of a number of significant issues. Some of the difficulties that need to be addressed include those related to secrecy and privacy, security, heterogeneous device management, and network restrictions. By [14], the security risks and difficulties of the Internet of Things are addressed in an innovative way [15].

As the most important goal of IoT security, data security is the basis for this categorization. Sleep Deprivation (SD) attacks have been studied in a variety of MANET setups. The Route Request (RREQ) flooding attack in MANETs was discussed in [16]. They came up with a way to avoid the RREQ flooding attack by relying on the monitoring of the neighbouring node. The incoming RREQs are kept in a priority queue. If RREQs are regularly produced by a single node, their priority is degraded. In a MANET environment, attackers may drain the power of wireless devices like PDAs and notepads in a variety of methods, as described in [17]. The battery life of notepads and PDAs was assessed in a series of trials done under this assault. Finally, they observed that this assault consumes the battery's energy faster. In order to defeat this attack, they've created a power-secure architecture that makes use of system energy monitoring and layered authentication. AODV protocol flooding may be reduced by using the session-based history table provided in [18]. The average number of RREQ packets is logged and compared to the discard threshold in this method to identify flooding attacks. In order to conduct a sleep deprivation assault, adversary nodes must become cluster chiefs, as predicted in [19]. Random vote, round robin, and hash-based schemes are examined for minimising sleep deprivation attack mitigation [20]. When a node detects an incursion, the agent then begins a global reaction. Analyzing the behaviour of neighbouring nodes may help assess the trustworthiness of other nodes in a MANET, as stated by [21]. An investigation into the use of mobile agents in MANETs led to the conclusion that mobile agent-based IDS is the best option for MANET IDS. The Dendritic Cell Algorithm is a well-known addition to the danger project (DCA). It takes advantage of the HIS's innate resistance to dendritic cells. The DCA method therefore proves that it can identify port scanning attacks, and it is thus argued that DCA may be utilised as an algorithm for anomaly detection. A comparison of MANET and sensor network environment features with innate immunity traits has been made by [12, 13]. This shows that the DCA may be used to identify various forms of assaults in dynamic settings like MANET.

3 RFID and IoT

To make the Internet of Things work, a few basic procedures are required [8]. In order to implement M2M and D2D communication, two major needs must be met: The first step is to recognise what you're looking at. In order to be uniquely identified, objects must have an integrated Auto-ID technology, which might be an RFID tag. Then there's the issue of communication. Devices using RFID tags can access the most important information. The tags' RFID readers may transmit the data they acquire to the internet. In Fig. 1, you can see how RFID technology may be linked to the Internet. Since the RFID readers are working as interpreters, they're useful in this situation. Radio-frequency identification tags may be passive or active. The radio signal provided by the reader serves as the only source of power for passive tags, which need no external power source. As a result, passive tags are far less expensive

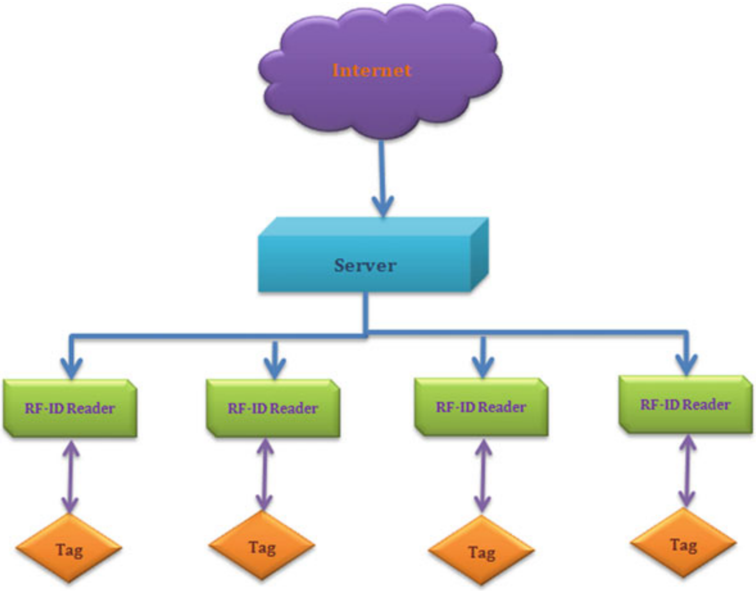


Fig. 1 RFID technology connected with the internet

than active tags. The Electronic Product Code is linked to RFID's involvement in the Internet of Things (IoT) (EPC). The MIT Auto-ID Center has developed EPC to serve as a global identifier for all physical objects on the planet.

Since RFID readers function as sensor nodes and RFID technology is the wireless connection used to gather data from tags, all the WSN techniques may be used to RFID communication. There is a server linked to the Internet through an RFID reader network.

4 Wireless Sensor Network

The sink node in a wireless sensor network collects data from the network's sensor nodes and distributes it to the rest of the network's nodes [9]. Sensor nodes are often used to gather environmental data, such as temperature, pressure, humidity, and proximity, among other factors. There are two ways to look at WSN in relation to the Internet of Things: The network as a whole is a single entity, with the sink node serving as the primary access point for all network information.

In terms of memory, compute, and battery power, WSNs are well-known to be constrained in their capabilities. Changing or recharging the small batteries that power WSN nodes is impractical in a WSN setup. To minimise frequent disconnections, nodes must save battery power by using efficient transmission techniques and implementing good MAC and routing protocols. In a monitored environment, WSN may be installed using a star, mesh, or tree topology. Because the nodes in a star topology are all within one hop of the sink node, data gathered from several sensors may be redundant. Such duplicate data is processed by the sink node in this case. With the usage of a MANET communications overlay in both mesh and tree topologies, multi-hop communication may take place. One method of cutting down on the network's power consumption is to use a graph theory-based algorithm, such as the minimal Connected Dominant Set (CDS) or minimum Spanning Tree. WSN has two significant variants: For the first time, a patient's bodily status may be monitored and sent to the Internet through a wireless body area network (WBAN) including a number of tiny biosensors [5]. Using the WBAN for a smart health monitoring system is the primary use case. The Wireless Sensor and Actuator Network (WSAN) [7] is the second option. Sinks, sensors, and actuators are the three main kinds of nodes in WSANs. Data from sensors is sent to the sink node, which processes the information and sends it to the backbone server through a link. Actuators are able to respond to orders from the server and control the environment.

5 Connecting WSN with IoT

Because sensor nodes are small and have limited resources, establishing a WSN connection to the Internet is very difficult. To connect WSNs to the Internet, many designs have been suggested. In the image, these designs may be divided into three classes [12]. Figure 2, IP over WSN, sensor over IP, and higher-level gateway overlays (i.e., the IP layer on top of the sensor layer). IP overlay over WSN: Sensor nodes are given a distinct IP address so that they may communicate with the Internet.

Due to the low processing capability of the sensor nodes, this strategy is not a viable one. This problem may be alleviated in the future if efforts are made to integrate IPv6 in sensor nodes. In this concept, each node is addressed using IPv6, making it a true Internet of Things (IoT). Sensor nodes may be identified using the 6LoWPAN protocol established by the IETF [14].

The sink node encapsulates the detected data in IP packets before sending them to the Internet through the sensor overlay over IP as shown in Fig. 2a. The sink node is online, and the sensor nodes are virtualized in this design Fig. 2b. WSNs and the Internet are treated as different networks by using a higher-level gateway overlay. Both networks' traffic is routed via the gateway to be redirected to the appropriate network. Internet routing information is transformed into WSN routing techniques by the gateway as in Fig. 2c. This is the most common way to link WSNs to the Internet. IP packet translation is the responsibility of the adapter at the sink node.

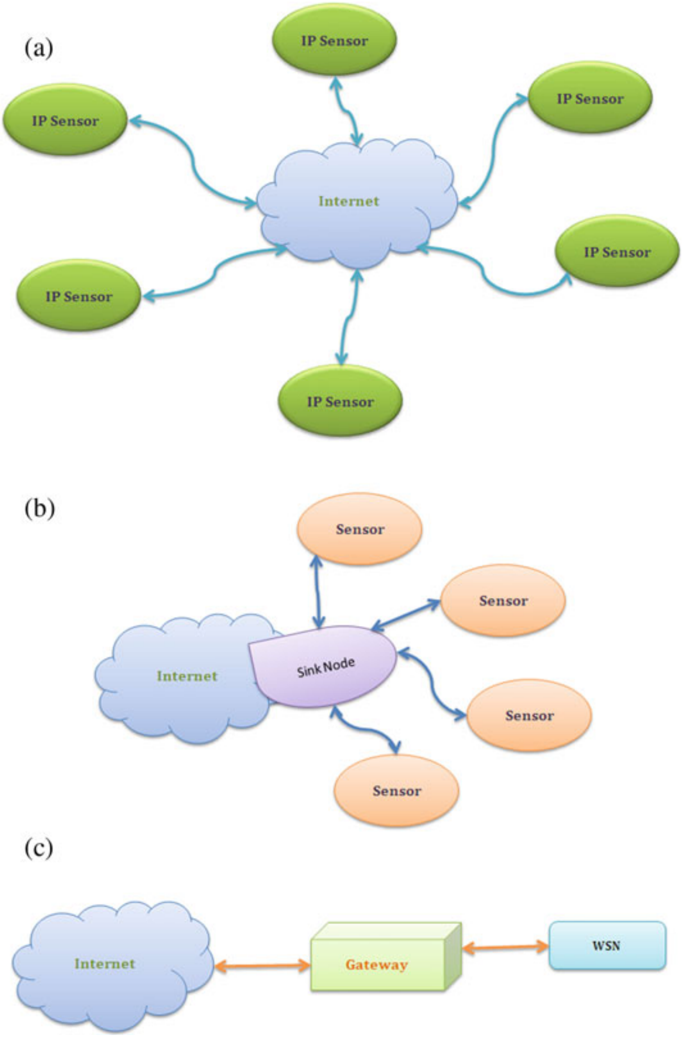


Fig. 2 Three schemes of connecting WSN with the IoT **a** the IP overlay over WSN, **b** the sensor overlay over IP, **c** the higher-level gateway overlays

6 Connecting MANET's to IoT

Mobile Ad hoc Network (MANET) is a self-configurable network of mobile nodes without the need for an infrastructure. These systems are often used in areas where infrastructure is scarce and patching it up is not an option. MANET has already shown its worth in disaster zones and on the battlefields. MANET's multi-hop communication and low implementation costs make it beneficial in the above-mentioned locations. Every node in a MANET functions as a router, forwarding packets to the next node on the network. Green communication, M2M, D2D, and the Internet of Things (IoT) are examples of recent research developments that demonstrate the value of ad hoc networks in reducing deployment and communication costs [11]. In order to link the MANET to the Internet, many technologies have been proposed. IP addresses are often assigned to nodes in a MANET so that data packets may be routed between them. As a result, it is always feasible to connect the MANET to the Internet. However, there are two major obstacles to overcome: (i) It is important for a node in a MANET to be able to quickly determine whether an address in the network is existent or not. (ii) Whether or not an access point or a gateway is required to connect to the Internet should be made very clear. Any node in the network may join or depart at any moment because to the nodes' mobility and flexibility. This makes it difficult to gather neighbouring nodes' IP addresses since nodes do not know where they are at any one moment. They use more time and message packets even if they are accessible. Connecting MANET to the Internet may be accomplished in a number of ways: Connecting a MANET to the Internet through an Access Point is the first option. However, the location of the access point is an issue. It is difficult to find the most efficient location for an access point or gateway. A mobile node might serve as a node of entry. Using two IP addresses, one for the MANET network and one for the Internet, is an alternative method for identifying nodes in the network. The target gateway, on the other hand, may be movable due to the nodes' mobility. It's possible that outbound connections might be disrupted if a node moves to a different gateway. Moving nodes may also be given IP addresses using the Dynamic Host Configuring Protocol (DHCP). Mobile nodes may now configure their IP addresses using this manner, however the DHCP server placement issue remains unresolved. As a result, MANET nodes may be automatically setup. In the literature, there is a lot of discussion on auto-configuration strategies. Connecting mobile nodes to the Internet via an appropriate auto-configuration approach in a MANET context is essential. From an IoT perspective, the capacity of nodes in a WSN to sense their surroundings and organise themselves into ad hoc networks to convey data is significant. However, three obstacles must be solved before the IoT may be used in more diverse ways [6]. Internet of Things (IoT) connectivity support for heterogeneous devices. In order for the Internet of Things to be a success, WSN nodes will need to have a number of common traits and functionalities. Sensor nodes' battery power. IoT battery depletion and the necessity for a regular battery replacement pose a significant hurdle to wider deployments, despite many efforts to improve energy efficiency at different levels of the system. Nodes integrated with sensors and microcontrollers are too large to be used in an IoT system that will be available to everyone. To fully exploit the IoT's potential, further progress must be made in the field of miniaturisation.

7 MANET-IoT Integration Protocol

IoT's WSN backbone requires two services from MANET: discovery and announcing. Discovery enables MANET nodes to look around the WSN topology

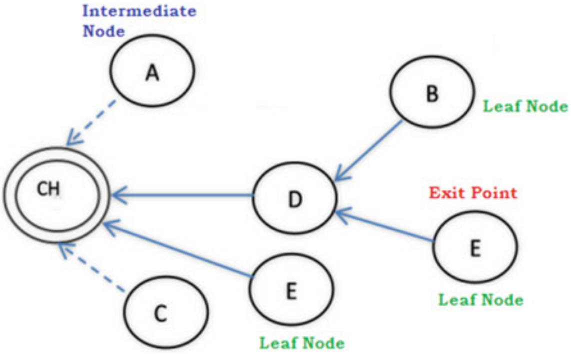


Fig. 3 MANET cluster formation

and choose a suitable node to connect to. WSN nodes are notified of the existence of MANET access points via the process of announcing. To prevent excessive power consumption, it is required to reduce the number of packet exchanges between MANET and WSN. Maintaining active communication and coordination with WSN packets for higher priority packets allows the MANET nodes to remain idle in typical scenarios. When a higher-priority data packet is detected in the WSN, any MANET node, which will be the cluster-head, initiates the cluster-formation procedure. As seen in the diagram, the procedure is divided into three sections. First, a MANET node identifies the higher priority WSN data packet as a cluster head and broadcasts a request to join the MANET group to other nodes. In the second phase, the MANET nodes that receive the cluster head's request send the discovery message to the WSN nodes to locate an entry point for IoT and MANET connection. Nodes that receive and connect with IoT nodes might join the MANET cluster in the third phase. The NS-2 network simulator is used to test the proposed MANET-IoT integration protocol's performance. MANET cluster formation is shown in Fig. 3.

8 Results and Discussion

Because MANET nodes are distributed across a larger region and might be placed in places that are disconnected from the WSN, simulation findings demonstrate that increasing the number of MANET nodes has a significant impact on the delay of higher priority packets. Figures 4 and 5 shows this outcome.

When MANET nodes move at 1.2 m/s, our protocol delivers over 90% of packets and drops over 60% of packets when nodes move at 12 m/s.

Fig. 4 Latency ratios between normal and higher priority packets

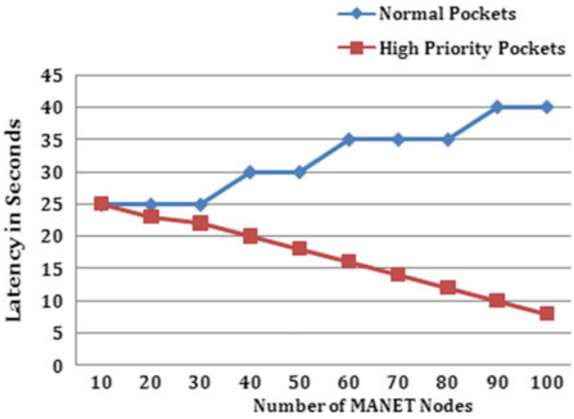
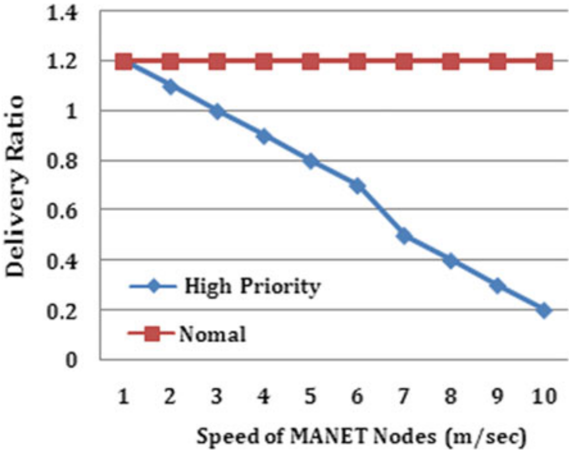


Fig. 5 The ratio of packets successfully delivered and the speed of MANET nodes



9 Conclusion

Wireless sensor networks (WSNs), radio frequency identification (RFID), and mobile ad hoc networks (MANETs) are all discussed in this study (MANET). Though the Internet of Things (IoT) is interconnected via a variety of standards, protocols, and other means of communication, several issues remain. Further study is needed on energy management, energy efficiency, and energy-related assaults. IoT integration with MANET and WSN will help us build smarter environments because of the increasing prevalence of IoT services and applications in our daily lives. However, the IoT is also affected by the faults or challenges associated with these technologies. To overcome the problems and secure the Internet of Things (IoT) against different threats, new solutions must be developed.

References

1. Andrea I, Chrysostomou C, Hadjichristofi G (2015) Internet of things: security vulnerabilities and challenges. In: Computers and communication (ISCC), IEEE symposium. IEEE, pp 180–187
2. Atakli IM, Hu H, Chen Y, Ku WS, Su Z (2008) Malicious node detection in wireless sensor networks using weighted trust evaluation. In: Proceedings of the 2008 Spring simulation multi conference. Society for Computer Simulation International, pp 836–843
3. Penna M, Jijesh JJ, Shivashankar (2017) Design and implementation of automatic medicine dispensing machine. In: RTEICT 2017—2nd IEEE international conference on recent trends in electronics, information and communication technology, proceedings, pp 1962–1966. https://doi.org/10.1109/RTEICT.2017.8256941
4. Christin D, Reinhardt A, Mogre PS, Steinmetz R (2009) Wireless sensor networks and the internet of things: selected challenges. In: Proceedings of the 8th GI/ITG KuVS Fachespräch Drahtlose sensornetze, pp 31–34
5. Kishore DV, Shivashankar, Mehta S (2016) MANET topology for disaster management using wireless sensor network. In: International conference on communication and signal processing, ICCSP 2016, pp 0736–0740. https://doi.org/10.1109/ICCSP.2016.7754242
6. Varun CA, Shivashankar, Sahana M, Varun RS, Rajesh T (2018) Implementation of swarm intelligence in obstacle avoidance. In: RTEICT 2017—2nd IEEE international conference on recent trends in electronics, information and communication technology, proceedings, 2017, pp 525–528. https://doi.org/10.1109/RTEICT.2017.8256652
7. Sridhara SB, Naveen KB, Ramesha M, Pai GN (2020) Internet of things: internet revolution, impact, technology road map and features. Adv Math Sci J 9(7):4405–4414. https://doi.org/10.37418/amsj.9.7.11
8. Ramesh Naidu P, Guruprasad N (2020) Design and implementation of cryptcloud system for securing files in cloud. Adv Math Sci J 9(7):4485–4493. https://doi.org/10.37418/amsj.9.7.17
9. Isaiadis S, Getov V (2005) Integrating mobile devices into the grid: design considerations and evaluation. In: European conference on parallel processing LNCS, vol 3648. Springer, pp 1080–1088
10. Kasinathan P, Costamagna G, Khaleel H, Pastrone C, Spirito MA (2013) DEMO: an IDS framework for internet of things empowered by 6LoWPAN. In: Proceedings of the 2013 ACM SIGSAC conference on computer and communications security. ACM, pp 1337–1340
11. Ramesha M, Sridhara SB, Naveena Pai G (2020) FPGA implementation of low power high speed BTED algorithm for 8 bit error correction in cryptography system. Int J Emerg Trends Eng Res 8(7):3893–3897. https://doi.org/10.30534/ijeter/2020/158872020
12. Kolias C, Kolias V, Kambourakis G (2017) TermID: a distributed swarm intelligence-based approach for wireless intrusion detection. Int J Inf Secur 16(4):401–416
13. Lazarescu MT (2013) Design of a WSN platform for long-term environmental monitoring for IoT applications. IEEE J Emerg Sel Top Circ Syst 3(1):45–54
14. Mamun MSI, Kabir AFM (2012) Hierarchical design based intrusion detection system for wireless ad hoc network. Int J Netw Secur Appl (IJNSA) 2(3):102–117
15. Nadeem A, Howarth MP (2013) A survey of MANET intrusion detection and prevention approaches for network layer attacks. IEEE Commun Surv Tutorials 15(4):2027–2045
16. Reina DG, Toral SL, Barrero F, Bessis N, Asimakopoulou E (2013) The role of ad hoc networks in the internet of things: a case scenario for smart environments. In: Internet of things and inter-cooperative computational technologies for collective intelligence. Springer, Berlin, Heidelberg, pp 89–113
17. Schumacher C, Kushalnagar N, Montenegro G (2007) IPv6 over low-power wireless personal area networks (6LoWPANs): overview, assumptions, problem statement, and goals. Available: https://tools.ietf.org/html/rfc4919
18. Verma SS, Patel RB, Lenka SK (2015) Investigating variable time flood request impact over QOS in MANET. Procedia Comput Sci 57:1036–1041

19. Watteyne T, Vilajosana X, Kerkez B, Chraim F, Weekly K, Wang Q, Pister K (2012) OpenWSN: a standards-based low-power wireless development environment. Trans Emerg Telecommun Technol 23(5):480–493
20. Smys S (2019) Energy-aware security routing protocol for WSN in big-data applications. J ISMAC 1(01):38–55
21. Chen JI, Iong Z (2020) Optimal multipath conveyance with improved survivability for WSN's in challenging location. J ISMAC 2(02):73–82

Chatbots: A Survey of the Technology

Hrithika Singh, Asmita Bhangare, Rashmi Singh, Shubhangi Zope, and Pallavi Saindane

Abstract In recent years, chatbots have become more widely used in a variety of fields, including marketing, customer service, support systems, education, healthcare, cultural heritage, and entertainment. Live chat interfaces have gained popularity as a way to engage clients in real-time customer care in many e-commerce contexts. Artificial intelligence-based chatbots, commonly referred to as conversational software agents, are created to have natural language conversations with human users. They are progressively taking the place of human chat operators (AI). In this study, we've covered a chatbot's fundamental definition as well as its numerous sorts and qualities. In this study, various methods for creating chatbots are also mentioned and described.

Keywords Chatbot · AI—artificial intelligence · HCI—human–computer interaction · NLP—natural language processing · AIML—artificial intelligence markup language · RiveScript · ChatScript · NLU—natural language understanding · Rasa · ChatterBot · IBM Watson assistant · Dialogflow

1 Introduction

Through the creation and examination of intelligent hardware and software, sometimes known as intelligent agents, artificial intelligence (AI) is gradually incorporating itself into our daily lives. The main motivation for doing research on this topic is that chatbots are trending in the modern world. The beginning of a new year is viewed as the start of new opportunities. Additionally, this suggests increased industry expansion, improvements to the current technology tools, and emerging AI technologies

H. Singh (✉) · A. Bhangare · R. Singh · S. Zope · P. Saindane
Vivekanand Education Society's Institute of Technology, Chembur, India
e-mail: hrithikasingh2501@gmail.com

P. Saindane
e-mail: pallavi.saindane@ves.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_49

with greater potential. According to chatbot trends, more complex real-time interactions based on computational modeling will replace simple customer-focused inquiries. In the last few years, chatbots have grown rapidly and have succeeded in replacing various industries and internal positions in the company. So this paper focuses on the various types of chatbots, their trends, their characteristics, and all the other information one might need to deeply understand the use of chatbots.

From simple labor to intricate operations, intelligent agents are capable of performing a wide range of tasks. One of the most fundamental and widely used types of intelligent Human–Computer Interaction is a chatbot, a type of AI system that is widely used (HCI) [1].

Computer programs called chatbots converse with users in everyday language. This technique, which dates back to the 1960s, sought to ascertain whether chat-bot platforms might deceive users into believing they were speaking to real people. A chatbot is defined as “A computer program designed to simulate interaction with human users, particularly over the Internet” [2]. Instead of offering direct touch with a real human agent, a Chatbot is a software application that conducts an online chat conversation via text or text-to-speech. In order to replicate a real-world conversation (or chat) with a user using messaging apps, websites, mobile apps, or the phone, chatbot systems use conversational artificial intelligence (AI) technology. It uses rule-based language applications to carry out live chat activities in response to user interactions occurring in real-time. Using Natural Language Processing (NLP) and sentiment analysis, it converses with users or other chatbots in human language via text or audio. In addition to chatbots, other names for them include interactive agents, intelligent bots, and digital assistants [2].

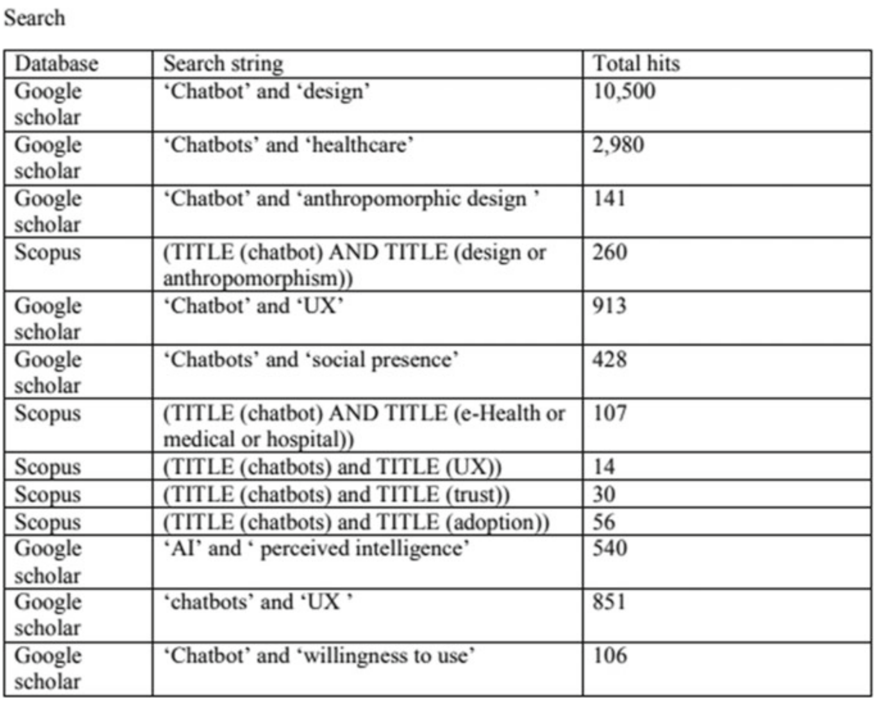
Virtual assistants and chatbots, which were formerly considered a niche concept in technology, have already become commonplace in India. Chatbots have primarily been relegated to consumer uses, according to cloud enterprise software vendor Ramco Systems. There are over 50,000 known bots on the market, the majority of which are geared toward consumers, and just a small percentage of them cater to business needs [3].

1.1 Chatbot Statistics

In recent times terms like ‘chatbot’, ‘healthcare’, and their design have been a part of many Google searches by users. The table below shows the statistics of these searches through various databases. For example, the search string ‘Chatbot’ and ‘design’ has appeared 10,500 times by the users using the Google Scholar database as in Fig. 1.

Figure 2 shows that after 2016, there was a noticeable growth in the use of chatbots. From 2000 to 2019, Scopus search results for the keywords “chatbot,” “conversation agent,” and “conversational interface” were organized by year [4].

The fastest-growing brand communication channel is chatbots: The use of chatbots as a brand communication channel has surged by a staggering 92% during 2019,

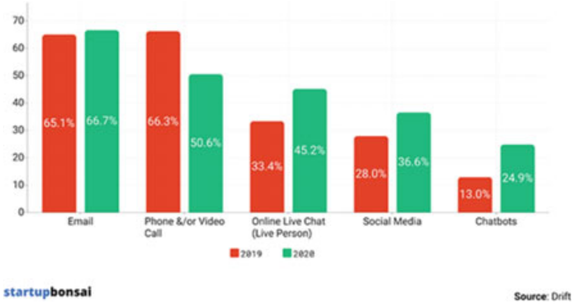


Search

Database	Search string	Total hits
Google scholar	'Chatbot' and 'design'	10,500
Google scholar	'Chatbots' and 'healthcare'	2,980
Google scholar	'Chatbot' and 'anthropomorphic design '	141
Scopus	(TITLE (chatbot) AND TITLE (design or anthropomorphism))	260
Google scholar	'Chatbot' and 'UX'	913
Google scholar	'Chatbots' and 'social presence'	428
Scopus	(TITLE (chatbot) AND TITLE (e-Health or medical or hospital))	107
Scopus	(TITLE (chatbots) and TITLE (UX))	14
Scopus	(TITLE (chatbots) and TITLE (trust))	30
Scopus	(TITLE (chatbots) and TITLE (adoption))	56
Google scholar	'AI' and ' perceived intelligence'	540
Google scholar	'chatbots' and 'UX '	851
Google scholar	'Chatbot' and 'willingness to use'	106

Fig. 1 Search results of chatbots

Fig. 2 Usage of chatbot as a brand communication



according to Drift's 2020 State of Conversational Marketing study. In 2020, 24.9% of customers interact with firms via chatbots, an increase from 13% in 2019.

Today's consumers want 24 h customer service in many different industries, including banking and finance, and health and wellness. Due to this, companies are scrambling to create chatbots and virtual assistants that can respond to inquiries from clients at any time of day. Virtual agents are preferred by almost 40% of internet users, and as more and more important sectors like retail and healthcare adopt digital technology, chatbots are expected to gain more and more traction in the coming years [5].

A lot of businesses have seen significant chatbot growth, and 1.4 billion individuals currently utilize them on a regular basis [6]. Chatbots are increasingly prevalent on many websites and social media pages, despite the fact that the industry is still developing. 80% of internet users have interacted with chatbots at least once during that year, according to figures released by Userlike [7].

1.2 History of Chatbots

Alan Turing suggested the Turing test in his 1950 work. If a panel of humans speaking with an unknown entity (via the keyboard, for example) thinks the entity is human yet the entity is a machine, the Turing test has been passed, according to the statement [2].

In 1966, the first chatbot ELIZA resembled the work of a psychiatrist, returning the user's statements in the interrogative form [1]. PARRY was introduced in 1972, playing the role of a schizophrenic sufferer [1].

Rollo Carpenter developed the Jabberwacky chatbot in 1988 which intended to be entertaining while simulating a natural human discussion [8].

Dr. Sbaitso is a Chabot that Creative Labs developed in 1992 for MS-Dos. It was one of the first attempts to merge A.I. into a chatbot, and it is renowned for its fully voice-operated chat program [9].

A.L.I.C.E. is a chatbot that uses universal language processing and heuristic pattern matching to carry on discussions. It includes a web-based discussion feature that allowed for longitude and covered any subject but lacked any real awareness of the entire discourse [1].

The SmartChild, developed in 2001, was a predecessor to Siri in many aspects [9]. In 2010, Apple developed Siri for iOS. It has a user interface that uses natural language and is both an intelligent personal assistant and learning navigator. All further AI bots were made possible because of it [9].

IBM invented a chatbot called Watson in 2011. Watson was capable of comprehending natural human language [1].

Google Now was launched in 2012. It responds to inquiries, executes actions via requests to a set of web services, and gives recommendations.

In 2014, Microsoft created the personal assistant Cortana. This program uses speech recognition and the proper algorithms to listen for and respond to spoken instructions [1, 9].

Google Assistant, which debuted in 2016, is the next version of Google Now. It boasts more advanced artificial intelligence and a nicer, more conversational interface, and it predicts user needs and delivers information to them [10–12].

Amazon's Alexa is an intelligent personal assistant. It was launched in 2014 [9]. It is embedded into home automation and entertainment equipment, making the Internet of Things (IoT) more accessible to humans [13].

Early in 2016, a breakthrough in Artificial Intelligence Technology transformed the way people communicate with manufacturers. Developers were able to construct chatbots for their brand or service on social media platforms, allowing customers to complete certain everyday actions within their messaging apps [14].

2 Types of Chatbots

The growth in chatbot technology is as dynamic as the evolution of chatbot capabilities. The types of chatbot is shown in Fig. 3.

Chatbots are categorized based on their level of involvement and response generation mechanism.

2.1 *Knowledge Domain*

Chatbots are categorized based on the amount of data they are trained on or the information they have access to **Open Domain**. These bots can converse about a variety of topics and answer accordingly. An open-domain dialogue system, as opposed to the traditional task-oriented bots, attempts to develop lasting relationships with users by fulfilling their need for social connection, affection, and communication. Meena, for example, is a Google open domain bot that can discuss anything **Closed domain**. These bots are focused on a specific knowledge topic, they may not be able to answer questions from other domains. A restaurant reservation bot, for example, will not inform you who America's first black president was. For example, Domino's ordering assistant bot as in Figs. 4 and 5.

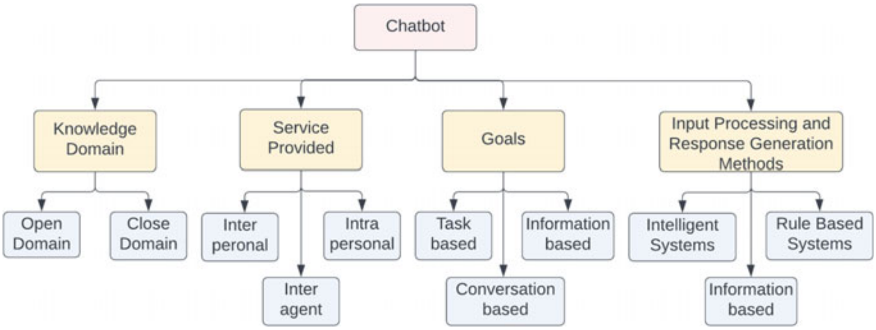
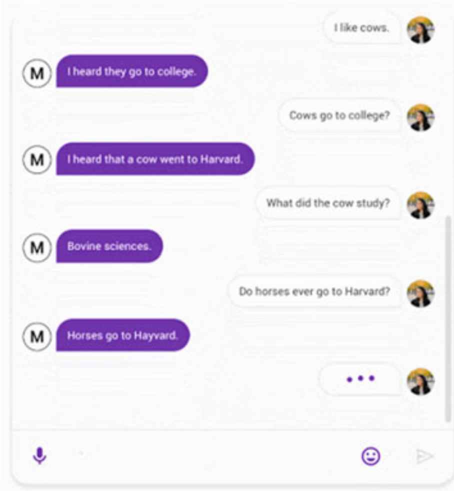


Fig. 3 Types of chatbot



I like cows.

M I heard they go to college.

Cows go to college?

M I heard that a cow went to Harvard.

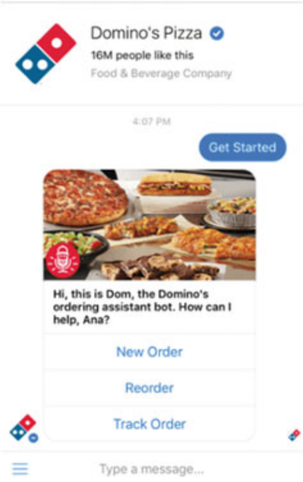
What did the cow study?

M Bovine sciences.

Do horses ever go to Harvard?

M Horses go to Hayvard.

Fig. 4 Meena



Domino's Pizza

16M people like this

Food & Beverage Company

4:07 PM

Get Started

Hi, this is Dom, the Domino's ordering assistant bot. How can I help, Ana?

New Order

Reorder

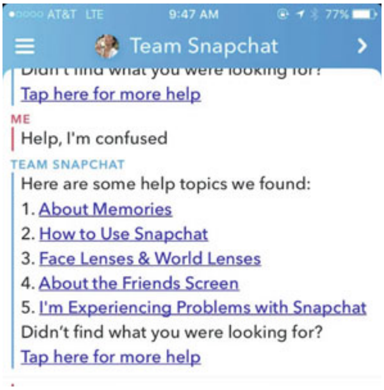
Track Order

Type a message...

Fig. 5 Domino's ordering assistant chatbot

2.2 Service Provided

Based on the bot's emotional proximity to the user they are classified into interpersonal, intrapersonal, and Inter-agent boats **Interpersonal Bots**. These include chatbots that operate in the area of communication that falls within the Proxemics chart's Social or Personal distance range. These chatbots aren't designed to be the user's companions; instead, they're supposed to collect data and send it on to the user; they're just enablers. For example, FAQ bots **Intrapersonal chatbots**. These operate in the user's domain and are typically seen in chat applications like Messenger, Slack, and WhatsApp and carry out tasks that are particular to the user's domain. For

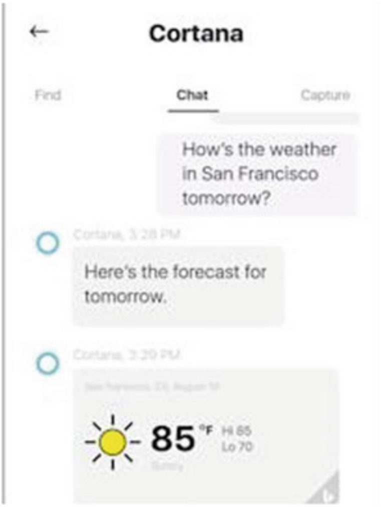


Didn't find what you were looking for?
Tap here for more help

ME
Help, I'm confused

TEAM SNAPCHAT
Here are some help topics we found:
1. About Memories
2. How to Use Snapchat
3. Face Lenses & World Lenses
4. About the Friends Screen
5. I'm Experiencing Problems with Snapchat
Didn't find what you were looking for?
Tap here for more help

Fig. 6 Ask Sunny



How's the weather in San Francisco tomorrow?

Cortana, 3:28 PM
Here's the forecast for tomorrow.

Cortana, 3:29 PM
85 °F Hi 85 Lo 70

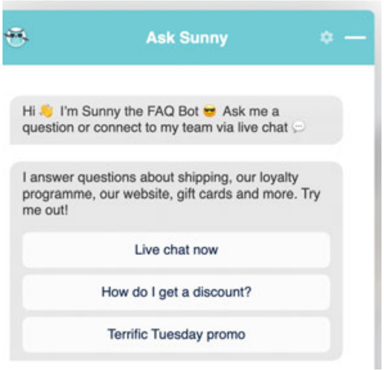
Fig. 7 Cortana

example, Team Snapchat of Snapchat **Inter-agent**. This type of bot will be common in IoT-heavy locations. In this scenario, two systems communicate with one another to complete a task. For example, Inter-agent communication is demonstrated through the Alexa-Cortana interface as in Figs. 6, 7 and 8.

2.3 Goals

Based on the primary goal, boats are classified into Informative boats, Conversational boats, and task-based boats **Informative**. These bots are intended to give the user data that has been saved or is available from a fixed source. For example, the IRCTC chatbot gives information about the services provided by IRCTC **Conversational**

Fig. 8 Snapchat



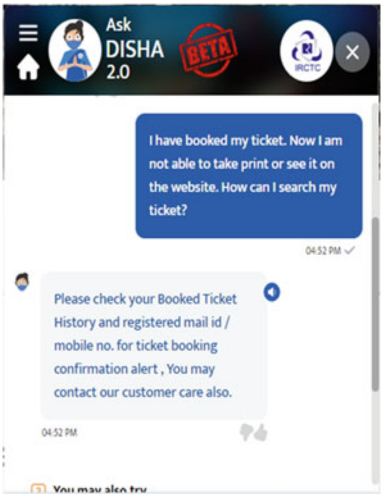
Hi 👋 I'm Sunny the FAQ Bot 😎 Ask me a question or connect to my team via live chat 💬

I answer questions about shipping, our loyalty programme, our website, gift cards and more. Try me out!

Bots. These converse with the user as if they were another human being. Their purpose is to answer the statement that has been presented to them accurately. For example Vainu—Enrich customer conversations without form fill-ups **Task-based**. They specialize in one job, such as arranging a flight or assisting you with shopping. For example, H&M—Recommend your users based on their interests in Figs. 9, 10, and 11.

Future of conversational AI. In the coming years, brands will continue to integrate AI and provide more meaningful customer service. Multilingual language capability and language switching are important priorities for brands if they operate from many regions or if the location has speakers of other languages. When it comes to customer service, emotional intelligence is a valuable asset. Conversational AIs are becoming more understanding of client intent, attitudes, and other emotions as a result of updated machine learning models. It enables them to respond more appropriately

Fig. 9 IRCTC



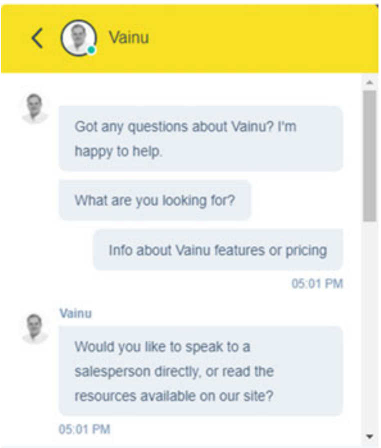
I have booked my ticket. Now I am not able to take print or see it on the website. How can I search my ticket?

04:52 PM ✓

Please check your Booked Ticket History and registered mail id / mobile no. for ticket booking confirmation alert , You may contact our customer care also.

04:52 PM

Fig. 10 Vainu



Vainu

Got any questions about Vainu? I'm happy to help.

What are you looking for?

Info about Vainu features or pricing

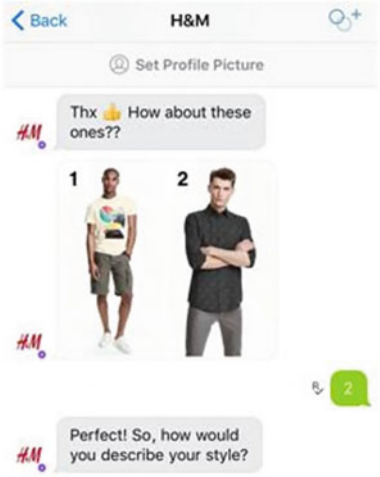
05:01 PM

Vainu

Would you like to speak to a salesperson directly, or read the resources available on our site?

05:01 PM

Fig. 11 H&M



Thx 👍 How about these ones??

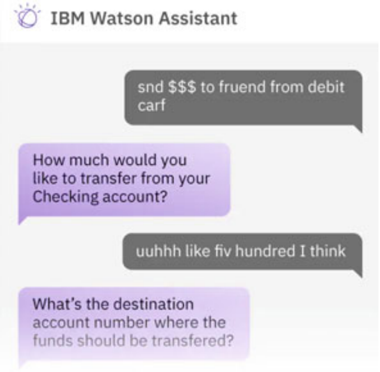
2

Perfect! So, how would you describe your style?

and gather crucial data. In the future years, AI may become self-aware enough to respond positively and perhaps create a welcoming environment. For years, these bots can remember past data, client emotional history during each engagement, and other sentiment-driven data. It can accurately select how to respond to the consumer and provide a safe atmosphere using this knowledge.

2.4 Method for Processing Input and Producing Responses

Based on input processing and generation of replies, chatbots are classified into Intelligent Systems, Rule-Based Systems or Hybrid systems. To understand the



IBM Watson Assistant

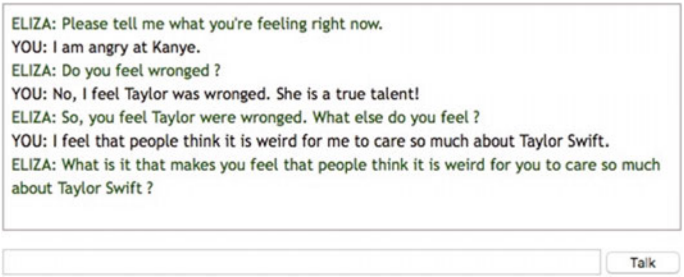
snd $$$ to fruend from debit carf

How much would you like to transfer from your Checking account?

uuhhh like fiv hundred I think

What's the destination account number where the funds should be transfered?

Fig. 12 IBM Watson assistant



ELIZA: Please tell me what you're feeling right now.
YOU: I am angry at Kanye.
ELIZA: Do you feel wronged ?
YOU: No, I feel Taylor was wronged. She is a true talent!
ELIZA: So, you feel Taylor were wronged. What else do you feel ?
YOU: I feel that people think it is weird for me to care so much about Taylor Swift.
ELIZA: What is it that makes you feel that people think it is weird for you to care so much about Taylor Swift ?

Fig. 13 ELIZA

inquiry, really **Intelligent systems**. They generate responses and use natural language comprehension. When the domain is narrow and there is plenty of data to train a system, these systems are used. For example IBM Watson Assistant **Rule-based systems as in** Fig. 12 and 13. These **use** Pattern matching, which makes them inflexible. When the number of possible outcomes is fixed and the number of situations is imaginable, these can be used. For example ELIZA **Hybrid systems**. These use a combination of rules and machine learning to achieve their goals. A system that manages the direction of conversation using a flow chart yet gives responses generated using natural language processing is an example [15].

3 Characteristics of Chatbot

Chatbots have evolved to meet several business models, and they are currently used in a wide range of industries and sectors. The chatbot can perform a range of functions depending on its goal.

3.1 Conversational Development

A great chatbot needs to have specialized natural language processing (NLP) skills to interpret and participate in conversations in several languages while also understanding the context of those conversations. To offer the appropriate first response, it may also identify the goal of a request and present alternatives to confirm or explain that purpose [16].

3.2 Autonomous Judgement

The chatbot can carry out complicated thinking without requiring human intervention. For instance, an effective chatbot for customer support should be able to infer answers from pertinent case studies [16].

3.3 Omni Channel

A chatbot is relevant enough to give users unified conversations across platforms or communication interfaces where the bot is connected. A chatbot can identify, recall, and relate to each client depending on their preferences, which are gleaned from past data, whether on the website or through a messaging channel [17].

3.4 Accessible

All clients, regardless of language competence or any visual or hearing disabilities, must now be able to use chatbot interfaces equally. For example, bots should be able to read text aloud to the visually impaired [16].

3.5 Free to Explore

Today's chatbots may search the internet and company intranets for answers to client inquiries. As the customer knowledge base expands, the chatbot's responses get better, and it might even figure out which articles were useful and which weren't [17].

3.6 Encourage Analytics

A robust dashboard is offered by a time-relevant chatbot, allowing company leaders to monitor talks and progress by providing aggregate interaction data. Statistical information and performance evaluations are available to business owners, allowing them to enhance their bots and keep up their exceptional customer service [17].

3.7 Secure

In the aftermath of recent data hacks, bots are safer than ever. Nowadays, a lot of businesses are quite transparent about the data they gather and how they use it. Users frequently have the choice to refuse data collection if they so choose [17].

3.8 Allows Upgrades

Since developers are constantly trying to improve the operational abilities of various chatbots, a superb chatbot can upgrade without the need to remove the previous version before installing the new one. Unless an administrator user forces a modification, a standard chatbot upgrade does not result in data loss or changes in technical and functional [17].

3.9 Pre-trained

The chatbot has already received training to identify brand- or industry-specific terminology. Even better, it's already configured to handle typical customer demands in a certain sector [16].

3.10 Emotionally Intelligent

The chatbot can understand sentiment and tone during a conversation to personalize the experience or escalate to a real agent if necessary. It can also deduce aspects of the client's personality [16].

3.11 Integrates with CRM

The chatbot may establish connections to crucial platforms and control workflows inside and outside the CRM. It can deal with everything, from a straightforward password update to a complex, multi-step workflow involving numerous programs, in real-time [16].

3.12 Sentences Are to the Point

Sentences in good chatbot scripts are always short. Customers and consumers are prone to become overwhelmed when confronted with huge blocks of text and lose interest in the dialogue. In your script, removing needless words and using bullet points instead of long answers will assist maintain a brief and simple interaction and provide better customer service [18].

3.13 A Distinct Personality

The script must consider what clients would find helpful, or motivating, just as it would in genuine interaction. The bot's personality should be consistent with the brand's. It's just as crucial to stay in character as it is to get rid of the chatbot script's bland and generic dialogue. This will make it easier to provide more effective real-time client service [18].

3.14 Handles the Unexpected

The chatbot script is set up to deal with consumer misconceptions or ambiguities. Suggestions for similar inquiries or FAQs, as well as the preparation of an error message in the event of a misunderstanding, could be used to allow customers to return to the conversation without having to restart it [18].

3.15 Saves Money and Time

Conversations that would otherwise necessitate an employee's response can be automated using a chatbot, saving the company time and money. This time, money, and personnel can then be put to better use, resulting in improved outcomes [19].

4 Chatbot Approaches, Tools and Architecture

This section discusses the many methods of chatbot development, as well as the numerous tools available for chatbot development and, subsequently, the architecture of conversational bots.

4.1 Approaches

Rule-based and machine learning approaches are two approaches to constructing a chatbot, depending on the algorithms and methodologies used.

Rule-based Approach. The rule-based approach is presented and compared in the basic functionalities of AIML, RiveScript, and ChatScript. The pattern matching mechanism is commonly employed in chatbots. It creates effective and accurate responses for user input using pattern recognition or pattern matching. These languages are the most prevalent because they are lightweight and simple to configure.

Machine learning-based approach. Machine learning-based approach because they examine the entire debate context rather than just the current turn, they do not require a predefined response for every possible user input. Natural Language Processing (NLP) is a branch of artificial intelligence that studies how computers use natural language text or speech.

Keys used in NLP ChatBot: Utterance; Intent; Entity; Context; Session.

The core of any NLP work is Natural Language Understanding (NLU). According to the user's intent, NLU attempts to accurately extract context and meaning from unstructured natural language user inputs and responses. In other words, the intent is a relationship between what a user says and what action the chatbot should do.

4.2 Chatbot Tools

Rasa is a Python-based framework for creating custom AI chatbots that understand natural language (NLU). ChatterBot is a Python package that makes creating automated responses to user input simple. Rasa is an AI chatbot development framework that employs natural language understanding (NLU). IBM Watson Assistant is a virtual assistant powered by IBM's Watson AI Engine and available on IBM Cloud. Dialogflow is used by businesses such as Dominos, Shelf, Vonder, and SnapEngage.

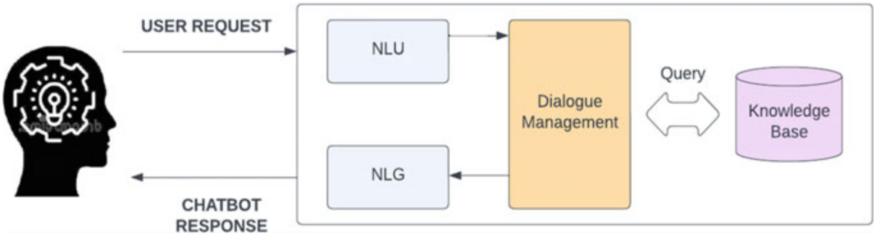


Fig. 14 Architecture of chatbot

4.3 Chatbot Architecture

Natural language understanding is a subfield of artificial intelligence that employs computer software to comprehend input in the form of sentences delivered via text or speech. When the Conversational Agent is in charge of a conversation, the DM system also acts as a state tracker, continuously maintaining the state of the conversation and triggering a transfer from one state to another. The NLU unit of dialogue systems generates the semantic representation for the dialogue task based on user input. The dialogue manager receives input from the NLU components, maintains some kind of state or interface with task managers, and then passes the output to the NLG. The architecture of chatbot is shown in Fig. 14.

5 Metrics for Evaluation of Chatbot

This section explains how to test the performance of your conversational agent by using leading indications. An indicator that can be used to organize training to assist your chatbot to become a valuable team member.

5.1 Chatbot Activity

This metric assists in determining the number of interactions, or the time it takes between a user asking a simple question and a constructive dialogue taking place [20]. This aids in determining the frequency with which chatbots are used and the number of users who use them [21].

5.2 Chatbot Interaction Rate

This metric aids in determining users' attention during a conversation with a chatbot. It will enable you to determine the average amount of messages sent per chat [21].

5.3 Chatbot Session Length

This metric aids in determining user interaction by comparing session length with other metrics. Session length provides additional information regarding conversational chatbot [22].

5.4 Chatbot Goal Completion Rate

This metric aids in determining how many inquiries a chatbot must be asked before providing the required information. The higher the number of questions that users must ask, the longer it will take [22].

5.5 Chatbot Retention Rate

This metric aids in determining how many people returned to utilizing the bot after a certain amount of time has passed. When compared to the normal frequency of client contact in your line of business, this is a significant difference. A high retention rate indicates that your strategy is working [22].

5.6 Chatbot Non-response Rate

This metric aids in determining how many times the chatbot fails to answer a query. The chatbot often fails due to a lack of information [21].

5.7 Questions Per Conversation

This metric aids in determining how many inquiries a chatbot must be asked before providing the required information. The more questions that users must ask, the longer it will take [21].

6 Comparison of Chatbots

AI chatbots come in a variety of shapes and sizes. Some have clever artificial intelligence (AI), while others simply have branching questions with answers. Whether you're utilizing an AI eCommerce chatbot or a simpler chatbot for assistance, the most important thing to consider is the user's experience and whether they require an engaging discussion or a smart approach to fill out the information in a form or ticket. When a user requires additional assistance, live chat is frequently used to assist the dialogue [23]. The comparison on different chatbot is shown in Table 1.

7 Chatbots in Healthcare

7.1 Can We Trust Health in the Hands of Chatbots?

Users' readiness to disclose personal information, accept advice, and follow suggestions can be defined as trust in the context of health chatbots. The trust-building process is influenced by reliability, transparency, and explainability. The chatbots' suggestions for recommendations must be consistent. The algorithms used by chatbots mainly rely on data. The data's integrity, correctness, privacy, and security are critical. Users will not disclose sensitive healthcare information to use the health chatbots if this is not the case. Chatbots are enhancing productivity for healthcare staff and creating trust with patients in a future where virtual medical appointments are becoming the norm. Medical chatbots have the ability to benefit both patients and clinicians by lowering the burden and increasing the likelihood of a correct diagnosis. Trust generally starts where knowledge stops, and it acts as a means of bridging knowledge gaps [24].

7.2 Benefits of Chatbot in Healthcare

By decreasing hospital visits, unnecessary treatments, and procedures, medical chatbots lessen the burden. Both time and money are saved. Doctor consultations are simple to schedule. By linking patients with the appropriate healthcare providers and assisting them in understanding their diseases and treatments without seeing a doctor, the chatbot reduces hospital readmissions. Additionally, medical chatbots are used by hospitals and private clinics to triage and clerk patients before they enter the consultation room. These bots use automatic responses to ask pertinent questions regarding the symptoms of the patient in an effort to provide the doctor with complete history [25].

Table 1 Comparisions of different chatbot

Title	Introduction	Pros	Cons
Ochatbot	Resolves inquiries and overcomes sales obstacles to engage consumers	Easy installation on eCommerce platforms Strong reporting to obtain insights and build business	At the time of posting, there is no SMS texting or Whatsapp interface
ManyChat	Is powered by Facebook Instant Messenger, via which they communicate with consumers. As a Facebook Messenger widget, it has been added to websites	Visual dialogue builder for simple branching conversations Works on Facebook Does have SMS messaging	Limited to Facebook The conversation is restricted to a pre-defined branched conversations
Botsify	Botsify customer supports chatbots with Live Chat options to generate leads. Businesses can employ their customized bots for teaching, financial, and customer service purposes	Connect to live chat Connects to WhatsApp, Facebook Messenger, and SMS Messaging	The AI cannot simply transition from one interaction to another to gain marketing insights Cart reminders and notifications are not included
LandBot	Landbot is a chatbot platform that converts leads and collects data to generate leads. Landbot can be launched on platforms such as WhatsApp, Facebook Messenger, and websites	It has flow diagrams and many more customisable elements to match the demands of a business When the bot is unable to answer the customer's questions, Landbot offers a function that allows the customer to connect with Live Chat	Landbot is not an AI chatbot and cannot change conversions Because it is mostly based on button clicks, it cannot engage in natural conversation like a flow bot No way for the user to type their question or search for products
Intercom	Is primarily a live chat system, but it is also evolving into a chatbot platform. Intercom's Live Chats are well-known in the enterprise market, and they are now branching out into the chatbot industry	Switching languages within a discussion Features like targeted chats, reporting, and follow-up notifications are available Code-free personalization	They are gradually evolving into a more automated chatbot, however, this is not their primary focus when interacting with clients For live chat, the charge by the seat

(continued)

Table 1 (continued)

Title	Introduction	Pros	Cons
Drift	Offers a pre-defined conversational experience that transitions to live chat for its consumers	One free seat on Live Chats' customizable chat system is available. This includes unrestricted consumer contact	Live Chats occur in real-time. Unlike chatbots, having a human representative on the chat is required for 24/7 customer care
MobileMonkey	MobileMonkey, which is similar to Manychat and is powered by OmniChat technology, is frequently used for marketing, sales, and customer care	Facebook Messenger, SMS, and WhatsApp are all used to communicate	It does not work without Facebook, SMS, or WhatsApp. It is impossible to communicate with a customer who is not using these platforms
Tidio	Tidio offers clients a way to communicate with them via Live Chat and AI chatbots	Abandoned cart, order status, and product availability are all eCommerce features	The AI can't hop from one chat to the next since it's programmed to do so
Maisie	Maisie is a remarketing and sales-oriented chatbot platform	When a chatbot becomes stuck or a live representative is requested, Maisie allows a live agent to take over the conversation	Reporting insights are not present
Chatfuel	Chatfuel uses Facebook Messenger to automate conversations between their chatbot and humans	Code-free building	Chatfuel has no predetermined pricing because it is dependent on the number of users

8 Open Challenges Related to Chatbot

The way that customers communicate with businesses has been changed by chatbots. As a result, the vast majority of businesses have jumped into the race to develop or enhance these virtual agents on their websites. But the developers must work around the obstacles that they present. The first issue is context in chatbots. The integration of context and meaningful responses is essential to the evolution of any chatbot because a conversation without context would be ambiguous. Limited User Attention is another problem, as users only have a short attention span and demand quick responses. Another major problem that contains the majority of the complexity is chatbot testing. Because natural language models are being improved, chatbots are always changing. Thus, testing and using chatbots becomes essential. All of your web data is often secure, but if you add specific chatbots to it, you can't be sure whether or not the API will be secure. If appropriate security measures are not followed, data leaks and hacks are likely to occur. In addition to being equipped with the capacity to contact your target audience through brand identity and awareness, chatbots should be enhanced with pertinent data.

9 Proposed System

During the pandemic, the demand for healthcare facilities increased. India is densely populated and does not adhere to the WHO standard of one doctor per thousand people. This kickstarted the idea of creating a Disease Prediction System using a Chatbot. We built a website and integrated a chatbot that predicts the disease the user is suffering from, as well as additional information and doctor recommendations for treatment. We chose to build a healthcare system in which patients can diagnose their illnesses by reporting symptoms to a chatbot, and they will also receive all necessary healthcare services for treatment. The primary goal of our project is to help people diagnose their illnesses early on by using our chatbot. So that they can take the necessary precautions. Along with disease prediction, our system aims to provide recommendations for doctors and medical laboratories in their area and also connect patients and doctors virtually through video conferencing.

10 Conclusion

As little human interference with technology use as feasible is the aim of our technological environment. Chatbots are more efficient than humans and can reach a wider audience. At the same time, they have the potential to develop into helpful instruments for acquiring information. Running customer care departments is quite cost-effective for them. We think that the information provided in this study about the core ideas behind chatbots is helpful. Chatbots can be better understood by users and developers, who can then learn how to build and employ them for the purposes for which they were designed. Unquestionably helpful are chatbots. Healthcare professionals can use them on their websites to assist patients to get prompt medical information, booking appointments, sending daily reminders, and even issuing invoices. It might be impossible to distinguish between a chatbot and a real-life agent as AI and machine learning improve.

Acknowledgements Nothing in this world is accomplished single-handedly. Every success is a result of much-needed guidance and support. We would like to thank all those people whose guidance and support have helped me make this paper. Firstly, we would like to thank our mentor Mrs. Pallavi Saindane who laid down the guidelines for this project, without which we would not have been able to complete this paper. Secondly, we would like to thank our parents and our siblings, who encouraged our work till the very end. Last, but not least, we would like to thank our dear friends, who have supported us in different parts of this project.

References

1. Adamopoulou E, Moussiades L (2020) An overview of chatbot technology. In: Maglogiannis I, Iliadis L, Pimenidis E (eds) Artificial intelligence applications and innovations. AIAI 2020. IFIP advances in information and communication technology, vol 584. Springer, Cham. https://doi.org/10.1007/978-3-030-49186-4_31
2. Khanna A, Pandey B, Vashishta K, Kalia K, Bhale P, Das T (2015) A study of today's A.I. through chatbots and rediscovery of machine intelligence. Int J u- and e-Serv Sci Technol 8:277–284. https://doi.org/10.14257/ijunesst.2015.8.7.28
3. Analyticsindiamag. https://analyticsindiamag.com/is-chatbot-a-boon-or-bane-heres-why-companies-are-using-them-despite-tech-glitches/
4. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S2666920X21000175#fig2
5. Insiderintelligence. https://www.insiderintelligence.com/insights/chatbot-market-stats-trends/
6. ChatBot. https://www.chatbot.com/blog/chatbot-statistics/
7. Bloggingwizard. https://bloggingwizard.com/chatbot-statistics/
8. Daffodil. https://insights.daffodilsw.com/blog/the-history-and-evolution-of-chatbots
9. Onlim. https://onlim.com/en/the-history-of-chatbots/
10. ChatterBot. https://chatterbot.readthedocs.io/en/stable/
11. Machine learning chatbot. https://neptune.ai/blog/building-machine-learning-chatbots-platforms-and-applications#:~:text=What%20is%20a%20machine%20learning,Natural%20Language%20Processing%20(NLP)
12. Harms J-G, Kucherbaev P, Bozzon A, Houben G-J (2019) Approaches for dialog management in conversational agents. IEEE Internet Comput 23(2):13–22. https://doi.org/10.1109/MIC.2018.2881519
13. Towards data science. https://towardsdatascience.com/5-reasons-why-your-chatbot-needs-natural-language-processing-ed20fb0a3655
14. GeeksforGeeks. https://www.geeksforgeeks.org/chatbots-using-python-and-rasa/
15. Nimavat K, Champaneria T (2017) Chatbots: an overview types, architecture, tools, and future possibilities
16. IBM. https://www.ibm.com/blogs/services/2020/01/15/seven-characteristics-of-a-great-chatbot/
17. Visight. https://www.visight.io/blogs/62
18. HappyFox. https://blog.happyfox.com/must-have-characteristics-for-an-effective-chatbot-script/
19. Ascentspark. https://www.ascentspark.com/blog/article/6-characteristics-ideal-chatbot
20. Kulkarni P, Mahabaleshwarkar A, Kulkarni M, Sirsikar N, Gadgil K (2019) Conversational AI: an overview of methodologies, applications and future scope. In: 5th international conference on computing, communication, control and automation (ICCUBEA), pp 1–7.https://doi.org/10.1109/ICCUBEA47591.2019.9129347
21. Measuring chatbot effectiveness. https://corporate.livingactor.com/en/measuring-chatbot-effectiveness-kpi/
22. Metrics for you to evaluate the success of chatbot. https://www.mjvinnovation.com/blog/8-metrics-for-your-chatbot/
23. Ometrics. https://ochatbot.com/best-ai-chatbots/
24. Petrova V (2020) Can we trust our health in the hands of chatbots?: An exploratory study investigating the effect of anthropomorphic design of e-Health chatbots on patients UX
25. Benefits of chatbot in healthcare. https://topflightapps.com/ideas/chatbots-in-healthcare/#3

An Improved Machine Learning Algorithm for Crash Severity and Fatality Insight in VANET Network

S. Bharathi and P. Durgadevi

Abstract A vehicular ad hoc network (VANET) is a wireless network that connects a group of moving or stationary vehicles together. VANETs were primarily used to provide safety and comfort to drivers in automotive environments until recently. Clustering is an important concept in vehicular ad hoc network (VANET) where several vehicles join to form a group based on common features. Clustering increases the complexity of data. The assessment of road accident strategies in Machine learning is presented in this paper. A road collision is the most unwanted and unexpected occurrence that may happen to a vehicle, due to the fact that they happen regularly. The goal of this study was to investigate the correlation between the concentration of collisions and injury. There are several factors that influence crashes, including weather, road conditions, driver distraction, and misread vehicle signals.

Keywords Vehicular Adhoc network · VANET · Machine learning · Accident data · Collision

1 Introduction

Much accident occurs in everyday activities and the number of victims increasing day by day. In the transport system, road safety is a crucial component. many people are harmed or died in accidents. The majority of these accidents are caused by human fault such as over speeding, violating the traffic rules, not wearing seat belt, drunk driving, and distracted driving [1]. Accidents can take place in variety of situations, and there is no guarantee that they might stop in the future because several people are controlled, such as traffic conditions, lighting conditions, wet roads, the number of people [2] on board, and the speed at the time of the collision. Figure 1:

S. Bharathi (✉) · P. Durgadevi
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Chennai 600026, India
e-mail: bs3778@srmist.edu.in

P. Durgadevi
e-mail: durgadep@srmist.edu.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_50

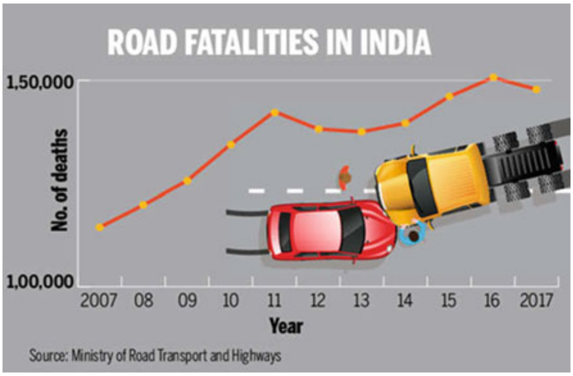


Fig. 1 Fatality statistic of Indian accident

Fatality Statistic of Indian accident In India, primarily cop officials gather, organize, and record accident data, therefore there is much scope for further improvement. They take extremely basic information and provide it in a non-analytical style for investigation reasons, and different researchers have done data analysis on that sort of dataset, but it becomes very tough to analyze particular circumstances.

2 Related Work

Clustering is a major initiative for data analysis and an unsupervised learning method. Clustering entails putting together a collection of comparable items that are distinct from one another [1]. Clustering begins with defining the number of clusters, followed by determining the cluster's centroid, and finally, ensuring that each cluster is the same size. In this study, a cluster of accident locations is created, which aids in the prediction and identification of risky places and scenarios in terms of road accidents. After that, take appropriate measures. Analyze as much data as possible and forecast important information [3–5]. The major goal of this study was to look at the elements that contribute to road Accidents and to forecast accident severity by applying Machine learning techniques. In this research, several classification algorithms were used to the accident visualization data set and the accuracy of three classifiers was compared with Decision tree, k-nearest neighbour, Naive Byes [6]. In this article, an analysis is carried out on incidents that occur at a given area on a regular basis or at the same spot every time [7, 8]. As a result, these types of data set analyses aid in the prediction of a given place, as well as the identification of characteristics that influence the likelihood of an accident in that region [9]. And for proceeding this, we used Kmeans algorithm for making a group of cluster about location and these clusters are divided into three different parameters depending upon the speed control such as low speed location, moderate speed cluster location and high speed cluster location [10–13] with this cluster classification the accident frequency count was then discovered as a metric to assist in the formation of location clusters. In this paper, analyze about light conditions during accidents. we used K-means algorithm for making a light cluster and these clusters are alienated into five different parameters depending upon the lightening conditions of the road such as full day, Night with public lightening on, night without public lightening on, Twilight or dawn and night with public lightening not lit. This cluster of clusters provided a study of accidents that occurred more frequently during lightning conditions. This Study presents an overview of road accident analysis by using the methodologies of Machine learning to predict and prevent the accident. There are many algorithms available in machine learning for classification and clustering the data. Among those algorithms, Random forest were utilized in this study for accident analysis. The literature survey of the existing methods is shown in Table 1.

3 Crash Severity and Fatality Insight

The severity of road accidents as defined by the number of people killed per 100 crashes went from 33.7 in 2019 to 36.0 in 2020. It emphasizes the importance of enhanced trauma healing and traffic-calming measures lowering the crash impact characteristics. Every year, over 1.3 lakh people are killed in road accidents in India, with over 3.4 lakh people wounded. The frequency of accidents and fatalities has decreased over time as a consequence of deliberate and coordinated road safety initiatives.

We intended to explore the data to determine whether there are any factors that correspond to the number of injuries or fatalities recorded. Some data is accessible at https://morth.nic.in/state-wise-data and https://data.gov.in/.

In this kernel, we will be studying road accident data from 2017 to 2019. Additionally discovered some statistical evidence for road accidents, such as weather conditions and road conditions. There is some more supplementary data available that you may utilize for better understand.

The given Fig. 2 shows the analysis data set of Indian road accident data in state wise.

India has the highest number of road fatalities in the world. In 2018 the latest year for which global figures are available, India accounted for more than a third of global road accident deaths. The World Health Organization says such deaths are under-reported and estimated that in 2016, the figure for India was likely twice as big as that reported by the government.

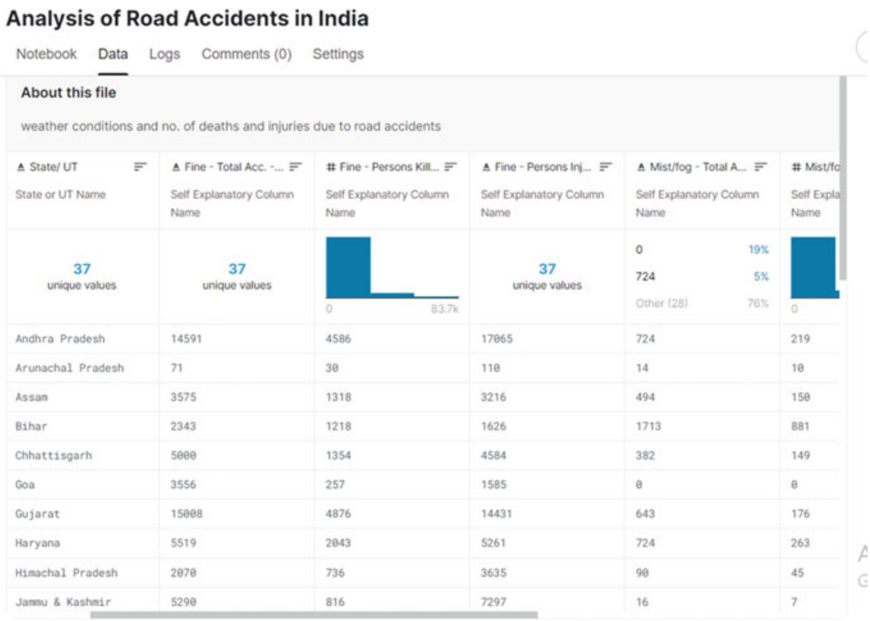
3.1 Accident Location Analysis

The ultimate focus is to get latitude and longitude intelligence in order to contribute an immediate help that lowers fatalities. Clustering is a relevant data analysis venture and an unsupervised machine learning tool. Clustering entails putting together a

Table 1 Overview of literature survey

Proposed research	Technique used	Task/description	Pros	Cons
[1]	Decision tree, KNN	Univariate feature selection, and feature importance	Reduces overfitting	Ignore the interaction between classifier
[14]	XGboost	Speed limit, collision type, vehicle model and vehicle movement	It is very fast and accurate than Ada boost	It is sensitive
[15]	Association rule mining	Finds the rules to associate between object and set of item	Useful for analyzing and predicting user on VANET	Too many parameters used leads to high computation
[9]	k-mean and association rule mining	Make an array of regions to determine the co-relationship between accident factors	Useful in analyzing data sets	Computational time is very high
[16]	Weka tool	Preprocessing, cataloging, grouping and visualization	Feature selection and data mining are integrated	It can handle only small datasets
[17]	IoT prototype, sensors	Detect accident and severity of the emergency level	Find out the location of the accident and nearest medical center. It sends a message for an ambulance after detecting basic information	All end user must be connected to the network
[18]	Self-organization map	Data are easily expounded and clarify	Grid clusters observes similarities in data	Neuron weights be necessary and enough to cluster input
[19]	Navie bayes, Ada boost	Recursive feature elimination	Noise data and outliers are eliminated	Extremely sensitive

collection of comparable items that are distinct from one another [1]. Clustering begins with defining the number of clusters, followed by determining the cluster's centroid, and finally, ensuring that each is the similar size. In this study, a cluster of accident locations is created, which aids in the forecast in addition identification of risky places and scenarios in terms of road accidents. After that, take appropriate measures. Analyze as much data as possible and forecast important information.



Analysis of Road Accidents in India

About this file

weather conditions and no. of deaths and injuries due to road accidents

State/ UT	Fine - Total Acc. -...	Fine - Persons Kill...	Fine - Persons Inj...	Mist/fog - Total A...	Mist/fo
State or UT Name	Self Explanatory Column Name	Self Explanatory Column Name	Self Explanatory Column Name	Self Explanatory Column Name	Self Expla Name
37 unique values	37 unique values	0 83.7k	37 unique values	0 19% 724 5% Other (28) 76%	0
Andhra Pradesh	14591	4586	17065	724	219
Arunachal Pradesh	71	30	110	14	10
Assam	3575	1318	3216	494	150
Bihar	2343	1218	1626	1713	881
Chhattisgarh	5000	1354	4584	382	149
Goa	3556	257	1585	0	0
Gujarat	15008	4876	14431	643	176
Haryana	5519	2043	5261	724	263
Himachal Pradesh	2070	736	3635	90	45
Jammu & Kashmir	5290	816	7297	16	7

Fig. 2 Data set of Indian road accident data in state wise

The major goal of this study was to look at the elements that contribute to road.

Accidents and to forecast accident severity by applying Machine learning techniques.

In this research, several classification algorithms were used to the accident visualization. The accuracy of different classifiers and the data set were compared using Decision tree and k-nearest neighbor, Naive Byes [6].

In this article, an analysis is carried out on incidents that occur at a given area on a regular basis or at the same spot every time. As a result, these types of data set analyses aid in the prediction of a given place, as well as the identification of characteristics that influence the likelihood of an accident in that region [9]. And for proceeding this, we used K-means algorithm for making a collection of clusters about site and these groups are divided into three different parameters depending upon the speed control such as low, moderate speed, high-velocity cluster location. With these clusters classification the accident occurrence total was then discovered as a metric to assist in the formation of location clusters. Figure 3 given below discuss about the location analysis of accident.

In accident location analysis, we taken different parameters such as Accident happened due to lighting conditions, weather conditions, lane analysis, over speed.

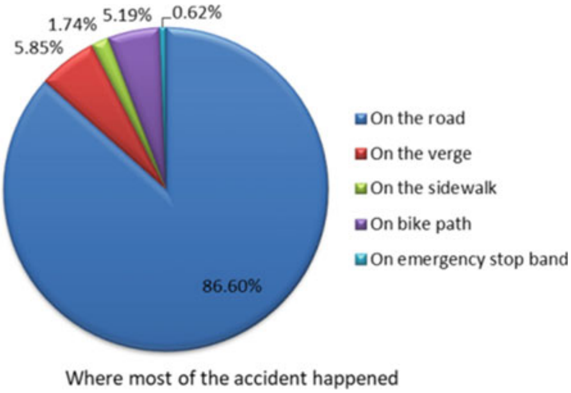
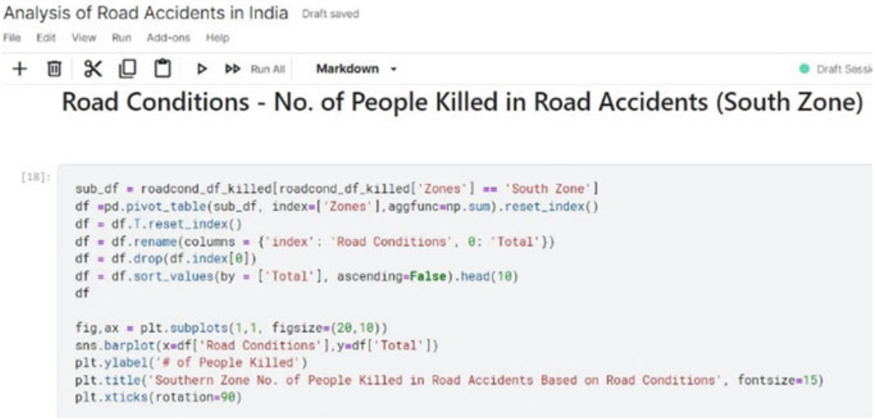


Fig. 3 Location analysis

3.2 *Accident Happened Due to Lighting Conditions*

This study analyzed about light conditions during accidents. We used K-means algorithm for making a light cluster and these clusters are alienated into five different parameters depending upon the lightening conditions of the road such as full day, Night with public lightening on, night without public lightening on, Twilight or dawn and night with public lightening not lit. This cluster of clusters provided a study of accidents that occurred more frequently during lightning conditions.

Besides that, we used the characteristics in our dataset to try to figure out what went wrong the accident. In Fig. 4, statistically we have found some feature conditions where the number of accidents gets increased. Figure 5 illustrates several statistically significant situations that lead to an increase in the number of accidents.


Analysis of Road Accidents in India
Draft saved
File Edit View Run Add-ons Help
Run All
Markdown
Draft Sessi
[18]:

Road Conditions - No. of People Killed in Road Accidents (South Zone)

```
sub_df = roadcond_df_killed[roadcond_df_killed['Zones'] == 'South Zone']
df =pd.pivot_table(sub_df, index=['Zones'],aggfunc=np.sum).reset_index()
df = df.T.reset_index()
df = df.rename(columns = {'index': 'Road Conditions', 0: 'Total'})
df = df.drop(df.index[0])
df = df.sort_values(by = ['Total'], ascending=False).head(10)
df

fig,ax = plt.subplots(1,1, figsize=(20,10))
sns.barplot(x=df['Road Conditions'],y=df['Total'])
plt.ylabel('# of People Killed')
plt.title('Southern Zone No. of People Killed in Road Accidents Based on Road Conditions', fontsize=15)
plt.xticks(rotation=90)
```

Fig. 4 Road condition analysis

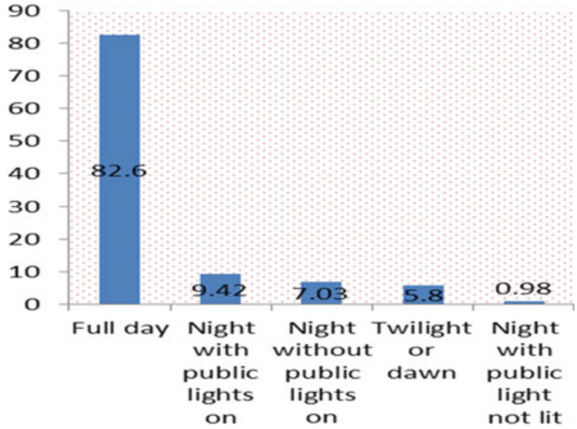


Fig. 5 Accident happened because of lighting

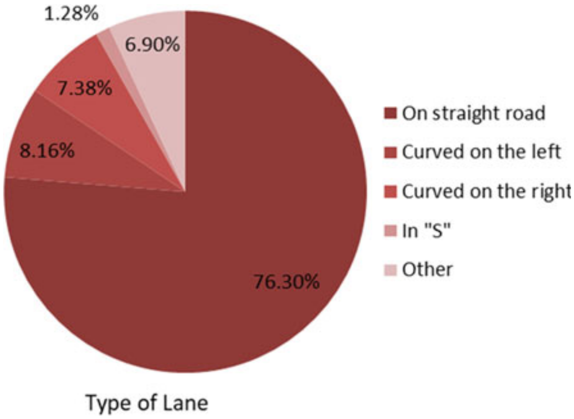
This was followed by a thorough examination of the lighting conditions in the areas where accidents were most common. 82.6% of accidents occur on a daily basis on average, and when public lights are on at night, the average accident rate is 9.42%.

3.3 Accident Lane Analysis

We performed a test on where most of the accident happened and the type of lane based upon few characteristics using linear regression. The graph in Fig. 6 given below shows the analysis that on straight road the accident level is very high and the least accident count is presented on emergency stop band.

In lane analysis, the accident rate is at very rate as 76.30% due to speeding and the drivers misleading transport constraints as shown in Fig. 6.

Fig. 6 Lane analysis

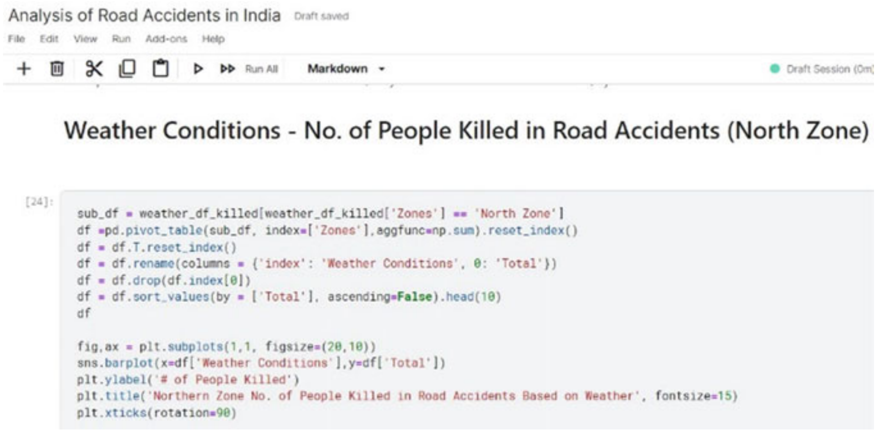


3.4 Analysis on Accident Happened Due to Weather

The fact is that bad weather significantly increases the chance of accident. Conditions such as fog, light rain, cloudy weather, heavy rain, dazzling weather, fog smokes and other weather conditions as in Fig. 7.

The chart given Fig. 8 shows the analysis of accident weather based on few features. The most accident happened on the normal day on an average 81.2%.

Weather conditions contributing 81.2% in normal rate at high and contributes less in other conditions.



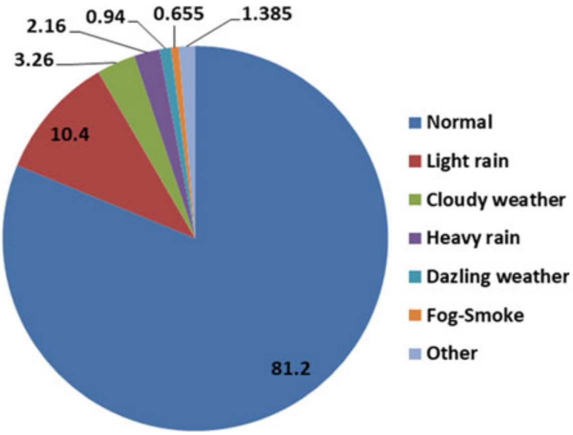
Weather Conditions - No. of People Killed in Road Accidents (North Zone)

[24]:

```
sub_df = weather_df_killed[weather_df_killed['Zones'] == 'North Zone']
df =pd.pivot_table(sub_df, index=['Zones'],aggfunc=np.sum).reset_index()
df = df.T.reset_index()
df = df.rename(columns = {'index': 'Weather Conditions', 0: 'Total'})
df = df.drop(df.index[0])
df = df.sort_values(by = ['Total'], ascending=False).head(10)
df

fig,ax = plt.subplots(1,1, figsize=(20,10))
sns.barplot(x=df['Weather Conditions'],y=df['Total'])
plt.ylabel('# of People Killed')
plt.title('Northern Zone No. of People Killed in Road Accidents Based on Weather', fontsize=15)
plt.xticks(rotation=90)
```

Fig. 7 Weather condition of fatality insight

Fig. 8 Accident weather analysis



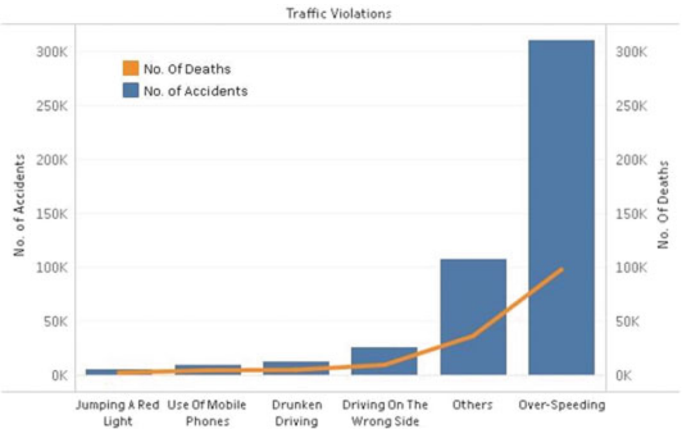


Fig. 9 Accident caused by over speeding

3.5 *Accident Caused by Over Speed*

Over speeding is a major factor in fatal accidents. To excel is part of the human mind as in Fig. 9. As long as man has a chance, he will attain infinite. But if we have to share the road with others, we'll constantly be a few cars behind them. In the event of an accident, speed increases the chance of damage and severity of harm. In addition, quicker cars are more likely to be involved in an accident than slower vehicles, and the severity of the collision will be greater in the case of the faster vehicles [18].

While a slower vehicle will stop instantly, a quicker one will take a long time to stop and skid over a long distance because to the law of idea. Speeding vehicles are more likely to be involved in accidents and cause more injuries. Speeding also impairs the driver's ability to predict what's coming next, resulting in errors in judgment and, collision in the end.

4 Performance Analysis and Result

In this research study to evaluate the performance of the offered techniques we conducted distinct experiments depending on the accident severity class. For accident severity classifications such as vehicle collision, speed, and lightning condition, the performance of each method has been determined. We observe that Decision tree and Random Forest algorithm works faster among other approaches. Table 2 shows the accuracy between the proposed approaches.

Table 2 Machine learning accuracy rate

Algorithm	Accuracy score
A-nearest neighbors	0.934712453515
Logistic regression	0.954389125349
Decision tree	0.968465327445
Random forest	0.9744156832658

5 Conclusion

It is impossible to live with the consequences of traffic accidents in a growing country like us. To decrease the accident rate, in our nation, it has become necessary to regulate and organize road traffic using an unconventional structure. Road accidents can be avoided by taking simple precautions based on forecasting or notifications from a sophisticated system [16]. Moreover, it's an urgent matter for our nation to report the fact that numerous individuals are killed in road accidents every day, and that the number is gradually growing. The use of machine learning is a reliable and effective approach to precise conclusion with the involvement to handle present circumstance, and the findings of the analysis section (Figs. 1, 2, 3, 4, 5 and 6) may be proposed to circulation powers that be for minimizing the sum of coincidences. We may utilise the recommended techniques to employ machine learning here since they have been proved to be more accurate in predicting traffic accident severity. By employing these techniques, we will also aim to create a recommender-system that can anticipate traffic accidents and alert road users. A mobile application based on this approach will be developed in the future to offer precise predictions for the user and make it extremely helpful.

References

1. Labib MF, Rifat AS, Hossain MM, Das AK, Nawrine F (2019) Road accident analysis and prediction of accident severity by using machine learning in Bangladesh. In: 7th international conference on smart computing and communications (ICSCC), pp 1–5. https://doi.org/10.1109/ICSCC.2019.8843640
2. Mamun MAA, Puspo JA, Das AK (2017) An intelligent smartphone based approach using IoT for ensuring safe driving. In: International conference on electrical engineering and computer science (ICECOS), Palembang, pp 217–223
3. Rajkumar AR, Prabhakar S, Priyadharsini AM (2020) Prediction of road accident severity using machine learning algorithm. IJAST 29(06):116–120
4. Habibullah KM, Alam A, Saha S, Amin A, Das AK (2019) A driver-centric carpooling: optimal route-finding model using heuristic multi–objective search. In: 4th international conference on computer and communication systems (ICCCS), Singapore
5. Satu MS, Ahamed S, Hossain F, Akter T, Farid DM (2017)Mining traffic accident data of N5 national highway in Bangladesh employing decision trees. In: IEEE region 10 humanitarian technology conference (R10-HTC), Dhaka, pp 722–725
6. https://www.kaggle.com/phip2014/ml-to-predict-accident-severity-pa-mont

7. Verma V, Bhardwaj S, Singh H (2016) A hybrid K-mean clustering algorithm for prediction analysis. Indian J Sci Technol. https://doi.org/10.17485/ijst/2016/v9i28/98392,pp.0974-6846
8. Ma J, Ding Y, Cheng JCP, Tan Y, Gan VJL, Zhang J (2019) Analyzing the leading causes of traffic fatalities using XGBoost and grid-based analysis: a city management perspective. IEEE Access 7:148059–148072
9. Kumar S, Toshniwal D (2016) A data mining approach to characterize road accident locations. J Mod Transport. https://doi.org/10.1007/s40534-016-0095-5,pp.62-72
10. Princess PJB, Silas S, Rajsingh EB (2020) Machine learning approach for identification of accident severity from accident images using hybrid features. In: International conference for emerging technology (INCET), pp 1–4. https://doi.org/10.1109/INCET49848.2020.9154079
11. Patil J, Prabhu M, Walavalkar D, Lobo VB (2020)Road accident analysis using machine learning. In: IEEE Pune section international conference (PuneCon), pp 108–112.https://doi.org/10.1109/PuneCon50868.2020.9362403
12. Kin T, Goto J, Oshima M (2019) Machine learning approach for gamma-ray spectra identification for radioactivity analysis. In: IEEE nuclear science symposium and medical imaging conference (NSS/MIC), pp 1–2. https://doi.org/10.1109/NSS/MIC42101.2019.9059618
13. Paul J, Jahan Z, Lateef KF, Islam MR, Bakchy SC (2020) Prediction of road accident and severity of Bangladesh applying machine learning techniques. In: IEEE 8th R10 humanitarian technology conference (R10-HTC), pp 1–6. https://doi.org/10.1109/R10-HTC49770.2020.9356987
14. Ting CY, Tan NYZ, Hashim HH, Ho CC, Shabadin A (2020) Malaysian road accident severity: variables and predictive models. In: Alfred R, Lim Y, Haviluddin H, On C (eds) Computational science and technology. Lecture notes in electrical engineering, vol 603. Springer, Singapore, 2020
15. Donchenko D, Sadovnikova N, Parygin D (2020) Prediction of road accidents' severity on Russian roads using machine learning techniques. In: Radionov A, Kravchenko O, Guzeev V, Rozhdestvenskiy Y (eds) Proceedings of the 5th international conference on industrial engineering (ICIE 2019). ICIE 2019. Lecture notes in mechanical engineering. Springer, Cham
16. Gothane S, Sarode MV (2016) Analyzing factors, construction of dataset, estimating importance of factor and generation of association rules for Indian road accident. In: IEEE international conference on advanced computing, pp 15–16. https://doi.org/10.1109/IACC
17. Khaliq KA, Qayyum A, Pannek J (2017) Prototype of automatic accident detection and management in vehicular environment using VANET and IoT. In: 11th international conference on software, knowledge, information management and applications (SKIMA)
18. Al Malki A, Rizk MM, Mousa AA (2016) Hybrid genetic algorithm with K-means for clustering problems. Open J Optim 5:71–78
19. Şahin DÖ, Şirin B, Akleylek S, Kılıç E (2018) Work accident analysis with machine learning techniques. In: 3rd international conference on computer science and engineering (UBMK), pp 215–219. https://doi.org/10.1109/UBMK.2018.8566564

Network Monitoring of Cyber Physical System

Mayank Srivastava, Aman Maurya, Utkarsh Sharma, and Shikha Srivastava

Abstract In the field of network and server security, every organization, no matter how big or small, has to adapt different monitoring techniques in relation with the servers and networks. Server attacks are a major threat in today's world if the systems and networks and the host servers are vulnerable. Network and server monitors or administrators have to keep an eye over different tools for different purposes. Therefore, this paper offers a methodology with the help of which we propose to develop a tool that would ease up the work of server administrators. This tool would provide all the services that multiple other tools provide. This would help enhance the monitoring section of the web servers in particular by using different tools and modules with the help of python. This paper also explains and concludes that the tool being developed would be effective and efficient, and therefore save a lot of time and ensures security from threats related to servers.

Keywords Server monitoring · Network monitoring · CPU · Wireshark · Network mapper (Nmap) · Memory consumption · Scapy

M. Srivastava (✉) · A. Maurya · U. Sharma
Department of Computer Engineering and Applications, GLA University, Mathura, UP 281406, India
e-mail: mayank.srivastava@gla.ac.in

A. Maurya
e-mail: aman.maurya_csf18@gla.ac.in

U. Sharma
e-mail: utkarsh.sharma_csf18@gla.ac.in

S. Srivastava
Department of Mathematics, Aligarh Muslim University, Aligarh, UP 202002, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_51

1 Introduction

Cyber Security Incidents—including Sophisticated Cyber Security Attacks—can and do occur in many different ways. When we talk about risk to an organization due to cyber-attack, chances are, the organization can lose a significant amount of data in no time, affecting the speed and performance and majorly, the reputation of that particular organization. Therefore, to deal with such suspected or actual cyber security attacks or threats to become incidents, we need to record the events related to it, monitor them in a refined way on continual basis so as to investigate a suspected cyber security breach thereby, reducing the chances of a loss of data, information, etc. It will also remediate the damage easily.

In this paper, an approach is presented in such a manner that it would make use of network packet sniffing, tracking, analyzing and extracting information, an easy task to perform. We have clustered modules such as wireshark, nmap, etc. with the help of python by creating a program that would run these modules in the background, which can be monitored effectively and efficiently. Section 2 includes related works of other researchers and authors in this field with their valuable contribution. Section 3 deals with the theoretical aspects of the background study of this paper and related topic's discussion. Section 4 explains different modules and commands used in this work and their demonstration. Section 5 is dedicated to methodology used in this paper and Sect. 6 is dedicated to represent the obtained results. Section 7 gives the importance of the proposed approach. The conclusion is given in Sect. 8.

2 Review of Literature

The area of Cyber Physical System is one of the ongoing technological research domain in which good number of researchers are working. Some of the major contribution of this area are as follows.

Zeng and Wang [1] introduces how to monitor numerous servers via a very unpretentious protocol called SNMP. They used multi-threading technological aspect for the collection and processing of data thereby improving the efficiency of the collection. Roblee and Berk [2] developed a system named PQS which enables user-space monitoring of servers and makes accurate and fast decisions regarding server and service state. Yucheng and Yubin [3] focuses mainly on real-time monitoring and as it is very difficult to understand the software performance related issues by historical data analysis and remote monitoring. Forrest et al. [4, 5] showcases an initial result which focus at creating a different definition for UNIX processes which elaborates synonymous treatment along with normal behavior for itself. Bohra et al. [6] proposed a remote monitoring technique and recovery of the state of the software with taking no help from the processors and OS resources of the of a computer system. Kephart [7] proposed an autonomic computing that has a grand-challenge revelation of the future in which calculating systems will accomplish themselves according to the high-level purposes quantified by humans.

Tsoa et al. [8] proposes an extensive survey on the management of server and network resources over Cloud infrastructures by highlighting key concepts and critically discussing their limitations and implications. Thirukonda and Becker [9] gives an automated architecture named WebSpy used to notify and take appropriate action when server downtime is identified. Sihyung et al. [10] shows the study of present network-monitoring technologies and identify different problems and suggested future directions. Suri and Batra [11] focusses on comparative study of different packet analyzers and criteria for choosing among them. Trimintzios et al. [12] presents the design and implementation of an API named DiMAPI used for managing flow creation and manipulation over distributed passive network monitoring. Fang et al. [13] proposed a method based on IA to complete dynamic network monitoring. The proposed method is developed using cross-platform language and includes automatic and manual mode. Bonelli and Giordano [14] presents Linux based approach named as PFQ that primarily fine-grained distribution to network applications and physical devices. Bashar and Smys [15] introduces an analysis of eavesdropping security issue in WSN environment. The paper proposed the notion of decreasing probability of interception and secure connection between the nodes. Haoxiang and Smys [16] discusses a concept based on optimization of energy related to Cyber-physical systems along with memory aware scheduling strategy and algorithms.

3 Basic Notions

This section presents the various basic notions related to the concepts discussed in this paper.

3.1 Networking

Networking, often called “computer networking”, is a well-known term in the field of computer science and also in cyber security domain. Usually understood as a practice, it involves moving and switching data between numerous nodes over a public medium of a statistics system. Network comprises mainly design, construction, operation, maintenance and management of the network.

The level of expertise vital to operate a network, associates directly to the intricacy of that particular network which is under supervision. Computer networking permits all the devices and their endpoints to each other on a LAN (Local Area Network) or WAN (Wide Area Network)-for larger network. Networking includes everything such as telephone calls, texts, streaming of videos, internet and also, internet of things (IoT).

3.1.1 Types of Networking

- **Wired Networking**: It necessitates the usage of a physical mode of transportation of data between two or more nodes. For example, copper Ethernet cables, optical fiber, etc.
- **Wireless Networking**: It makes the use of radio waves and frequencies to transport data by air, empowering devices to be linked to a network without any cables. For example, microwave transmission, satellite, cellular and Bluetooth, wireless LAN, etc.

3.1.2 Some Basic Concepts

- **TCP SYN Request**: It is a form of denial-of-service attack in which an attacker hastily recruits a connection to a server without concluding the connection. The server has to spend resources to come up for half-opened connections, which can consume enough resources to make the system impassive to authentic traffic.
- **TCP Reset Connections**: It represents an unforeseen closing of the session. It causes the resources assigned to the connection to be instantly released and all other information about the connection is removed. TCP reset is symbolized by the RESET flag in the TCP header set to 1.
- **TCP Half-Open Connections**: The term half-open refers to TCP connections whose state is out of harmonization between the two interactive hosts, possibly due to a clatter of one side. A connection in the process of getting launched is also known as embryonic connection.
- **HTTP GET Requests**: HTTP GET request method is used to regain data from a stated URL. The GET is the most popular HTTP request technique. GET requests should only collect data and should not disturb the state of the server.

3.2 Networking

System performance monitoring is a method of collecting and analyzing the performance parameters of a system such as memory usage, I/O, CPU usage summary of node, etc. It involves presenting the data in such a way that it can be easily and effectively administered and understood by an administrator. Such services are very important and critical for the stable working of large clusters as it gives a green light to the administrator to investigate and find possible problems well before damage occurs. Also, other system software parts are benefitted with this information provided by the administrator after monitoring.

3.2.1 Server Monitoring

In our daily lives, we have witnessed that computer network and size of communication has increased in a rapid fashion. As with this kind of pace, network has become a very important and irreplaceable asset for all of us. Therefore, network monitoring is as essential, which is evident from the fact that every organization now focuses on monitoring its network. Majority of all the network management software are mainly focused on the particular link and network equipment only, and not on server. As it appears, servers should be taken into consideration much like the networks or more. Therefore, this paper gives a brief explanation on how to monitor not only the networks and links, but also the servers and information related to it by developing a systematic monitoring tool. Main objectives to monitor server performance are as follows:

- Server availability can be monitored.
- Responsiveness of the server is easily monitored.
- Server capacity and speed are also monitored effectively.
- Detection and prevention of issues that might affect the server proactively.

3.2.2 Network Monitoring

Network monitoring involves discovering, mapping and monitoring a computer network in a quest to ensure optimal performance and availability. Therefore, the only way to learn about what's on a network is by monitoring it with the help of various network monitoring tools. The reason why we monitor a network is because it is the lifeline of IT infrastructure and almost all the critical information floats over some network. If there is any problem arising on a network due to a possible attack or malfunctioning, that critical information comes under a huge risk. A network monitoring system usually performs the following basic functions:

- Discovering
- Mapping
- Monitoring
- Alert
- Reporting.

Network administrators ensures everything is working in its supposed manner so that there may be no malfunctioning or threat for a disrupt. Routers, switches, ports, etc. usually comes under the scrutiny in case of any disruption. Hence, Administrator processes a report based on what is wrong and a critical action is then taken.

4 System Modules

In the process of developing the desired web server monitoring tool, we have studied many tools and techniques which are being used to monitor different aspects of a server and a network. Such tools, or we can say modules, includes wireshark, nmap, netstat command, scapy, socket, psutil, shutil and xampp stack package.

Let us understand the basic functionalities of these modules and how we have compiled them based on their functionalities.

Wireshark: It is an open-source packet analyzer, used for many purposes in the fields of communication protocol, network troubleshooting, software development, etc. to track the network packets in a manner as per need. Most commonly, Wireshark is considered as a sniffer tool or a network protocol analyzer. Wireshark also monitors unicast circulation which is directly not transmitted to the network's MAC address interface, but it does various network taps which is also known as port mirroring to extend the capture at any point of time or traffic. The use of Wireshark in the implemented tool is shown in Fig. 1.

How Wireshark command has been used in this tool

- Wireshark captures packets, analyses them and displays every information related to it.

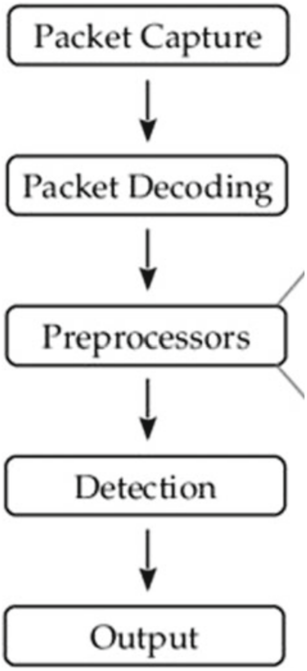


Fig. 1 Use of Wireshark in the tool

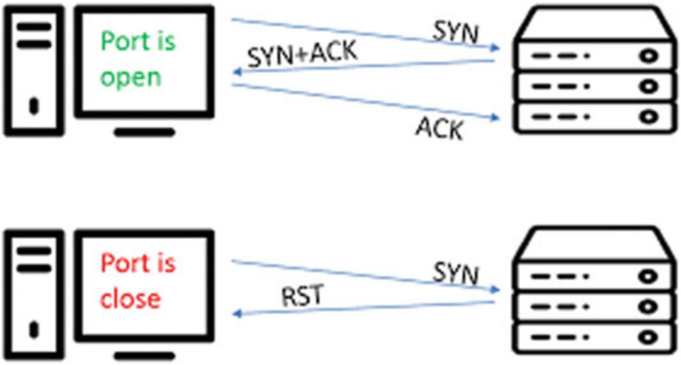


Fig. 2 Use of Nmap in the tool

- This information is stored in a.pkt file which is stored in the systems library with a certain number of packets and with different IP addresses as well.
- We make use of these files to extract information about the running traffic on a network.

Nmap: It is also an open-source utility for discovering a network. A Network Mapper (Nmap) is used to perform security auditing and scanning of a network. Its main use is to manage network inventory, monitor hosts and server uptime and downtime. We can use it over windows or linux or even Mac. Saving and comparing scan results is one of the main features of this tool. The use of Nmap in the tool is shown in Fig. 2.

How nmap has been used in this tool

- Nmap has a predefined task to map the network traffic by providing information regarding all the ports and services.
- With the help of python and nmapscanner, we have collaborated nmap and our tool effectively to display all the information about the ports and services they offer.

Netstat Command: Netstat Command is used in linux operating system which stands for "Network Statistics". It provides information on different interface statistics, which includes open sockets, routing tables and connection information as well. Netstat command is also considered handy when used to display all the socket connections like TCP, UDP. It displays all those packets which are pending to get connections.

How netstat command has been used in this tool

- Netstat command is a replacement to many other statistics tools.
- With just one command, we can capture live port analysis and information about the network we are operating at.

Scapy: Scapy is a communicating packet handling program which is proficient for deciphering the packets of a wider quantity of protocols, while it sends them on the wire and captures them, makes ample number of requests and so on. The performance of scapy is commendable in its true sense for handling classical tasks like scanning, tracerouting, probing, unit tests, etc. Other specific tasks include invalid frames transfer.

How scapy has been used in this tool

- Scapy is a library in python which captures network packet. Therefore, the packets from wireshark that were first obtained are extracted using scapy into python.
- We have run scapy using python on linux, but it can also run over windows, OSX and on most Unixes using libpcap.

Socket: Sockets are basically the endpoints of a two-directional communication channel which communicates amongst a process or between two or more processes on the same system or between processes over different machines on different locations as well. We can implement sockets over different channel types, for example, Unix domain, TCP, UDP, etc. Its collection efficiently provides with precise classes for handling some public transports and generic interface to handle the rest others.

Psutil: Also known as "Process and System Utilities", Psutil is basically a cross-platform library used for retrieving information about running processes in python as well as system utilization like CPU, Memory, etc. The main use of Psutil is system monitoring, profiling and limiting process resources as well as management of running processes.

How psutil has been used in the tool

- Psutil utility modules has been integrated with python in the development of this tool in order to extract the information regarding the CPU usage.
- CPU usage such as RAM usage, memory usage, etc. is extracted using **pyautogui** library.

Shutil: Shutil is used to automate copying files and directories. Shutil helps in saving a lot of steps of opening, reading and writing and also closing of files, if there exist a condition when no actual processing is being done. This utility module is also used to achieve tasks like copying, moving or removing directory trees.

How shutil has been used in the tool

- Shutil has been used to extract information about the disk usage of the system being used for monitoring.
- As we provide path of directory where python files are saved to shutil utility module, it extracts the disk usage and displays it.

Xampp: Xampp is an open-source cross-platform web server solution stack package which makes transitioning from a local test server to a live server. It offers an apache application server, mariaDb database server, a php server, etc. and makes it easier to host. We used Xampp to host a dummy website's application server and then ran all the monitoring tools with the help of a python concentric technique to get the desired outputs as part of our work.

How Xampp has been used in our tool

- Xampp offers many services like web server hosting, database server hosting, etc.

We have used Xampp to host a web server that is to be monitored for various purposes.

5 Proposed Methodology

Organizations bank on numerous servers such as core App servers, Database servers, Web servers, and caching servers for regular communiqué and serious business maneuvers. Servers are, undoubtedly, one of the most critical elements of the IT infrastructure as they're largely used to bring about resources in a network. Therefore, we created a tool using python which monitors the performance of server and helps organization to proactively detect issues at early stages.

We basically merged a few packet sniffing and tracing tools such as Wireshark, Nmap, etc. which gives the desired results as per our needs. To help matters, we constructed a python code to trace all the TCP packets using Wireshark. This helped us in getting information about what all packets are live on the server. Using Nmap, we tracked down the ports that we open on the server, in order to get information about the incoming and outgoing packets through those open ports. Secondly, server resource utilization is also observed using psutil and shutil modules with the help of python. It gives information regarding the CPU usage and RAM usage and also gives information about the disk usage statistics. Also, server uptime and downtime information are also gathered in the same manner. The basic process of implemented tool in terms of Use-case is shown in Fig. 3.

In the build-up to this tool's GUI, following are the key features of this tool:

Features

- It has four different buttons with their specified functionalities.
- System-Server Information section provides all the information regarding the system and server.
- Port status section gives a comprehensive information regarding all the ports open and closed on the server.
- Packet analysis section delivers a detailed information regarding the packets transmitted, types of packets, flags used, etc. It also gives a count of total packets transmitted and total data transmitted in bytes.
- Live Packet Analysis section captures live packet transmission.
- Server Bandwidth section displays downloading and uploading speed on the server.

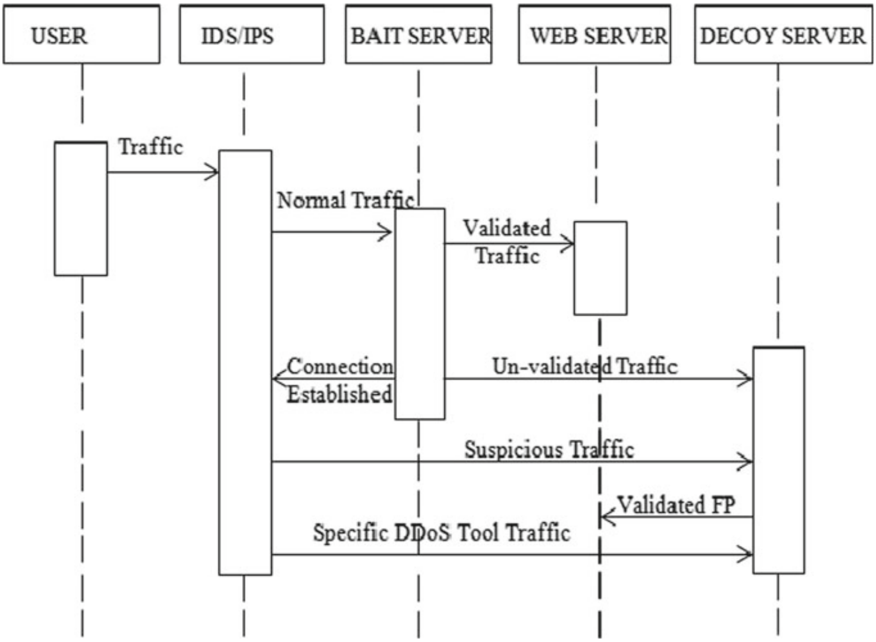


Fig. 3 Use case diagram

6 Simulated Results

The GUI of the tool—Web Server Monitoring is given in Fig. 4. Using the tool, we can monitor the server location, information of system and its architecture, bandwidth, uptime, CPU and RAM usage with disk usage and the server port status. The IP address of the server used—139.167.198.38.

Monitoring Status

"Connection Acquired"

Connection acquired at: 2022-03-28 18:44:09
Monitoring started at: 2022-03-28 18:44:09
Disconnected at: 2022-03-28 18:44:017.

System-Server Information Statistics

Server Information: Location: 28.6519,77.2315, Region: Delhi, City: Delhi, Country: IN.

Server System Information: System: Windows, Node Name: Utkarsh-Sharma, Release: 10, Version: 10.0.19044, Machine: AMD64.

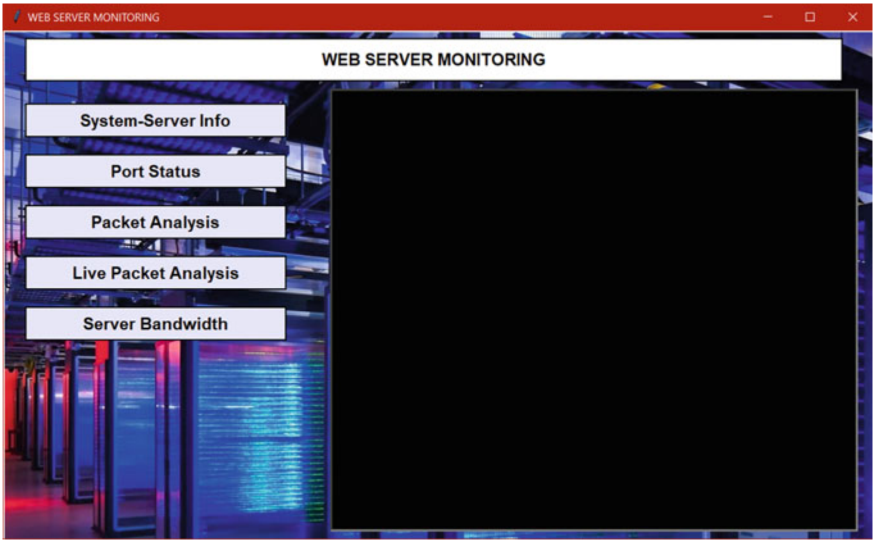


Fig. 4 GUI of the tool

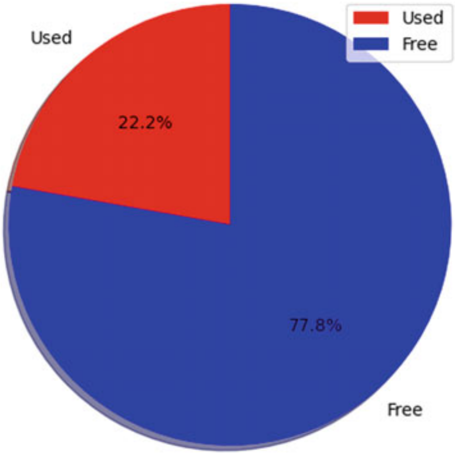
Disk Usage Statistics: Disk usage statistics: usage (total = 250,025,603,072, used = 134,773,084,160, free = 115,252,518,912). The statistics is also shown in Fig. 5.

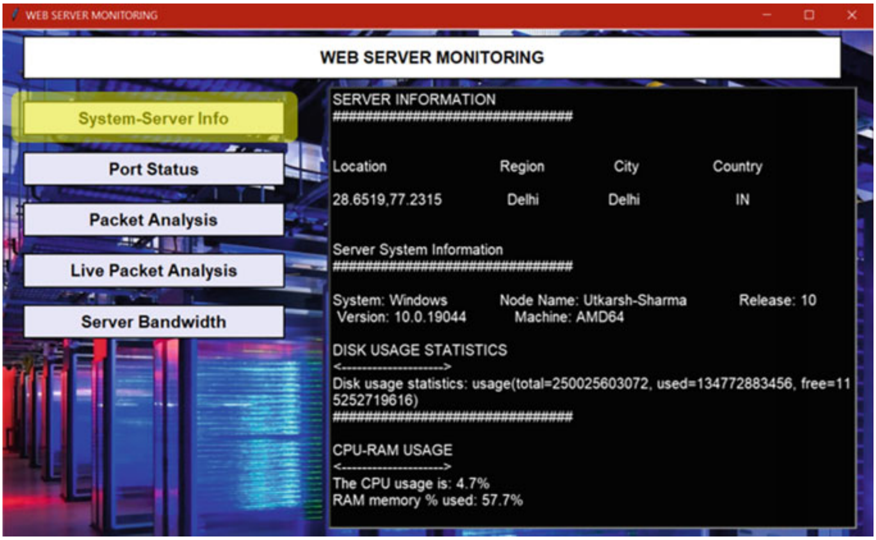
CPU-RAM Usage: The CPU usage is 12.4% and RAM memory % used 55.6%
The snapshot of the above statistics is also shown in Fig. 6.

Server Port Status

The port status of the Server is shown in Table 1. The snapshot of the tool which gives Server port status is given in Fig. 7.

Fig. 5 Disk usage statistics





```
SERVER INFORMATION
###################################

Location              Region       City        Country
28.6519,77.2315        Delhi       Delhi          IN

Server System Information
###################################

System: Windows        Node Name: Utkarsh-Sharma        Release: 10
 Version: 10.0.19044     Machine: AMD64

DISK USAGE STATISTICS
<-------------------->
Disk usage statistics: usage(total=250025603072, used=134772883456, free=11
5252719616)
###################################

CPU-RAM USAGE
<-------------------->
The CPU usage is: 4.7%
RAM memory % used: 57.7%
```

Fig. 6 Server system ınformation

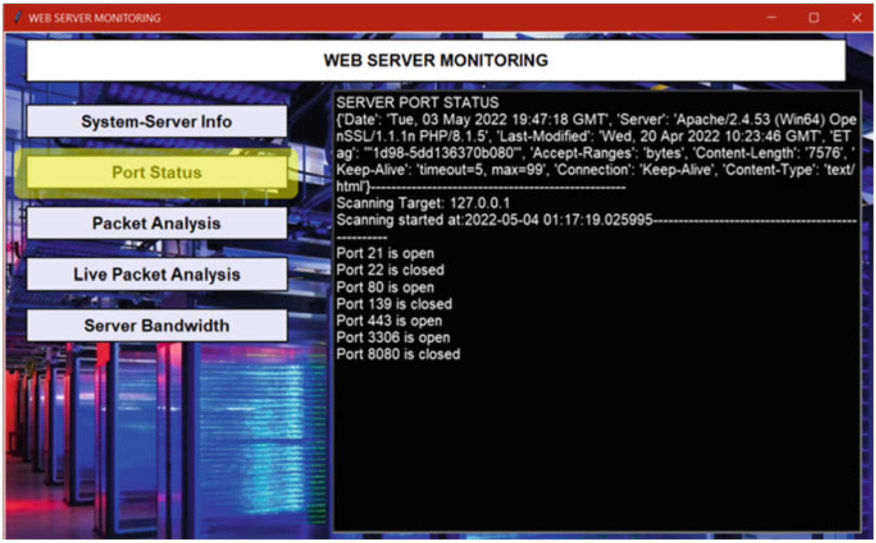
Table 1 Port status of server

Port	Status	Service
Port 21	Open	FTP
Port 22	Closed	NSH
Port 80	Open	HTTP
Port 139	Closed	Netbios-SSN
Port 443	Open	HTTPS
Port 8080	Closed	HTTP-alt

Packet Analysis

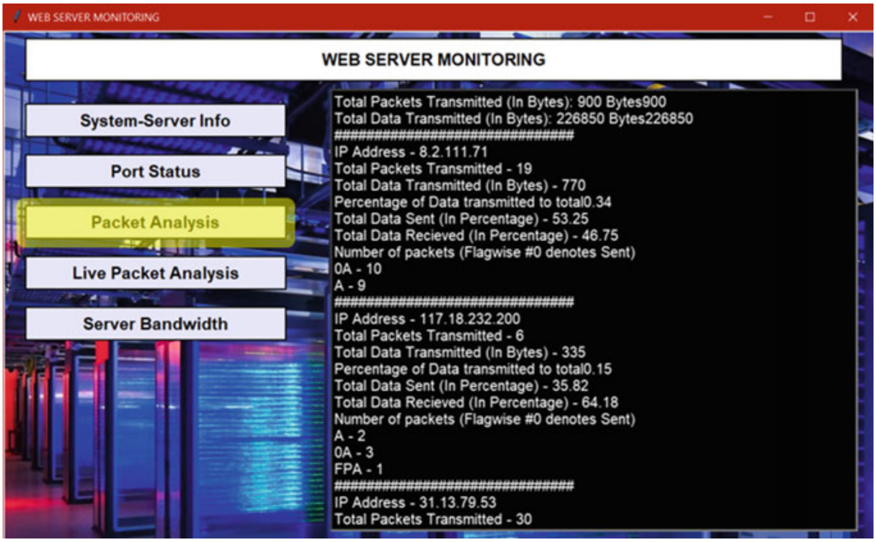
The Server packet analysis is given below. The snapshot of this statistics is also given in Fig. 8.

Total Packets Transmitted (In Bytes) 900
Total Data Transmitted (In Bytes) 226,850
IP Address—8.2.111.71
Total Packets Transmitted—19
Total Data Transmitted (In Bytes)—770
Percentage of Data transmitted to total 0.34
Total Data Sent (In Percentage)—53.25
Total Data Received (In Percentage)—46.75
Number of packets (Flag-wise #0 denotes Sent)
0A–10



```
SERVER PORT STATUS
{'Date': 'Tue, 03 May 2022 19:47:18 GMT', 'Server': 'Apache/2.4.53 (Win64) Ope
nSSL/1.1.1n PHP/8.1.5', 'Last-Modified': 'Wed, 20 Apr 2022 10:23:46 GMT', 'ET
ag': '"1d98-5dd136370b080"', 'Accept-Ranges': 'bytes', 'Content-Length': '7576', '
Keep-Alive': 'timeout=5, max=99', 'Connection': 'Keep-Alive', 'Content-Type': 'text/
html'}----------------------------------------
Scanning Target: 127.0.0.1
Scanning started at:2022-05-04 01:17:19.025995----------------------------------------
----------
Port 21 is open
Port 22 is closed
Port 80 is open
Port 139 is closed
Port 443 is open
Port 3306 is open
Port 8080 is closed
```

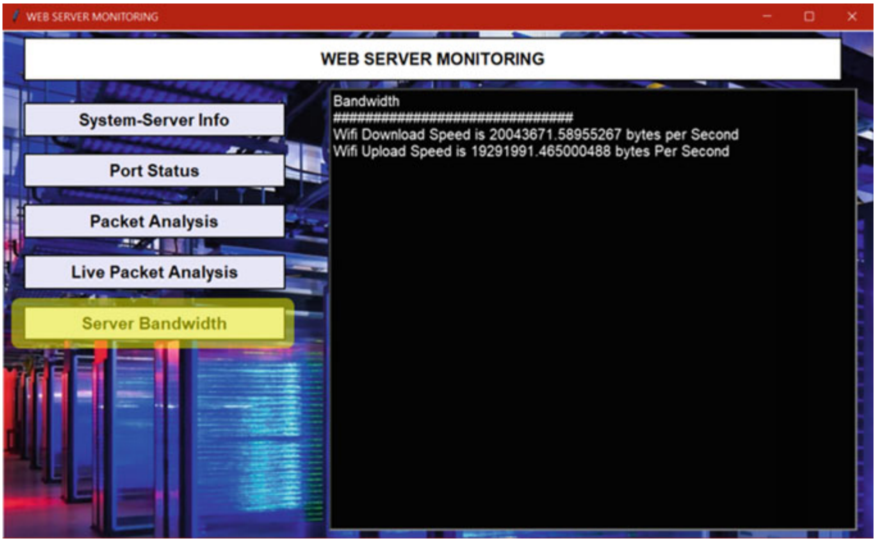
Fig. 7 Server port status



```
Total Packets Transmitted (In Bytes): 900 Bytes900
Total Data Transmitted (In Bytes): 226850 Bytes226850
#################################
IP Address - 8.2.111.71
Total Packets Transmitted - 19
Total Data Transmitted (In Bytes) - 770
Percentage of Data transmitted to total0.34
Total Data Sent (In Percentage) - 53.25
Total Data Recieved (In Percentage) - 46.75
Number of packets (Flagwise #0 denotes Sent)
0A - 10
A - 9
#################################
IP Address - 117.18.232.200
Total Packets Transmitted - 6
Total Data Transmitted (In Bytes) - 335
Percentage of Data transmitted to total0.15
Total Data Sent (In Percentage) - 35.82
Total Data Recieved (In Percentage) - 64.18
Number of packets (Flagwise #0 denotes Sent)
A - 2
0A - 3
FPA - 1
#################################
IP Address - 31.13.79.53
Total Packets Transmitted - 30
```

Fig. 8 Server packet analysis

A–9.



```
Bandwidth
########################################
Wifi Download Speed is 20043671.58955267 bytes per Second
Wifi Upload Speed is 19291991.465000488 bytes Per Second
```

Fig. 9 Server bandwidth status

Server Bandwidth

The Server bandwidth status is given below. These statistics are also given in terms of snapshot in Fig. 9.

Wifi Download Speed is **21734586.81508718** bytes per Second
Wifi Upload Speed is **3065543.6396214203** bytes Per Second

Therefore, this shows how the tool works where we don't have to look around for multiple tools but a single platform is sufficient enough.

7 Importance of the Monitoring Tool

Web Server Monitoring techniques represents the use of techniques and tools for security reasons. If you want to check what ports are open or closed, what services these ports are providing (in case you want to assess your server), then you have to use nmap for network traffic mapping. Similarly, if you want to go check how many data packets and of what type are travelling to and fro on your server, then you look for wireshark and so on. Therefore, this constant use of different tools and techniques takes a lot of time and complications. This paper comes up with a tool which is developed to give an efficient solution in this regard.

Key important features of this project and monitoring tool are:

1. There will be no need of playing around with different tools and techniques.

2. Web server monitoring becomes easier and efficient.
3. With efficiency, time taken to check and analyze a web server is reduced.
4. User friendly interface that provides easy and quick results.
5. One domain and multiple services.

However, there are certain parameters that are still need to be focussed more precisely like data loss, throughput, efficiency etc. These parameters will be added as a part of future work of this paper.

8 Conclusion

This paper gives a concrete solution as how we can manage different tools, techniques and modules in order to combine them significantly in a manner that every possible information about the server can be extracted easily and efficiently. With the help of python, we have achieved the objective of getting maximum information about the hosted server. Therefore, our objective is somewhat to develop a tool which can provide us with different information regarding the web server which helps ease up the monitoring when it comes to web server security. A significant action can be taken as all the information regarding the server is available on a single platform, rather than exploring different tools for different information. Further, the scope of these types of tools is huge. Time is money in today's time and with increasing digitalization, we need to be very specific as to what need to be done to reduce the downtime. In this regard, this paper proposes efficient tool for server and network monitoring for reducing downtime.

References

1. Zeng W, Wang Y (2009) Design and implementation of server monitoring system based on SNMP. In: IEEE international joint conference on artificial intelligence, vol 3, pp 857–860
2. Roblee C, Berk V (2005) Implementing large-scale autonomic server monitoring using process query systems. In: IEEE proceedings of the second international conference on autonomic computing, pp 123–133
3. Yucheng L, Yubin L (2010) A monitoring system design program based on B/S mode. In: IEEE international conference on intelligent computation technology and automation, pp 184–187
4. Forrest S, Hoffmeyr S, Somayaji A, Longstaff T (1996) A sense of self for unix processes. In: IEEE symposium on security and privacy, pp 120–128
5. Nemeth E, Snyder G, Seebass S, Hein TR (1995) UNIX system administration handbook. Prentice Hall
6. Bohra A, Neantiu I, Gallard P, Sultan E, Iftode L (2004) Remote repair of operating system state using Backdoor. In: IEEE proceeding of the international conference on autonomic computing, pp 256–263
7. Kephart JO (2005) Research challenges of autonomic computing. In: Proceedings of the 27th international conference on Software engineering, IBM Thomas J. Watson Research Center, USA

8. Tsoa FP, Jouet S, Pezaros DP (2016) Network and server resource management strategies for data centre infrastructures: a survey. Comput Netw 106:209–225
9. Thirukonda MM, Becker SA (2002) WebSpy: an architecture for monitoring web server availability in a multi-platform environment. Technical report CS-2002-07, Computer Science Department, Florida Institute of Technology
10. Sihyung L, Kyriaki L, Hyong SK (2014) Network monitoring: present and future. Comput Netw 65:84–98
11. Suri S, Batra V (2010) Comparative study of network monitoring tools. Int J Innovative Technol Exploring Eng (IJITEE) 1(3):63–65
12. Trimintzios P, Polychronakis M, Papadogiannakis A, Foukarakis M, Markatos EP, Oslebo A (2006) DiMAPI: an application programming interface for distributed network monitoring. In: Conference on network operations and management symposium, pp 382–439
13. Fang W, Zhijin Z, Xueyi Y (2008) A new dynamic network monitoring based on IA. In: IEEE international symposium on computer science and computational technology. IEEE, pp 637–640
14. Bonelli N, Giordano S (2016) Network traffic processing with PFQ. IEEE J Sel Areas Commun 34(6):1819–1833
15. Bashar A, Smys S (2021) Physical layer protection against sensor eavesdropper channels in wireless sensor networks. IRO J Sustain Wirel Syst 3(2):59–67
16. Haoxiang W, Smys S (2020) Secure and optimized cloud-based cyber-physical systems with memory-aware scheduling scheme. J Trends Comput Sci Smart Technol (TCSST) 2(3):141–147

Impact of Security Attacks on Congestion in Wireless Sensor Networks

Divya Pandey and Vandana Kushwaha

Abstract Security is an essential factor that must be addressed in wireless sensor networks (WSNs) since they are resource-constraint, may handle sensitive data, and operate in hostile inaccessible regions. The Purpose of security attacks in networks is mainly to breach the confidentiality, authenticity and, integrity of data which consequently raises the energy consumption rate and lessens the network lifetime. Several methods and techniques have been applied by attackers to achieve their goals. However, one of the easiest methods is to prompt congestion which causes packet delay, packet loss, re-transmissions and ultimately jeopardizes the network. Unfortunately, aggravation of network congestion by the faulty behaviour of malicious nodes has not been discussed to great extent in literature. In this paper, we explore and discuss different types of attacks to analyse their impact on congestion occurrence and related problems through simulation in different network scenarios. According to this analysis, different types of attacks could have devastating effects on networks. This demonstrates the need for developers to create more secure WSNs.

Keywords Congestion · Security-attacks · Wireless sensor networks · Malicious nodes · Network lifetime

1 Introduction

Researchers are paying a lot of attention to wireless sensor networks (WSNs) [1] due to their vast applications. Network security is one of the necessary requirements of these applications. However, WSNs are susceptible to various types of security attacks due to its stringent constraints and features [2]. Based on the target

D. Pandey (✉) · V. Kushwaha
Department of Computer Science, Banaras Hindu University, Varanasi, India
e-mail: divya.pandey4@bhu.ac.in

V. Kushwaha
e-mail: vandanakus@bhu.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_52

of attackers, security attacks in the network have been divided into two broad categories—Active and Passive attacks. An active attack disrupts the WSN by injecting defective data into it, impersonating, modifying resource and data streams, breaking security protocols, destroying sensor nodes, introducing bugs into the protocol, and overloading it with traffic. Rather than intercepting data communications, passive attacks focus on privacy (eavesdropping on data communications or monitoring packets within a WSN). Both types of attacks have adversarial effects on networks. As a result, security is a critical concern in WSN in order to maintain the confidentiality, availability, and integrity of sensed and sent data. Attacks on networks are primarily intended to deny the quality of service. Denying services in WSNs can be accomplished in several ways, including destroying packets and signals, congesting the network, dropping packets, or depleting the energy of network resources.

Many congestion control algorithms for Wireless Sensor Networks ignore the impact of malicious nodes in network congestion as they presume that all nodes are authentic and operate properly, which is not a feasible assumption [3]. There can be various internal and external attacks in the network which might interrupt the usual operation of the network by compromising the transmitted data packets. Sensor nodes are prone to failure and malicious nodes aggravate congestion by sending multiple fake messages or dropping packets unnecessarily. Moreover, they can even generate heaps of packets abruptly. A network's security can be improved by identifying the most harmful attacks that it can face. Keeping this fact in mind, in this paper, we explore and discuss different types of attacks that can cause congestion. Classification of these attacks has been done based on how they create congestion as shown in Fig. 1. Effects of security attacks on important network parameters such as PDR, throughput and network lifetime have been shown through simulation. In this way, theoretical as well as experimental aspects of security attacks are presented in this paper.

2 Related Works

Security is one of the prime concerns due to its catastrophic consequences. As a result, it is one of the major research topics in the field of wireless sensor networks. There have been several attempts by researchers to explain what are the different kinds of attacks, as well as their detection techniques and countermeasures. A review of some of the relevant existing research articles in which security attacks in WSNs have been discussed in detail can be found below.

Dewal et al. [4] have presented a series of challenges, security issues, and breaches in wireless sensor networks along with defensive measures that can be taken in response to these threats.

Abasikeleş-Turgut et al. [5] have investigated the effects of blackhole and sinkhole attacks on clustered WSNs using various performance metrics.

Gavric and Simic [6] have discussed several common DOS attacks as well as potential prevention methods. They observed that DOS attacks can be divided into

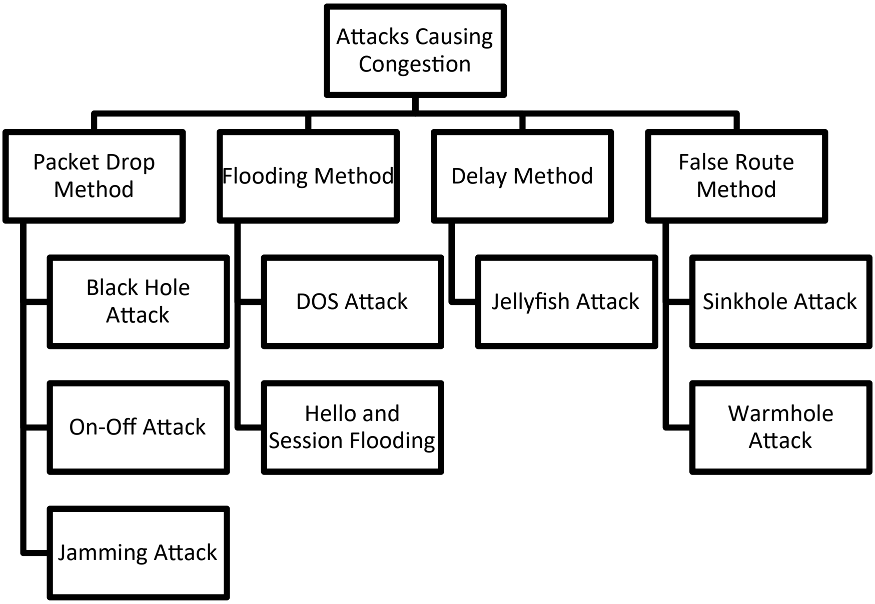


Fig. 1 Attacks causing congestion

multiple categories and they categorized DoS attacks based on the layers of protocol stacks.

Diaz and Sanchez [7] have presented a methodology for security analysis in WSNs that involves attacker modelling and attack simulation, as well as performance analysis (estimation of node's software execution time and power consumption). Following an examination of several WSN attack variants, an attacker model is developed. This model specifies three categories of attackers capable of simulating most WSN assaults.

Tripathi et al. [8] have investigated the performance of the leach protocol in WSN when it was compromised by a black hole and a grey hole assault. They thoroughly explored different network characteristics with different node density. It has been noticed that the black hole attack has a greater impact on network performance than the grey hole attack.

Suma [9] has analyzed the impact of the sybil attack on localization error in wsn and proposed a countermeasure by employing a detection and defence technique based on distance vector hop and the simulation result showed that proposed technique is able to decrease the average localization error buy a solid 4% when compared with previous methodologies.

Vivekanadam [10] has presented a hybrid technique of hopfield neural network and firefly algorithm employing leach for preventing WSN from denial of sleep attack (DoSA) in which there is a significant loss of energy at the nodes due to their inability to enter sleep or power conservation modes. To address this issue, he proposed a hybrid approach which results in a large improvement in network lifespan and energy usage patterns.

Ganguly et al. [11] have proposed a novel Trust Integrated Congestion-aware Energy Efficient Routing algorithm (TCEER). They have calculated node potential using a Fuzzy Logic Controller based on its trust value, congestion condition, the remaining energy of the node, and distance from the current packet-transmitting node and the base station using a Fuzzy Logic Controller. In this way, they have combined security and congestion aspects together in their approach. The simulation results showed that security and congestion, when combined, yield positive network performance outcomes.

Yan and Qi [12] a introduced congestion-aware routing algorithm (CARA). The method evaluates four route parameters: forward rate, node load factor, cache remaining rate, and forward average cache remaining rate, taking into account both geographical relationship and traffic load. Routing decisions are made using the multi-parameter fusion approach. As a consequence, the CARA algorithm recognises the sensor nodes and the surrounding area's congestion perceptions and optimises network transmission performance.

It is evident from the literature that although researchers have made significant efforts to provide security to WSNs against various kinds of attacks. Adversarial impact of these attacks on different network parameters have also been studied. However, congestion and security both aspects have not been given much consideration together, although they are somewhat inversely proportional. Therefore, in this paper efforts have been made to demonstrate the correlation between security attacks and congestion and it is observed that the impact of attacks on congestion is severe.

3 Attacks that Contribute to Congestion Occurrence

Some internal attacks can cause congestion on sensor nodes, links or on both. There are variety of methods for attackers to make network congested. Therefore, depending on the approach of creating congestion these attacks have been classified into 4 major categories. A short description of these methods has been given in the following subsection.

3.1 Packet Drop Method

The idea behind this type of attack is basically to destroy the packets or signal so, as to prevent them from reaching to the destination so that they must be retransmitted resulting in a huge number of packet drops and retransmissions. In this way, this method hampers QoS parameters such as packet delivery ratio, network throughput and end-to-end delay etc. Attacks that follow packet drop pattern are as follows

- **Black hole Attack**

  Blackhole attack tries to divert network traffic towards the malicious node by advertising zero or low cost towards the Sink node. Basically, it tricks neighbouring nodes into sending packets to the compromised node, where another attack takes place. The blackhole attack is usually combined with **Selective Forward attack** which discards some or all the packets it receives by neighbour nodes [9]. In worst case, a malicious node imitates a black hole, discarding all data packets that travel through it as matter and energy vanish from our cosmos. If the attacking node is a connecting node between two network connecting components, the network is essentially divided into two separate components.
- **On–Off Attack**

  In On–off attack as its name suggests, malicious nodes might opportunistically act good or bad. Nodes in the network are compromised in such a way that bad behaviour would go unnoticed [10]. As a result, malevolent nodes can maintain their trust despite behaving negatively. So, this type of attack can only be identified based on the analysis of misbehaviour history.
- **Jamming Attack**

  The purpose of jamming attack is to destroy a signal or a packet. The term "destroy" here refers to rearranging a signal or modifying a few bits of a packet by interfering with transmission [5]. Jammers are one of the earliest and most well-known WSN attacks. Jamming can occur at any OSI layer of a network.

3.2 Flooding Method

In flooding technique attackers make lots of replica of packets and overwhelm the sensor node with huge number of incoming packets making it incapable of processing them further so it become congested. A list of flooding based attacks are as follows

- **Denial of Service (DoS) Attack**

  A DOS attack seeks to interfere with the proper operation of a network.in DoS attack compromised node makes lots of replica of received packets and transmit it to its neighbour node. And the receiver node becomes unable to handle this multitudinous packet simultaneously so it gets congested [6]. Thus, DoS assault affects a target by flooding it with massive volumes of packets thereby degrading or rendering the target's service useless.
- **Hello and Session Flooding**

  The motive of this type of attack is to exhaust the node's resources or drain its energy. Channels are continuously flooded with requests or transmissions from malicious nodes. Using a large transmission power, a network attacker could broadcast "HELLO" packets (used in many protocols for node discovery) in order to fool all nodes into thinking the adversary is within one-hop communication range, thus wasting excessive energy sending packets to an imaginary neighbour. Sensor nodes periodically send out a special data packet (message) called a HELLO packet to establish and confirm network relationships with other sensor nodes. Once routers successfully exchange HELLO packets, they can automatically establish adjacency and can begin to route data between them.

3.3 Delay Method

Motive of delay-based method is to cause high end-to end latency thereby reducing throughput and QoS.

- **Jelly fish Attack**
  Jellyfish attack is similar to blackhole and grey hole attack but the difference is that in a jellyfish attack, the packet will first be delayed before and after transmission and reception in the network then it may drop packets also, whereas in a black hole attack the packet is only dropped. Jellyfish attack may also jumble the order in which packets are received before sending them in random order. The regular flow control method utilised by nodes for reliable transmission is disrupted as a result of this attack. A jellyfish assault can cause a long end-to-end latency, lowering QoS.

3.4 False Route Method

False route technique, as its name implies, entails sending data packets over the routing path in the incorrect direction, rendering the destination unreachable. Adversary nodes can do this by disseminating fake routing information or overflowing routing table with routes that do not exist. In this section we have briefly explained these attacks which directly or indirectly give their contribution to congestion occurrence are listed below.

- **Sinkhole Attack**
  During a sinkhole attack, an attacker compromises nodes in the network and advertises fictitious routing updates to lure network traffic towards the compromised node. It advertises fake messages saying it is the node with zero hop distance from the sink i.e., it imitates as sink in order to capture all the packets that were meant to receive by actual sink node [11]. A sinkhole attack can also be used to launch other attacks, such as selective forwarding, acknowledging spoofing, and dropping or altering routing information. A base station can also receive bogus information from it as the data packets can be modified by compromised node before relaying it to the actual sink node [12].
- **Wormhole Attack**
  The attack involves more than one malicious node and is carried out at the network layer. Compared to normal nodes, the nodes used in this attack are capable of establishing better communication channels over a wide area [13]. This attack involves tunnelling data between one malicious node and another at the other end of the network [14, 15]. Consequently, the other nodes in the WSN may believe

they are closer to other nodes than they actually are, causing routing algorithm problems. Compromised nodes may also interfere with data packets. It is also possible to combine wormhole attack with sinkhole attack in order to enhance its effectiveness [16, 17].

4 Experimental Results

We have setup a WSN in 100 * 100 m^2 region with 16 static sensor nodes that are randomly deployed and one sink node which is different from all the sensor nodes in terms of power, storage etc. We assume all the sensor nodes have equal initial energy, same hardware configuration and interface. For analyzing the impact of security attacks on congestion and associated network parameters such as packet delivery ratio (PDR), energy consumption, network lifetime etc. we have purposefully injected malicious nodes rendering different types of attacks into the network. The simulation also takes network deployment into consideration as attacks may have a different effect depending on the topology of the network, the software on the nodes, the hardware components or even the configuration of the nodes/networks. In this way, the WSN simulation will assist in identifying the most problematic attacks and the most vulnerablearts of the network. Table 1 shows the simulation parametrs and its corresponding values.

Table 1 List of simulation parameters

Network parameters	Value
Network area	100 * 100 m^2
Number of nodes	16
Initial energy of nodes	0.5 J
Round of simulations	200 rounds
Malicious nodes	1–4 nodes
Sink position	(50, 50), (50, 100)
Buffer size	10 packets
Data packet size	1024 Bits
Control packet size	200 Bits

4.1 Effect of Attacks in Different Network Scenarios

In this section, we examine how attacks behave based on the distances from sinks and network topology.

Experiment-1

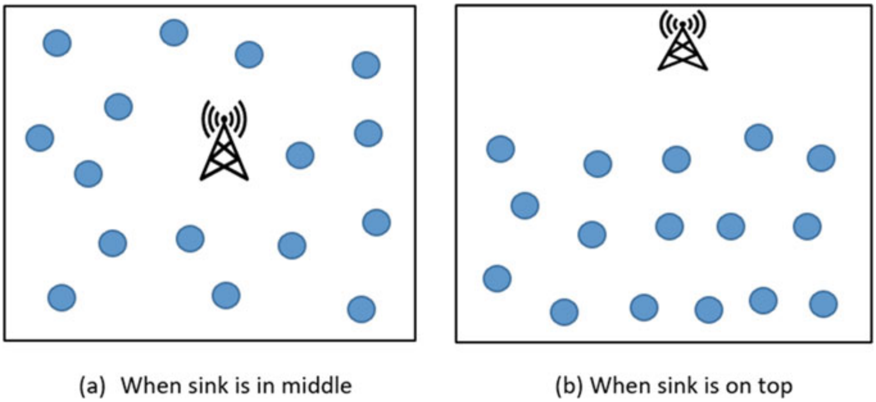
In this experiment firstly, we simulated the network without attacking node, then we intentionally introduced malicious nodes launching attack into the network and gradually minimised its distance from the sink to see the effect of attack when malicious nodes approach the sink. During this test, it is observed that interference increases whenever compromised node approaches the sink, and that the number of packets that are received decreases significantly. Table 2 outlines the whole experiment scenario. Firstly, in the network without attack total 50 packets were sent and all were received that means no packet drops happend. When network is victimzed with attack, position of attacker node from sink node is gradually decreased and it was seen that as distance between sink node and attacker node reduces number of packet drop increases.

Experiment-2

In this experiment we intend to analyse the behaviour of malicious nodes by changing the sink positions in the network or we can say by changing the network topology. We have taken two test cases; in first case we have deployed sink inside the sensing region and placed it in middle of the network area. (represented by Fig. 2a) and in second case sink has been placed farther from the sensing region (represented by Fig. 2b) thereby placed in center top position. And it is observed and shown in Fig. 3 that if sink is placed inside the sensing region it is less affected by some attacks as compared to sink which is placed farther from the sensing nodes, possible reason behind this could be that sensor nodes can find more alternate routes for packet transmission when sink is placed between the sensor nodes. The experiment involved transmitting 30 data packets to the sink when the network was free of attacks and when attacks were launched on it. And, it was witnessed that sink positions do not have any significant effect on the network in the absence of attacks as in both cases when (sink was postioned in middle and when sink was on top) almost all the data packets were successfully received by it. However, When the network contained malicious nodes, then a sink located inside the sensing region was less vulnerable to attacks than a sink located far away sensor nodes.

Table 2 Impact of security attacks on network when it approaches sink

	Without attack	With attack		
		Distance from sink 50 m	Distance from sink 30 m	Distance from sink 10 m
Total successful transmission	50	34	27	15
Total packet drops	0	16	23	35


(a) When sink is in middle

(b) When sink is on top

Fig. 2 Sensor network topology

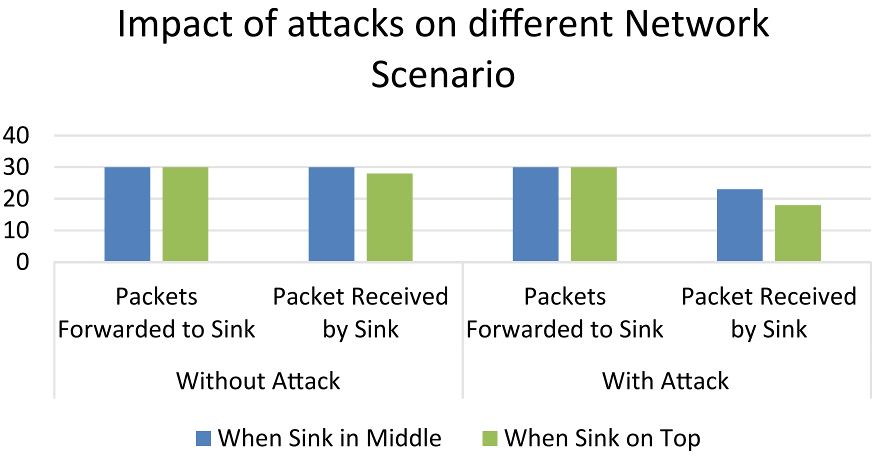


Fig. 3 Effect of attacks on different sink positions

4.2 Effect of Attacks on Important Network Parameters

This section shows the impact of various types of security attacks on crucial network parameters such as packet delivery ratio, Throughput and Network lifetime with the help of graph.

- **Impact of Attacks on PDR**

  "The packet delivery ratio (PDR) is defined as the ratio of total packets delivered to total packets sent (including RREP and RREQ) from a source node to a destination node in a network" [18]. The aim is to deliver maximum number of data packets sent to the destination. Figure 4 depicts what impact do the different

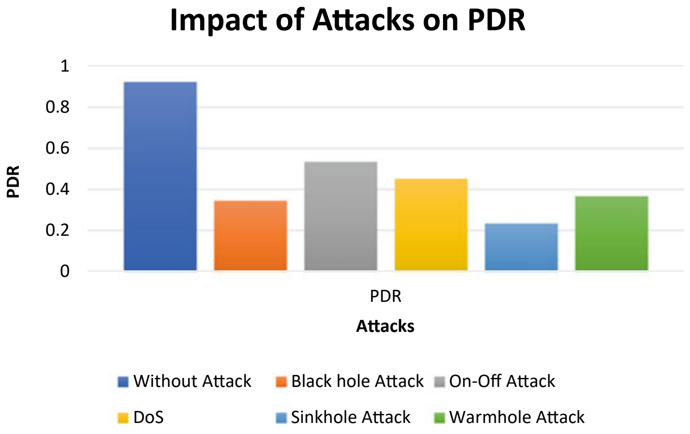
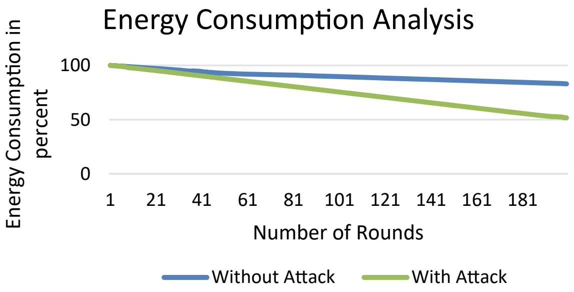


Fig. 4 Effect of attacks on packet delivery ratio

types of attacks have on the value of PDR. Network with all the legitimate nodes get the average PDR value between 0.91 and 0.98 whereas in presence of attacks it gradually decreases. We have observed that black-hole attack shows the worst effect on PDR.

- **Impact of Attacks on Energy Consumption**
  Some security attacks exploit WSN's limited energy resource by replicating massive amounts of data packets or by dropping lots of data packets that require multiple retransmissions, so they increase network energy consumption per transmission, which results in a shorter network lifetime. In the following graph (Fig. 5) impact of attacks on energy consumption has been shown. As demonstrated by the experimental results, security attacks can have extensive adversarial effects on energy consumption.

Fig. 5 Energy consumption analysis



5 Conclusion

Some security threats have a direct influence on network congestion which brings more processing and communication overhead, as well as an increase in energy consumption, which essentially limits network lifetime. The majority of existing congestion management techniques for Wireless Sensor Networks ignore the impact of security threats by malicious nodes on network congestion. Malicious nodes exacerbate congestion by delivering fraudulent messages. This paper examines possible network attacks that are crucial to understand while developing congestion countermeasures. A systematic analysis of the different WSN attacks has been carried out and existing research results have been critically analysed in this study.

References

1. Zheng J, Jamalipour A (2009) Wireless sensor networks: a networking perspective. Wiley, Hoboken
2. Yick J, Mukharjee B, Ghosal D (2008) Wireless sensor networks survey. Comput Netw
3. Pandey D, Kushwaha V (2020) An exploratory study of congestion control techniques in wireless sensor networks. Comput Commun 157:257–283
4. Dewal P, Narula GS, Jain V, Baliyan A (2018) Security attacks in wireless sensor networks: a survey. In: Cyber security. Springer, Singapore, pp 47–58
5. Abasikeleş-Turgut I, Aydin MN, Tohma K (2016) A realistic modelling of the sinkhole and the black hole attacks in cluster-based WSNs. Int J Electron Electr Eng 4(1):74–78
6. Gavric Z, Simic D (2018) Overview of DOS attacks on wireless sensor networks and experimental results for simulation of interference attacks. Ing Inv 38(1):130–138
7. Diaz A, Sanchez P (2016) Simulation of attacks for security in wireless sensor network. Sensors 16(11):1932
8. Tripathi M, Gaur MS, Laxmi V (2013) Comparing the impact of black hole and gray hole attack on LEACH in WSN. Procedia Comput Sci 19:1101–1107
9. Suma V (2021) Detection of localization error in a WSN under Sybil attack using advanced DV-hop methodology. IRO J Sustain Wirel Syst 3(2):87–96
10. Vivekanadam B (2020) A novel hybrid HNN and firefly algorithm to overcome denial of sleep attack on wireless sensor nodes. J Ubiquit Comput Commun Technol (UCCT) 2(04):223–227
11. Ganguly S, Chakraborty A, Naskar MK (2013) A trust-based framework for congestion-aware energy efficient routing in wireless multimedia sensor networks. arXiv:1312.4071
12. Yan J, Qi B (2021) CARA: a congestion-aware routing algorithm for wireless sensor networks. Algorithms 14(7):199
13. Gaware A, Dhonde SB (2016) A survey on security attacks in wireless sensor networks. In: 3rd international conference on computing for sustainable global development (INDIACom). IEEE, pp 536–539
14. Patel MM, Aggarwal A (2013) Security attacks in wireless sensor networks: a survey. In: International conference on intelligent systems and signal processing (ISSP). IEEE, pp 329–333
15. Kibirige GW, Sanga C (2015) A survey on detection of sinkhole attack in wireless sensor network. arXiv:1505.01941
16. Kavitha T, Sridharan D (2010) Security vulnerabilities in wireless sensor networks: a survey

17. Le A, Loo J, Luo Y, Lasebae A (2013) The impacts of internal threats towards routing protocol for low power and lossy network performance. In: IEEE symposium on computers and communications (ISCC). IEEE, pp 000789–000794
18. Pandey D, Kushwaha V (2019) Performance parameter analysis of congestion control in wireless sensor networks. In: 4th international conference on information systems and computer networks (ISCON). IEEE, pp 656–661

IoT Weather Forecasting Using Ridge Regression Model

Karthik G. Dath, K. E. Krishnaprasad, T. S. Pushpa, and K. P. Shailaja

Abstract Weather Forecasting is an Internet of Things (IoT) based initiative that attempts to provide weather forecasting via the internet via a website. Using several sensors, our project Weather Forecasting collects data such as temperature, humidity, rain, and pressure. Our project also includes a facility for measuring atmospheric conditions to give information for weather forecasts and to study weather and climate. As a result, Weather Forecasting using the Internet of Things is proposed to assist consumers in accessing weather data anywhere in real-time. For the data storage, a real-time database has been used using Firebase and will be displayed in the dashboard which was designed using React JS.

Keywords Internet of Things · Weather forecasting model · Regression model · Sensors

1 Introduction

Weather forecasting is an application of science and technology to forecast weather conditions based on prior data. Weather forecasting has been attempted formally since the nineteenth century and informally for millennia. Weather forecasts area unit created by assembling quantitative knowledge on this state of the atmosphere, land, and ocean, and so applying meteorology to forecast however the atmosphere can modification at a particular place.

K. G. Dath (✉) · K. E. Krishnaprasad · T. S. Pushpa · K. P. Shailaja
B.M.S College Engineering, Bangalore, India
e-mail: karthikgd.mca20@bmsce.ac.in

K. E. Krishnaprasad
e-mail: krishna.mca20@bmsce.ac.in

T. S. Pushpa
e-mail: pushpa.mca@bmsce.ac.in

K. P. Shailaja
e-mail: shailaja.mca@bmsce.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_53

Our project Weather Forecasts is an Internet of Things-based project that intends to provide weather information and forecasting using the Ridge Regression (RR) Model, a machine learning method. Using sensors from the surrounding environment, our project captures data such as temperature, humidity, rain, and pressure. Our project also includes the ability to measure atmospheric conditions to provide data for weather forecasts. We used Firebase's real-time database to store sensor data, and a ReactJS-based website to display information.

We store the weather data in ThinkSpeak, which is an open-source IoT analytics platform. It visualises data and allows us to download the recorded data together with the timestamp. The API supports data retrieval and logging by connecting to devices and websites. It can also perform online analysis and processing of the data it comes from.

Using the RR model, we use the acquired data to forecast the weather. As a result, we advocated that people have real-time access to weather information wherever they are. Also, because we collected data in our home setting, the forecasts may not be completely accurate.

2 Hardware Requirements

Node MCU (ESP8266) as in Fig. 1 is a low-cost and powerful open-source microcontroller board which connects objects and lets data transfer using Wi-Fi protocol [5]. It is easy to use and can be programmed with Arduino IDE, it is available as an access point or station and also consist of an internal antenna.

The absolute pressure of the air around Barometric Pressure sensors is measured as in Fig. 2. This pressure is affected by both the weather and altitude. Depending on how you interpret the data, you may monitor weather changes, measure altitude, or do any other operation that necessitates an exact pressure reading [6]. The ultra-low power consumption down to 3 μA makes the BMP 180 the leader in power saving for the devices. BMP 180 is also distinguished by its very stable behaviour (performance) regarding the independence of the supply voltage [7].

The DHT11 in Fig. 3 is a low-cost sensor, used to measure temperature and humidity [9]. The DHT-11 Sensor is a straightforward, incredibly affordable digital

Fig. 1 NodeMCU ESP8266

Fig. 2 Barometric pressure (BMP) 180

thermometer. It measures the air around it using a capacitive humidity sensor and a thermistor, and it spits out a digital signal on the data pin (no analogue input pins are needed).

The raindrop sensor in Fig. 4 detects rain. It has two modules one that detects rain and another the control module that compares and converts analogue values to digital values. The analogue output is employed as a section of discovery of drops within the life of rainfall related to three 3V/5V power provided and the device works in lightweight of the amount of the water interfacing the raining board, the output voltage of the gadget varies on the length of the raining board being wet that is modified over to digital through ADC chip [8].

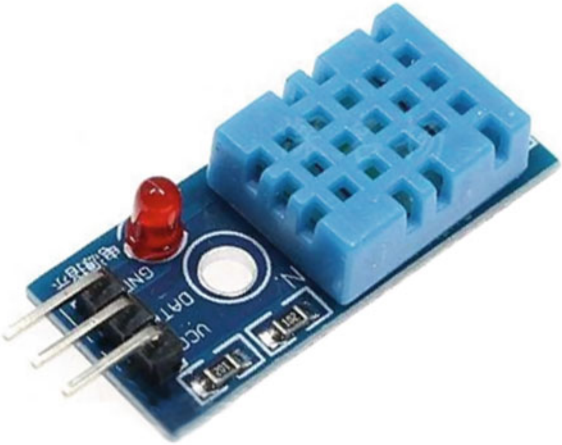


Fig. 3 Digital humidity temperature (DHT 11)

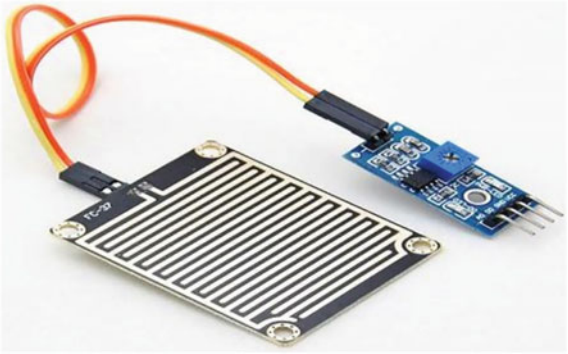


Fig. 4 Raindrop sensor

3 Software Requirements

Arduino IDE: Arduino Software (IDE) is an open-source editing software for writing code and uploading code to any compatible Arduino microcontroller board. Arduino board can be also used with other third-party boards by installing appropriate libraries [1].

ReactJS: ReactJS is a JavaScript library for creating user interfaces, single-page applications, and reusable UI components [2].

Firebase: Firebase is a Google back-end and cross-platform application development that helps you build and deploy. Firebase provides services like real-time database, storage, hosting, authentication, function and ML. In our project, we have used Hosting and real-time database services. [3].

Jupyter Notebook: it is an open-source tool for writing and execution of the program. It is used to create and share documents that contain live code, equations, visualizations, and text [4].

ThinkSpeak: It is a cloud-based IoT platform for the analysis and visualization of IoT data. It visualises data and allows us to download the recorded data together with the timestamp. The API supports data retrieval and logging by connecting to devices and websites. It can also perform online analysis and processing of the data it comes from.

4 Working and Design

Our project starts with the sensors collecting weather data in its surrounding environment and pushing it to firebase real-time-database and think to speak cloud. The data stored in firebase is used to display our React website and the data collected in

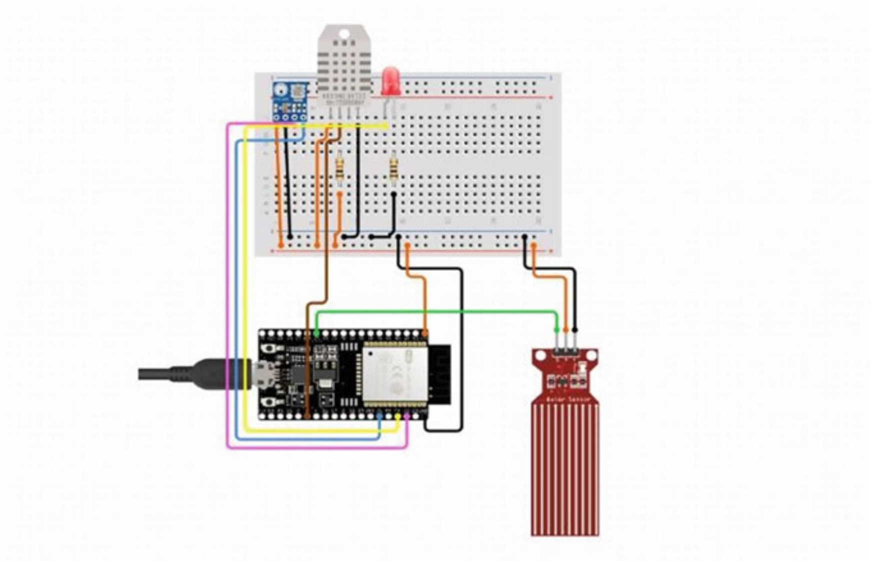


Fig. 5 Model design

think speak is used to visualize and export the collected data as a CSV file which is later used to train our model using the Ridge Regression (RR) model.

We have used the Ridge regression (RR) model to predict the weather. Ridge regression may be a model standardization methodology that's accustomed to analyses any knowledge that suffers from multiple correlations. This method will use L2 regularization [12]. Uses the maximum temperature (Tmax) and minimum temperature (Tmin).

Figure 5 describes the connection of sensors on a breadboard, where DHT 11, BMP 180, Rain sensor and ESP8266 Wi-Fi modules are connected.

5 Literature Survey

Paper 1: **Weather Prediction Using Machine Learning**.

Publication Date: 05 May 2021.

Authors: Abhishek Patel, Pawan Kumar Singh and Shivam Tandon.

Publisher: Galgotias University-Galgotias University School of Computing Science and Engineering.

Abstract

Climate conducts a completely critical function in many key production sectors, e.g., farming. Climate change with high charging these days, which is why old weather forecasts are getting closer and less powerful and continue to be annoying. Miles is therefore very important to decorate and modify the weather forecast model. those predictions affect the country's financial system and people's lives. A system of information and statistical analysis algorithms has been used that includes a wooded area used for weather forecasting [10].

Methodology

This paper mainly focuses on the basic Machine Learning model, which has been an automated system to gather historical data. The paper focused to help future generations of people where prediction of weather helps lives to get prepared for the weather changes, taking precautions of bad weather conditions etc. Paper concentrated on format effective weather forecast and performing a function of accurately predicting the weather [10].

Paper 2: **Smart Weather Forecasting Using Machine Learning**.

Publication Date: 25 August 2020.

Authors: A H M Jakaria, Md Mosharaf Hossain and Mohammad Ashiqur Rahman.

Publisher: A Case Study in Tennessee.

Abstract

Traditionally, weather predictions are performed with the help of large complex models of physics, which utilize different atmospheric conditions over a long period. These conditions are often unstable because of perturbations of the weather system, causing the models to provide inaccurate forecasts. The models are generally run-on hundreds of nodes in a large High-Performance Computing (HPC) environment which consumes a large amount of energy. The paper presents a weather prediction technique that utilizes historical data from multiple weather stations to train simple machine learning models, which can provide usable forecasts about certain weather conditions for the near future within a very short period. The models can be run in much less resource-intensive environments [11].

Methodology

This paper predicts the weather by collecting the historical data from multiple weather stations train simple machine learning models, which can provide usable forecasts about certain weather conditions for the near future within a very short period. The models can be run in much less resource-intensive environments [11].

Paper 3: Temperature Forecast Using Ridge Regression as Model Output Statistics.

Publication Date: 30 April 2020.

Authors: Niswatul Qona'ah, Kiki Ferawati, Muhammad Bayu Nirwana and Sutikno.

Publisher: Universitas Sebelas Maret.

Abstract

This study uses the maximum temperature (Tmax) and minimum temperature (Tmin) observation at 4 stations in Indonesia as the response variables and Numeric Weather Prediction (NWP) as the predictor variable. The results show that the performance of the model based on Root Mean Square Error of Prediction (RMSEP) is considered to be good and intermediate. The RMSEP for Tmax in all stations is intermediate (0.9–1.2), Tmin in all stations is good (0.5–0.8) [12].

Methodology

This uses the maximum temperature (Tmax) and minimum temperature (Tmin) observation at 4 stations in Indonesia as the response variables and NWP as the predictor variable. The result from Ridge Regression is more accurate than the Numerical Weather Prediction model and also it corrects up to 90.49% of the biased NWP for Tmax forecasting [12].

Paper 4: Device and Development of Automatic Microcontroller-based Weather Forecasting Device.

Published Date: March 2020.

Author: Dr Bindhu V.

Publisher: PPG Institute of Technology, Tamil Nadu.

Abstract

The proposed method utilizes the sensors to monitor the weather changes and engages the raspberry pi to process the information gathered and convey it to the end user. The proposed system was tested by implementing it in the Indian delta districts and the accuracy, precision and flexibility in the forecasting were evinced by the data output observed over and done with the ThinkSpeak [13].

Methodology

The proposed model develops a weather monitoring device to measure the meteorological parameters by employing the sensors, the sensors employed measure every minute changes in the atmosphere and convey it to the web server for the direct access of the users, proffering an accurate forecast of weather [13].

Paper 5: A Methodology of Atmospheric Deterioration Forecasting and Evaluation Yjrough Data Mining And Business Intelligence.

Published Year: 2020.

Author: J V Anand.

Publisher: PACE Institute of Technology and Sciences, Ongole.

Abstract

This paper emphasis on an instinctual and efficacious method to forecast and analyze the condition of different atmospheric determinants all over the world. The difficulty in the extant remedies or the apparatus is its incapability to provide comprehensive information regarding the evaluation of the attributes of the atmosphere. The proposed methodology in the paper gathers the actual information about the atmospheric attributes such as the water, air, the forest and the tree cover etc. from the government bases and processes the collective information [14].

Methodology

The methodology does the extrication transformation load over the original collective data that are in its raw format. The converted information sets are imported into the database to develop a dashboard with the multiple information displayed on it. This allows having evaluated data about the various atmospheric factors. To forecast the deteriorations and the conditions of the atmospheric attributes the methodology proffered utilizes the Fuzzy C means clustering, R-studio, and the ARIMA framework [14].

6 Equations

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{1}$$

Ridge regression places a particular form of constraint on the parameters **(β’s): β^ bridge** is chosen to minimize the penalized sum of squares. which is equivalent to minimization of **∑i = 1n(yi − ∑j = 1pxijβj) ^2** subject to, for some **c > 0, ∑j = 1pβj2 < c**, i.e. constraining the sum of the squared coefficients.The related equation is shown in Eq. (1).

Therefore, ridge regression puts further constraints on the parameters, βj’s, in the linear model. In this case, what we are doing is that instead of just minimizing the residual sum of squares we also have a penalty term on the **β’s**. This penalty term is **λ (a pre-chosen constant)** times the squared norm of the **β** vector.

This means that if the **βj’s** take on large values, the optimization function is penalized. We would prefer to take smaller **βj’s, or βj’s** that are close to zero to drive the penalty term small [15].

7 Result Screenshots

Figure 6 shows our implementation of the weather station using IoT components.

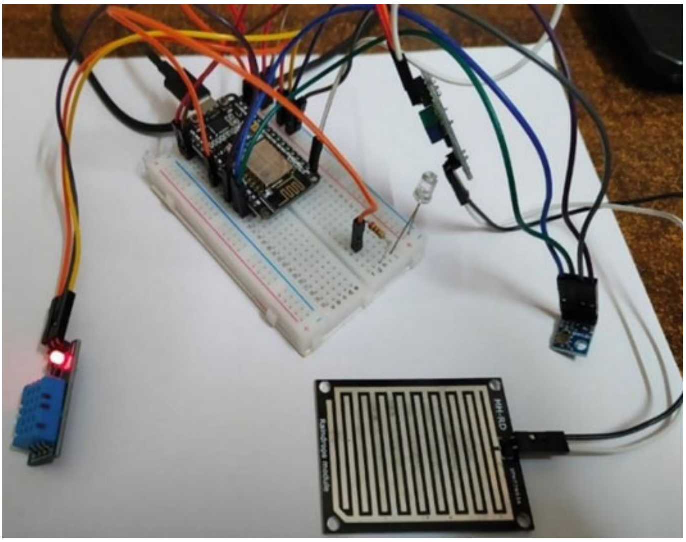


Fig. 6 Model and connections of project

Figure 7 shows Arduino serial monitor which is displaying the messages which are written in the program.

Figure 8 shows the Firebase Realtime Database, which is used as a backend for Reacts website.

Figure 9 is the website which is disapplying of the weather information.

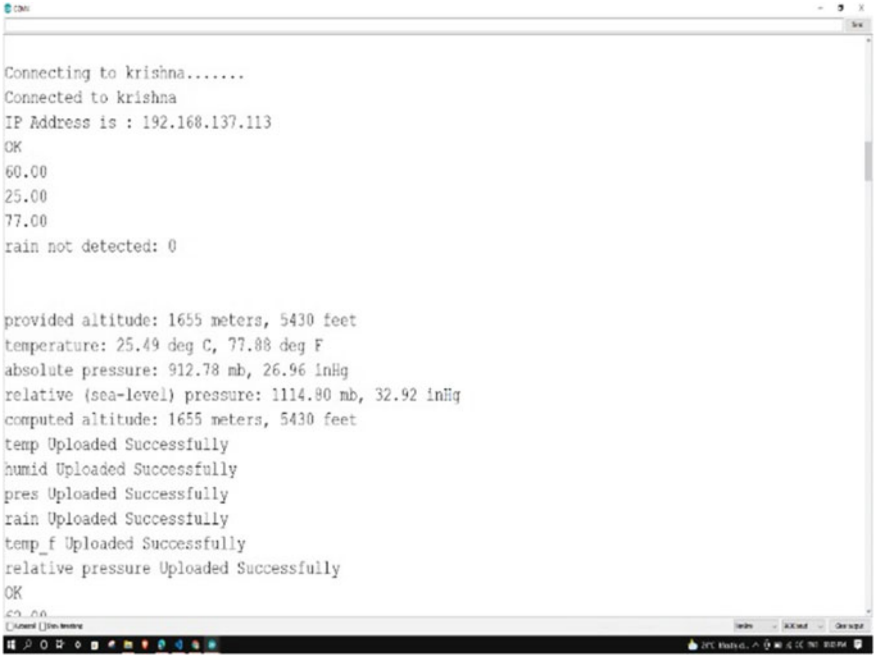
Figure 10 shows the ThinkSpeak cloud which we have used to store the weather information.

Figure 11 shows our data set.

Figure 12 shows the graph of prediction from the data collected. The graph show the actual data and predicted data.

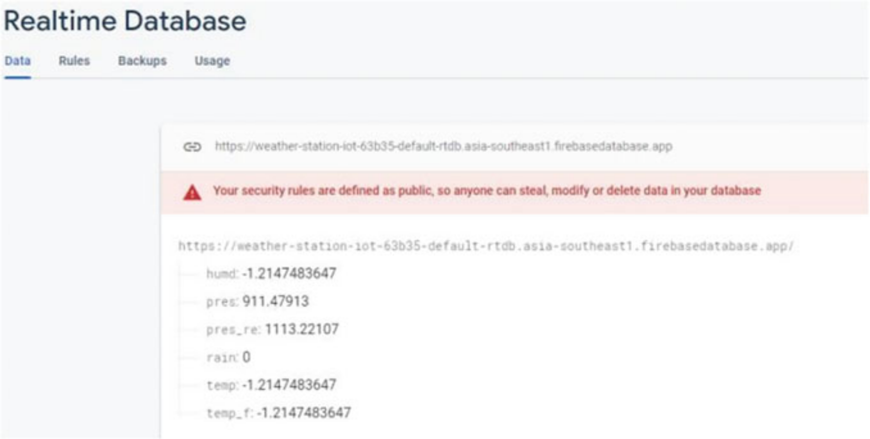
8 Conclusion

Weather Forecasts is a project based on the Internet of Things that aims to give weather information and predictions using the Ridge Regression (RR) Model, a machine learning method. Ridge regression is a model tuning technique used to analyze data with multicollinearity. The challenges we faced while doing this project include collecting data from the sensors and storing it. Most of the tools used in this project are open-source tools. Rain is one of the data that we were unable to collect and the data collected is from constant locations instead of multiple locations. We aim at improving the project in future for better results.



```
Connecting to krishna.......
Connected to krishna
IP Address is : 192.168.137.113
OK
60.00
25.00
77.00
rain not detected: 0


provided altitude: 1655 meters, 5430 feet
temperature: 25.49 deg C, 77.88 deg F
absolute pressure: 912.78 mb, 26.96 inHg
relative (sea-level) pressure: 1114.80 mb, 32.92 inHg
computed altitude: 1655 meters, 5430 feet
temp Uploaded Successfully
humid Uploaded Successfully
pres Uploaded Successfully
rain Uploaded Successfully
temp_f Uploaded Successfully
relative pressure Uploaded Successfully
OK
```

Fig. 7 Serial monitor

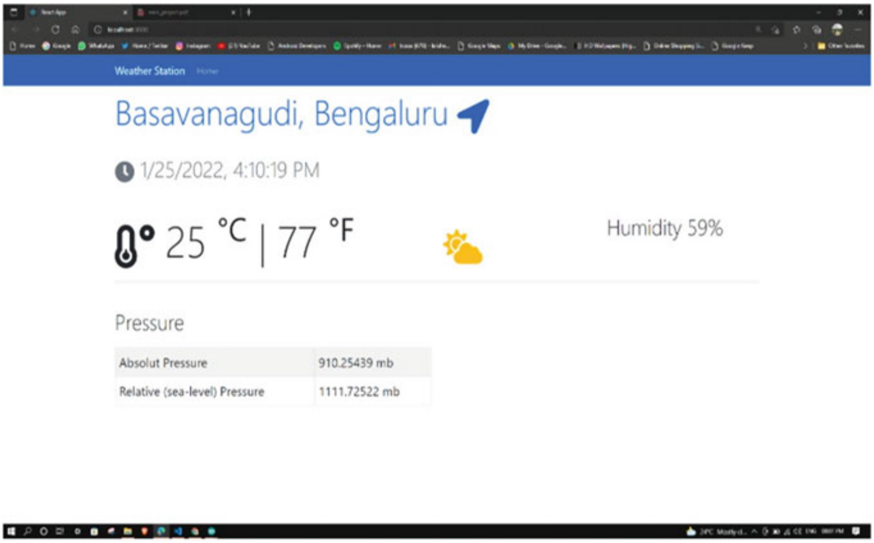


https://weather-station-iot-63b35-default-rtdb.asia-southeast1.firebasedatabase.app

Your security rules are defined as public, so anyone can steal, modify or delete data in your database

```
https://weather-station-iot-63b35-default-rtdb.asia-southeast1.firebasedatabase.app/
    humd: -1.2147483647
    pres: 911.47913
    pres_re: 1113.22107
    rain: 0
    temp: -1.2147483647
    temp_f: -1.2147483647
```

Fig. 8 Realtime database



Basavanagudi, Bengaluru

1/25/2022, 4:10:19 PM

25 °C | 77 °F

Humidity 59%

Pressure

Absolut Pressure	910.25439 mb
Relative (sea-level) Pressure	1111.72522 mb

Fig. 9 Dashboard created using React JS

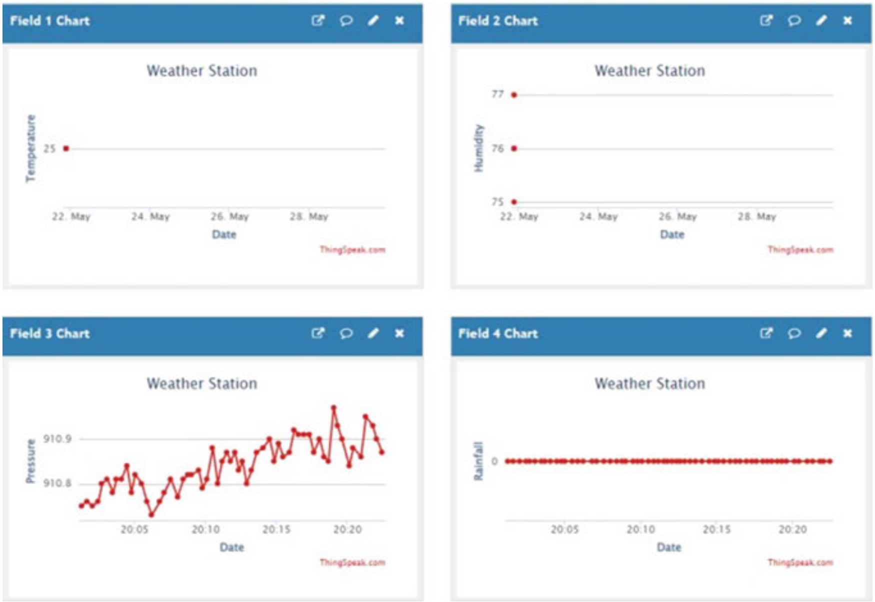
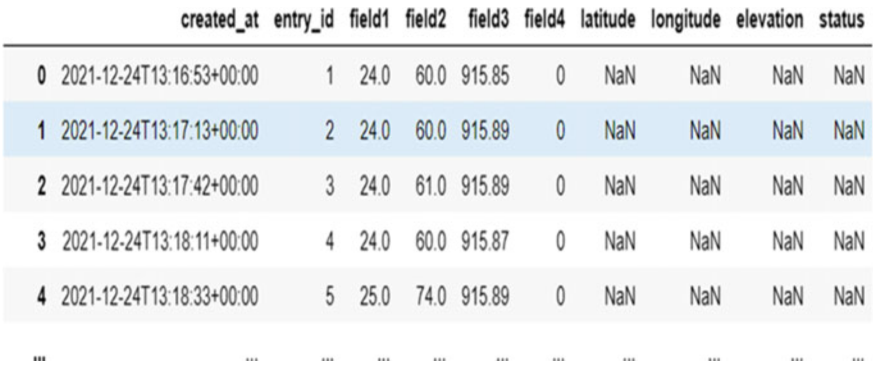


Fig. 10 Data collected and visualized by thing speak cloud



	created_at	entry_id	field1	field2	field3	field4	latitude	longitude	elevation	status
0	2021-12-24T13:16:53+00:00	1	24.0	60.0	915.85	0	NaN	NaN	NaN	NaN
1	2021-12-24T13:17:13+00:00	2	24.0	60.0	915.89	0	NaN	NaN	NaN	NaN
2	2021-12-24T13:17:42+00:00	3	24.0	61.0	915.89	0	NaN	NaN	NaN	NaN
3	2021-12-24T13:18:11+00:00	4	24.0	60.0	915.87	0	NaN	NaN	NaN	NaN
4	2021-12-24T13:18:33+00:00	5	25.0	74.0	915.89	0	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...

Fig. 11 Dataset used

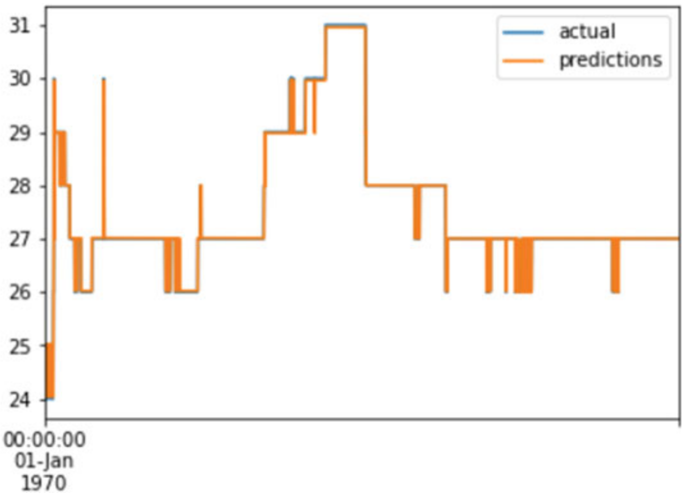


Fig. 12 Predicted value with the actual value

9 Future Enhancement

1. Collecting weather data from multiple locations.
2. UV index.

3. Displaying predictions on the website.
4. Analyze wind speed and pattern.

References

1. The Arduino Team (n.d.) Getting started with Arduino products. https://www.arduino.cc/en/Guide/
2. React (2019) Getting started—React. https://reactjs.org/docs/gettingstarted.html
3. Google (n.d.) Firebase realtime database | firebase documentation. Retrieved 18 Mar 2022 from https://firebase.google.com/docs/database
4. https://realpython.com/jupyter-notebook-introduction/
5. Nodemcu ESP8266. Components101 (n.d.) Retrieved 18 Mar 2022 from https://components101.com/development-boards/nodemcu-esp8266-pinout-features-and-datasheet
6. Industries A (n.d.) BMP180 barometric pressure/temperature/altitude sensor-5V ready. Adafruit industries blog RSS. Retrieved 18 Mar 2022 from https://www.adafruit.com/product/1603
7. https://ae-bst.resource.bosch.com/media/_tech/media/product_flyer/BST-BMP180-FL000.pdf
8. Rain drops sensor module. Components101 (n.d.) Retrieved 18 Mar 2022 from https://components101.com/sensors/rain-drop-sensor-module
9. DHT11 humidity and temperature sensor-Mouser.com (n.d.) Retrieved 18 Mar 2022 from https://www.mouser.com/datasheet/2/758/DHT11-Technical-Data-Sheet-Translated-Version-1143054.pdf
10. Patel A, Singh PK, Tandon S (2021) Weather prediction using machine learning. SSRN: https://ssrn.com/abstract=3836085 or https://doi.org/10.2139/ssrn.3836085
11. Jakaria AHM (2020) Smart weather forecasting using machine learning:a case study in Tennessee. arXiv.Org. https://arxiv.org/abs/2008.10789
12. http://sunankalijaga.org/prosiding/index.php/icse/article/view/533
13. Bindhu V (2020) Design and development of automatic micro controller based weather forecasting device. J Electron Inf 2(1):1–9
14. Anand JV (2020) A methodology of atmospheric deterioration forecasting and evaluation through data mining and business intelligence. J Ubiquit Comput Commun Technol (UCCT) 2(02):79–87
15. The Pennsulvania State University Ridge regression model equation. https://online.stat.psu.edu/stat857/node/155/

Automated Cloud Monitoring Solution: Review

Ishwari Deshmukh and Jayshri D. Pagare

Abstract The Cloud adoption statistics clearly depicts that considerable number of organizations has hosted their infrastructure on cloud, most of the major migrations from on premise server to cloud are done and some are in progress. With this Cloud Computing era, Infrastructure as a service requires more efficient management. It comprises of IT resources for storage, networking, computing etc. Optimized and efficient use of the rented resources is the need of hour as resources are on rented basis and they have cost associated with it. As industry is growing, users are growing, resources are growing followed by volume and traffic. To manage any cloud infrastructure there is need of cloud monitoring solution with proactive alerting. This paper will provide entire overview with focus of providing open-source solution to monitor cloud environment.

Keywords Cloud computing · Cloud monitoring · Monitoring phases · Beat agents · Alerting · Monitoring metrics · Parsers · Data shippers

1 Introduction

Cloud Computing is a model that provides many on demand services. XaaS is the common terminology used for this purpose. It means that anything can be outsourced and integrated with existing application to improve performance or quality of service. As Cloud model comprised of three main services as Infrastructure as a Service, Platform's a Service, Software as a Service. We also have technologies as Communication as a Service, Database as a Service, and Monitoring as a Service. To emphasis on Monitoring as a Service which is used to improve quality of Service and to attain service level agreements at accuracy. Apart from that it also helps to troubleshoot, find out root cause analysis, identify various vulnerability and threats over the application environment.

I. Deshmukh (✉) · J. D. Pagare
Department of Computer Science and Engineering, JNEC, Aurangabad, Maharashtra, India
e-mail: deshmukhishwari9@gmail.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_54

With each deployment model i.e. public cloud, private cloud, hybrid cloud there comes the need of monitoring the computing resources to use resources in optimised way. For public cloud resources are made available over the internet to multiple known or unknown users. Similarly private cloud exposes and provides access to limited and most of the known audience. However, hybrid cloud is combination of above-mentioned cloud, but monitoring aspects remain constant irrespective of deployment model. Scaling up of resources or scaling down of resources plays an important part. Organization often terms it as capacity management or capacity planning. Entire moto of the proposed system its t monitors the environment with least cost and best quality of service with minimal or no failure of computing resources.

2 Motivation to Design

With increasing complex cloud infrastructure, it is especially important to have a stable monitoring solution applied over the cloud cluster. Moreover, existing resources are on rented basis i.e., already cost is to be paid for them. Resources here refers to the processors, storage, or network [1]. There is a need to identify and design an architecture that is efficient to monitor the cloud infrastructure with minimal or no cost. This solution should be effective and reliable as very less of negligible cost is associated to build and maintain it [2].

Cloud Computing involves many activities for which monitoring is an essential task. The most important ones are like.

(1) Capacity and Resource Planning and Management.
(2) Security management.
(3) Data Center Management.
(4) SLA Management.
(5) Billing.
(6) Performance Management.

Traditional monitoring or availed monitoring services are mostly domain specific in nature. This is an attempt to provide a versatile, low cost and effective solution for cloud infrastructure [3]. Using Monitoring as a service from cloud vendor will ultimately require cost of two thing. Firstly, it will be all the rented resources or platform, or it can be software service that is used directly [4]. Second part will comprise of Monitoring service which will continuously run over the cloud infrastructure, and it will cost per minute [1, 5]. In Order to provide cost effective open-source solution that can be designed and used as per domain specific need one can use various automation tools combine them and create a complete open-source solution. Hence Monitoring has become an important aspect that is to be considered while designing as well as maintaining cloud infrastructures.

Monitoring the cloud infrastructure and analysing the gathered statistics helps for capacity management. Resources or instances can provision based on the need and as required they can be decommissioned. This ultimately defines on demand pay as used model [6]. Automation helps to upscale and downscale the system. No human intervention is required in entire process and the accuracy is high. This defines the term rapid elasticity which the monitoring solution will provide. Used services can be measured, monitored, and reported [5].

3 Architecture of Monitoring System

Monitoring architecture can be broadly classified in three major steps.

(1) Data Collection
(2) Data Analysis
(3) Data Visualization.

These three steps show high level picture of any monitoring solution. From granular perspective there are multiple operations that happens within this phase. These included collection of data, transportation of data, processing the data at required processing time, presenting thee data in most effective way [1, 7]. Apart from that every phase must be associated with some security protocol so that security breach can be avoided. Cloud monitoring architecture is shown in Fig. 1.

Basic requirement to set up any monitoring solution is environment that is to be monitored, light weight shippers to ship data, on demand parser, storage for gathering the stats, UI to visualize the gathered data. These are basic primary requirement for any monitoring solution that is to be built [8]. Initial stage design of architecture plays a significant role. Infrastructure to be monitored need to be measured as all the instances will forward the statistic to main centralized server. Average no of hits per minute need to check in order to build the centralized server [1, 5, 7]. One can design it as master server architecture, where the master comprises of universally available and efficient parser and slaves, or the servers are the devices that needs to be monitored [9].

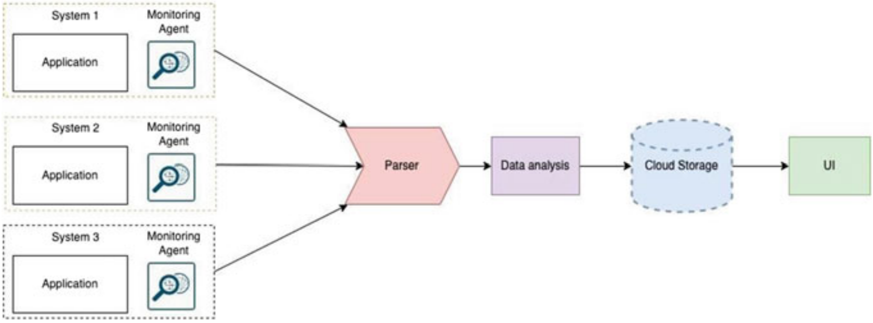


Fig. 1 Cloud monitoring architecture

Phase one comprises of data collection. Here data can be referred as files, log files, metrics, information stats of cloud services, health rules for the instances, etc. [3]. Broadly this monitoring solution can be classified into two types based on the data. It can Application monitoring or Infrastructure monitoring. Application monitoring is collection of application logs, application of grok patterns on it and storage in form of attribute and value. E.g.: A java application has various exceptions. One can setup alert if any such exception string occurs in log. Infrastructure monitoring is where system related metrics are gathered e.g.: CPU, Memory, Disk, Process, etc. Alerts can be added if any of the mentioned parameter breaches the threshold value that are defined already [10].

Overall, any event that is to be captured and monitored or visualized can be considered as data set. Although it depends upon design architecture what data to collect and what part of data to be extracted. To gather this data once can use beat agents [1, 11]. These agents are lightweight programs that run on the server. These shippers run on server as service, and they run throughout in the background without consuming resources. Main function of this agent is to push data to centralized server or the parser. In some architecture where parser is absent it could be direct storage. It can be static database of cloud availed database service. One can gather events from automation scripts. Also, various API can be used to fetch data. Collection of data can also be done with plugins or protocols such as Nagios or NRPE. Ready to install beat agents are available which are mainly written in java or GO language one can ship data without constraint of compatibility [2, 6, 7, 12]. Transportation of events or logs are mainly taken care by the shippers. It's important to have encrypted mechanism if data is shipped with hops over the cluster. This will help to avoid all the security breaches [13].

Phase two comprises of as centralized server where all the data is accumulated and sent over to storage device or storage service. Each and every event that gets generated with help of beat agent is undergone through a parser device. Here various grok patterns can be placed. Segregation of data can be done at this level. Key value pair format can be used to store data. This parsing helps to visualize the data on graphical UI easily. One can plot various visualizations on segregated data and analyse the behaviour [2, 14].

The centralized server designed in this phase should be capable to oversee all the network load. Network load here refers to the events that are transferred all the way from cloud infrastructure. Parser or number of parsers should be defined based on no of events that we are going to receive. One can receive any number of events with any frequency. Mostly frequency started from 2 s. However, this is relative to the use case, if application is critical one can reduce the time to seconds otherwise its ca be in minutes interval [15]. Based on that infrastructure should be designed. Parser should be designed in such a way that it should oversee all the load. Multiple nodes of the parser can be installed. Nodes should be available with HA mode i.e., high availability mode. If one of the instances goes down other should be capable enough to handle all the traffic in the environment. Also, for larger clusters load balanced parsers can be applied. Events load will be routed and distributed evenly over the load balance parser [2, 16].

Phase three mainly focus to store and visualize the data. As mentioned earlier storage device should be efficient in terms of volume. It can be static storage or simple database or cloud service as s3. Elasticsearch database can be used here to store the data in Json format. One data get segregated with key value pair we can use it easily to analyse the data behaviour [1, 2].

Other important aspect is to hire the storage that can oversee substantial number of read write operation within low frequency and should provide efficient output with near real time. Time lapse parameter should be considered as further alerting or visualizing process is dependent on this. Data can be aggregated and stores in attribute value format. This can be depicted in form of graph, line chart, pie chart, heap map, etc. Various dashboards and canvas reports can be made out of it [17]. Stored data can be aggregated on any configured attribute that is present in logs which makes fault finding and debugging easy. Alerts can be configured on stored data in form of email. Also, integration can be extended to any destination or any ticketing tool for further analysis. Also, Machine learning modules can be applied over the data. Module can predict number of resources required based on prior knowledge also they can predict the failure in applications [1, 2, 18].

4 Results and Performance Analysis

This architecture can be built in both ways, i.e., centralized or decentralized as in Fig. 2. A solution can have hardware device with CPU that has processor greater than two cores with Memory more than 4 GB. Any Linux flavours will suffice the need. Light weight agent is installed on the system. Agent is java-based agent that is built in GO lang. Agent varies as per need. E.g. If system statistics are to be gathered then metric beat type of agent to be installed on the system. If file data or log files need to be stored, parsed, and monitored then file beat type of gent is required to be installed. Similarly, if application health or URL are to be monitored then heartbeat type of agent to be installed.

Events that are getting collected can be parsed. The required strings are to be captured and stored in database. Unusual data can be discarded at initial stage. Parser collects the data at specific port e.g., 5044. Once event is received segregation of string happens so as to store the data in attribute value format. Time stamp data or real time data is parsed. Configuration file can be considered of three important parts namely input, filter, output. Input part listen to mentioned port. Filter part segregates the data and directs it to desired index mentioned in output part. Here the output is Elasticsearch type of database where data is stored. All graphical interfaces run on this.

Gathered data is seen in form of indices. Indices are used to create aggregated Dashboards. Canvas reporting the real time reporting is built on top of it. Java program watchers can be used to monitor the data. Every 5 min a program gets executed to identify if all nodes are reporting and all the nodes in the environment are performing well. Integration with SMTP server is required if Elasticsearch data alerts are to be sent

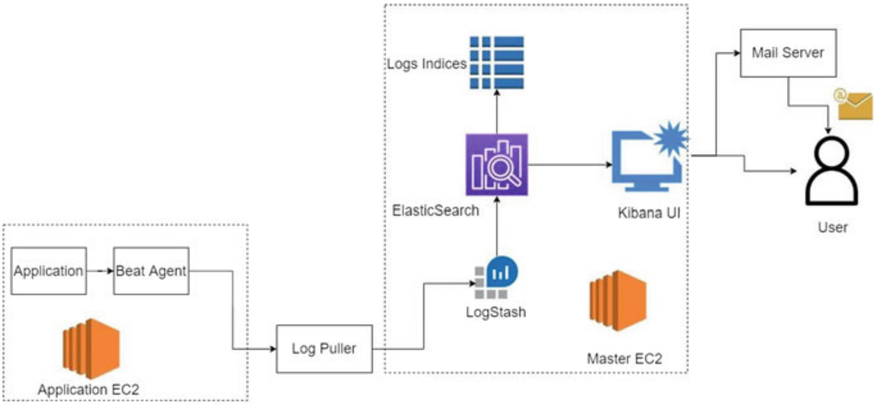


Fig. 2 Designed architecture

in email format. This gives complete pictorial view of application and infrastructure on single dashboard.

Replication of proposed solution is easy as time taken for development is one time effort. Solution being scalable in nature can be scaled using various automation tool. Agents or monitoring shippers can be installed across entire infrastructure in using single playbook. Expansion of parser requires replication of existing parser. To generalize the fact, time complexity to expand the solution is minimal.

5 Advancement

As we have pass dated information that is parsed and stored over the cloud which is nothing but the data set for further uses. Here we can use the data set in two ways as prediction and dynamic alert configuration.

Prediction comprises of analysing the data. Identifying the trends for past dates and predicting the future based on past data. This can be used for early resource planning. For example, streaming platform where traffic gets increased on holidays. We can refer to the trends that are predicted by the machine learning model to plan server capacity that can fulfil request without delay. Similarly, this can be used for applications business management [4].

With the same predictions, adjustment of thresholds can be done in dynamic manner. Alerting part disused in phase three above has threshold values that were define. These thresholds are generalized as certain number of resources that should be consumed for application to be active. However, some positive or negative deviation in this consumption is sign of some discrepancy. Dynamic threshold definition will help for proactive monitoring [19].

6 Comparison of Cloud Monitoring Solutions

In comparison to existing solution, we have below mentioned highlights that make the system more effective. Here existing solution mainly refers to the on-demand services provided by cloud service providers as Amazon, Google or Microsoft. It can be also any monitoring tool as Nagios or AppDynamics. Tool Nagios uses NRPE kind of protocol that acts as beat agent to transfer data from source to destination. Nagios allows users to design and deploy customised plugin over the infrastructure. This designing of plugins is easy and ca be used. AppDynamics is has generic architecture wherein shippers are associated to transfer data. However, dynamic real time monitoring can be obtained using it. In comparison to all these tools given solution has below mentioned benefits [8, 11].

Cost Effective: To use any of the service one has to pay the cost with pay per use module. For monitoring service, the use is very high as infrastructure or application should be always under monitoring to avoid outages. Above given solution is totally open source and no cost is required and hence monitoring is free. However minimal storage cost is required if data storage to be done.

Customized Monitoring: One can monitor application logs and system logs. Grok pattern can be applied on data to find error, warnings or any of the desired string. One can parse and find various error codes. Application end points can be polled at certain interval to check if they are up. This all leads to proactive monitoring and zero downtime.

Easy to Setup and Maintain: Installation of agents over the servers is simple job with automation tools as ansible or chef. Single server with playbook can install agents on multiple servers within less time. Similarly, maintenance and upgrade become one step activity.

Scalable Solution: This solution can be implemented on any of the cloud infrastructure. Scaling of application is quite easy as one has to replicated the parser among various load balancers if required. Also, storage can be scaled if traffic is high. However, with all the changes implementation logic remains same.

Static/Dynamic Proactive Alerting: As mentioned earlier, alerting works on static thresholds as well as it works with dynamic threshold. Alerting being proactive i.e., alerting user before issue occurs. This will provide minimum or no outage over the application business [1, 4].

7 Conclusion

Client machines are the small nodes that will be idle but treated as application servers. These machines will have shippers that are placed over. These shippers will be lightweight and can be installed manually or in automated way as well. Small programs will be running all the time to monitor the instance. Solution being open source will be easy to manage and maintain in cloud infrastructure. Now we have multiple client that will reside in the cloud with that we will have centralized server that will be in same environment. Metrics will be fetched from clients and can be stored on server. Analysis can be done on saved data. Data can be visualized into gauge, graphs, line charts, bar charts. This dashboard can be plotted as and when required. Moreover, alerting, and proactive alerting can be set up based on data requirements.

This alerting feature will help the user to get alerts when something is going wrong or few of the features in application are not working as expected. This will help end user to identify and fix the failure before system gets crashed. One can analyses entire cluster in once screen. This solution is scalable and can be replicated over any cluster. This solution being open source can be replicated. It can be placed on two nodes clusters to two thousand with infra design that is capable to handle the load.

References

1. Birje MN, Bulla C (2019) Cloud monitoring system: a review. Int J Eng Sci Manage 1(1):49–55. A Multidisciplinary Publication of VTU
2. Sun Y, Xiao Z, Bao D, Zhao J (2019) An architecture model of management and monitoring on cloud services resources. In: 3rd International conference on advanced computer theory and engineering, ICACTE, vol 3. IEEE, pp V3–207–211
3. Ali A (2022) An overview of cloud computing for the advancement of the e-learning process. J Theor Appl Inf Technol 100(3)
4. Bulla C, Bhojannavar S, Danawade V (2013) Cloud computing: research activities and challenges. Int J Emerg Trends Technol Comput Sci
5. Alwadan T (2018) Cloud computing and multi-agent system: monitoring and services. J Theor Appl Inf Technol 96(09)
6. Meera A, Swamynathan S (2013) Agent based resource monitoring system in IaaS cloud environment. In: International conference on computational intelligence: modelling techniques and applications (CIMTA)
7. Ward JS, Barker A (2014) Observing the clouds: a survey and taxonomy of cloud monitoring. J Cloud Comput Adv Syst Appl. Springer
8. Cloud Computing (2019) National Institute of Standard and Technology, 05 Feb 2019. https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf
9. Aversa R, Tasquier L, Venticinque S (2017) Agents based monitoring of heterogeneous cloud infrastructures. In: 10th International conference on ubiquitous intelligence and computing and 2013 IEEE 10th international conference on autonomic and trusted computing. IEEE
10. Zubac DA, Kravchenko TV, Maiatin AV, Khagan MV (2018) A multi-agent approach to the monitoring of cloud computing system with dynamically changing configuration. In: FRUCT-ISPIT. IEEE
11. Birje MN, Bulla C (2019) Cloud monitoring system: basics, phases and challenges. Int J Recent Technol Eng (IJRTE) 8(3). ISSN: 2277-3878
12. Zhang C, He Y, Du B, Yuan L, Li B, Jiang S (2020) Transformer fault diagnosis method using IOT based monitoring system and ensemble machine learning. Futur Gener Comput Syst 108:533–545

13. Barot V, Kapadia V, Pandya S (2020) QoS enabled IOT based low-cost air quality monitoring system with power consumption optimization. Cybern Inf Technol 20(2):122–140
14. Alsubari SN, Deshmukh SN, Alqarni AA et al (2022) Data analytics for the identification of fake reviews using supervised learning. Comput Mater Continua 70(2):3189–3204
15. AlKhunzain A, Khan R (2021) The use of M-learning: a perspective of learners' perceptions on M-blackboard learn
16. Hu S (2022) Research on monitoring system of daily statistical indexes through big data. Recent Adv Comput Sci Commun 15(5):731–740
17. Klaver L et al. (2021) Towards independent runtime cloud monitoring. In: Proceedings of 2021 ACM/SPEC international conference on performance engineering (ICPE'21), Virtual Event, France, 19–23 Apr 2021. ACM, New York
18. Toczé K, Madon M, Garcia M, Lago P (2022) The dark side of cloud and edge computing: replication package. https://doi.org/10.5281/zenodo.6390610
19. Khalil MM et al. (2022) Measuring the effect of monitoring on a cloud computing system by estimating the delay time of requests. J King Saud Univ Comput Inf Sci 34(7)

A Secured Framework Against DDoS Attack in Wireless Networks

O. K. Vismaya, Ajay Kumar, Arya Paul, and Albins Paul

Abstract In the modern world, wireless networks are more prominent. Due to the dynamically changing nature, mobile networks are more vulnerable to several attacks. Data security is a significant problem since data flow across a wireless channel that is naturally exposed, making possible for malicious attackers to obtain sensitive data. Finding and resolving security problems in wireless networks are essential steps in ensuring secure data transmission. One of the main threats that can be performed against wireless networks to stop legitimate nodes from sending and receiving data packets is distributed denial of service (DDoS). The proposed framework includes implementation, detection and elimination of DDoS attacks in the wireless network. To provide further security, an authentication approach is introduced which will grand access/denial to the receiver based on successful authentication of the nodes. Finally, analysis of the proposed model is performed in 3 modes i.e., normal mode, attacker mode and controlled mode.

Keywords Distributed denial of service · Wireless networks · Authentication · Ad hoc on demand distance vector routing

O. K. Vismaya (✉) · A. Kumar · A. Paul · A. Paul
Adi Shankara Institute of Engineering and Technology, Vidya Bharati Nagar, Ernakulam, Kerala, India
e-mail: vismaya.ok@gmail.com

A. Kumar
e-mail: ajay.ec@adishankara.ac.in

A. Paul
e-mail: arya.ec@adishankara.ac.in

A. Paul
e-mail: albins.ec@adishankara.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_55

1 Introduction

Today, the internet has evolved into a need for everyday life and is utilized for more than just amusement. It also facilitates ordinary tasks like money transfers, bill payments, reservation of tickets, educational analysis, viewpoints on learning, commerce, media coverage, etc. There are primarily two fundamental varieties of networks: wired networks and wireless networks, and the main goal of computer networks is to exchange resources. In computer networks, nodes will make a connection with the use of cables like fiber-optic, twisted pair, and coaxial, to exchange data through a connection known as a data link. A wireless network (WLAN) allows access to the network without a cable connection because the network is set up via frequency signals. It refers to using a certain domain (range), such as WIFI, to access internet services without requiring a physical connection (stands for wireless fidelity). Desktops, mobile phones, and servers are examples of nodes or hosts in computer networks. Each has a special code known as a MAC address. Early on, diversity develops as a result of the market sales of switches, routers, and other equipment by many network manufacturers [1].

Depending on the need, a variety of wireless systems including WLAN, WPAN, WMN, and WSN are available. However, there is a significant security problem with wireless networks, particularly with regard to specific attacks that depend on the medium and are not present in the older counterpart. Ping-pong effect, MITM, exhaustion, collision, radio jamming, signal jamming, noise jamming, unfairness, PAN id conflict, capture, and tempering attacks are examples of traditional DDoS attacks that are conducted in the PHY and MAC layers[2]. The wireless media introduces a number of attacks that are difficult for standard protection techniques to effectively counter. Distributed denial-of-service (DDoS) assaults are a prominent class of these attacks that target system or user domain buffers. These attacks launched against wireless networks and will prevent legitimate nodes from transmitting and receiving. However, in the wireless world, attackers might be able to seize the communication channel and stop genuine nodes from communicating. Because wireless networks are built using an open medium, intruders have an easy way to implement such attacks. A straightforward DDoS assault can defeat wireless network defenses like cryptography and pass-phrase sharing, effectively shutting down the entire network. Thus, in order to prevent the effect of attacks, an efficient counter measurement system must be developed and make the network more secure[3, 4].

2 Related Works

Poongodi et al. [5] proposed a method to effectively protect data against a selective drop assault, this study provides a resistant to selective drop attack (RSDA) technique. The neighboring nodes won't faithfully pass the information to the neighboring nodes during a selective drop attack. A rogue node, however, that has placed in a data transmission route, has the ability to block certain forwarding messages. Malicious nodes that are overtaxing a host and rendering it completely unusable must be found. The hostile nodes would refuse to forward messages travelling through them in a selective drop attack. Finally, the assault reduced the host's throughput to its minimal value. Shivam Dhuria and Sachdeva [6] introduced a model to protect Wireless Sensor Networks in an efficient way. In this work, two approaches are presented. The first is a light-weight two-way authentication method that will shield WSNs from the majority of attacks, and the second is a traffic analysis-based data filtering method that will identify and shield WSNs from DDoS attacks. Several performance metrics have been used by the Network Simulator 2 (NS2) to verify the findings. i.e. throughput, delay, lost packets, energy consumption and PDR. Qazi et al. [7, 8] proposed a method to secure wireless Sensor Networks (WSNs). This study primarily targeted security challenges in WSNs. Using Elliptic Curve Digital Signature (ECDSA) algorithm, the introduced model not only secures the communication from one node to the other node network but also stores space of memory in one node to provide the correct method for calculating the time for generating key, number of hello message, and size of the packet. Additionally, key management with suitable key length is offered by the Algorithm for Wireless Secure Communication (ASCW). Additionally, ASCW aids in safeguarding node-level communication, aiding greater and more effective network security. With the aid of the authentication process, ASCW also lowers the cost of risk and security concerns on the network. Kshirsagar et al. [9] proposed a model for mobile ad hoc network (MANET). It is a mobile node communication network that lacks any preceding communication infrastructure[10]. The suggested technique is compatible with the current routing protocol. Secure communication paths use the trusted list idea. The trusted list and trust values display the number of times each node engaged in communication. The energy level of mobile components is used in MANET to distinguish between altruism and selfishness. Security and the distinction between selfishness and altruism are determined using the trust and energy models, respectively. Kolandaisamy et al. [11] introduced a method for detecting Distributed denial of service attack (DDoS). The approach determines the trustworthiness of the packet using all these variables and takes it into consideration while making decisions [12–14]. The proposed SPPA model bases its determination of the detection of a DDoS assault on the CCA's estimated value. Every single vehicle's behavior is evaluated, and the true weight of the car or node is determined. Every single vehicle's behavior is evaluated, and the true weight of the car or node is determined. The assault detection is carried out using this valid weight by determining whether the vehicle is an intruder or a regular node. Amrish et al. [15] proposed a solution to determine DDoS attacks. Cyberattacks of the Distributed Denial of Service (DDoS) variety aim to overload a target server in an effort to stop regular traffic. While being subjected to a DDoS assault, the system is still busy processing requests from bots rather than serving actual users. In this work, normal traffic and DDoS attack traffic are separated using a machine learning method. There are four machine learning classification approaches used to find DDoS attacks. The Artificial Neural Network (ANN) produces the greatest results when measured against KNN, Decision Tree, and Random Forest. Smys [16] introduces a model to detect DDoS attck in the communication network, which is made up of the terminal nodes. The telecommunications networks are vulnerable to a variety of cyberthreats, with distributed denial of service (DDOS) being the one that affects customers' access to services the most frequently. Therefore, the study offers a detection and classification approach for DDOS attacks in the communication system using a neural network and support vector machine combination. Performance evaluation of the proposed model is performed using NS2.

3 Proposed Model

A number of computer users have been increased dramatically and tremendously along with the interest in internet usage. The rapid increase of laptop computers and PDAs has increased the locations where individuals undertake computing tasks, including schools, colleges, business centres, and even homes. Everyone wants to join wireless networks because they provide users with mobility. While considering this type of network security, the message is a primary concern because there are many users in the network. The wireless networks are vulnerable to various attacks that can harm the network. Security of messages is a primary concern because there are a large number of users. Distributed denial of service is one of the main attacks in the network which can disrupt the services provided by the network. It is an active type of attack, which have the ability to make changes to the data while routing. The DDoS attack could be considered as one of the main threats to mobile networks. To provide security to the network, a secured framework is introduced in the wireless networks against DDoS attack. Block diagram for the proposed model is shown in Fig. 1.

In this proposed model DDoS attack will be detected and will get eliminated. In addition to eliminating attack from the network, an authentication mechanism is introduced which will authenticate the data and will ensure more security to the network. After authenticating the nodes, the data will be routed using the AODV protocol. Thus, as a whole the proposed model is divided in to four phases they are:

- Phase I: DDoS attack implementation
- Phase II: Attack detection
- Phase III: Attack elimination
- Phase IV: Authentication.

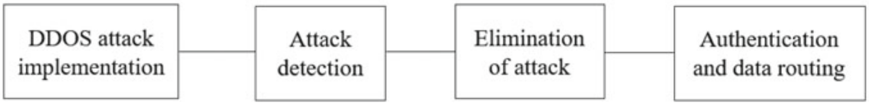


Fig. 1 Block diagram of proposed model

3.1 Phase I: DDoS Attack Implementation

The most well-known assaults used to bring down the entire network are DoS. By delivering malicious requests, DoS or DDoS attacks can fully deplete the network's resources. It is also a form of active attack that affects the network's availability. DDoS refers to DoS that can be performed by several nodes, increasing the number of network enemies. The attack model is employed in the wireless network to better understand the impacts of DDoS assault on a network. Initially, an attack model is implemented over the normal ad hoc on-demand distance vector. For the implementation of the first phase, 20 normal nodes and 5 attacker nodes are defined from the total of 25 nodes.

3.2 Phase II: Attack Detection

DoS attack detection is performed in the phase II. This detection is performed based on the rate of packets that are dropped from the network during data transmission. For each packet drop, route flag will become 1, which indicates the packet drop. At the same time, an initial trust value of zero is kept for each node. This trust value will get incremented for every positive flag 1 and the node which having the trust value above a particular threshold value will be considered as an attacker node. All the identified attacker nodes will be then saved to a text file.

3.3 Phase III: Attack Elimination

After detection of attack from the network, the detected attacker nodes will be eliminated from the network which is performed in the phase III. In this phase, initially the system will read the text files by which attacker nodes are saved. Then the system will exclude those attacker nodes from the network. In addition to that, there may be a chance of including the attacker nodes by default to the routing table. In such a case the system will again check the presence of attacker nodes from the routing table and will eliminate those identified attacker nodes from the network. In such a way the attacker nodes can be eliminated effectively from the network.

3.4 Phase IV: Authentication

To make the network more secure after the elimination of attacks from the network an authentication step is also introduced, which is considered phase IV. Proposed model for this authentication phase is shown in Fig. 2. In this phase, it includes both sender and a receiver. The sender will send an authentication message code. Then the message code will be converted to hash code using rotational hashing. After that a private key X will be generated using Elliptical Curve Digital Signature Algorithm (ECDSA). ECDSA will be using a sign digital signature value, they are two values r and s. Each node will be having a public and private key. The private key used in this model is X and public key is the combination of 3 values they are P, G and Y. P is a randomly generated prime number, G is a constant and Y is a value that is generated by using the private key. The public key Y is generated by finding the power of g and x. An authentication key is generated by using the public key and previously generated hash code. Then, using ECDSA algorithm generate the signature S and R. Finally compute $T1$, $T2$, and $T3$ using the digital signature, private key and hash code.

The receiver will then reconstruct the private key generated by the sender using the public key variables P, G and Y. Then compute TT1, TT2, ans TT3 using the signature S and R. Finally reconstruct the key and hash code. Finally compare the receiver side code (TT3) with the sender code (T3). If both the value is equal, then the authentication will be successful and the nodes will be permitted for communication otherwise the nodes will deny the communication. All the nodes in the network will participate in the communication.

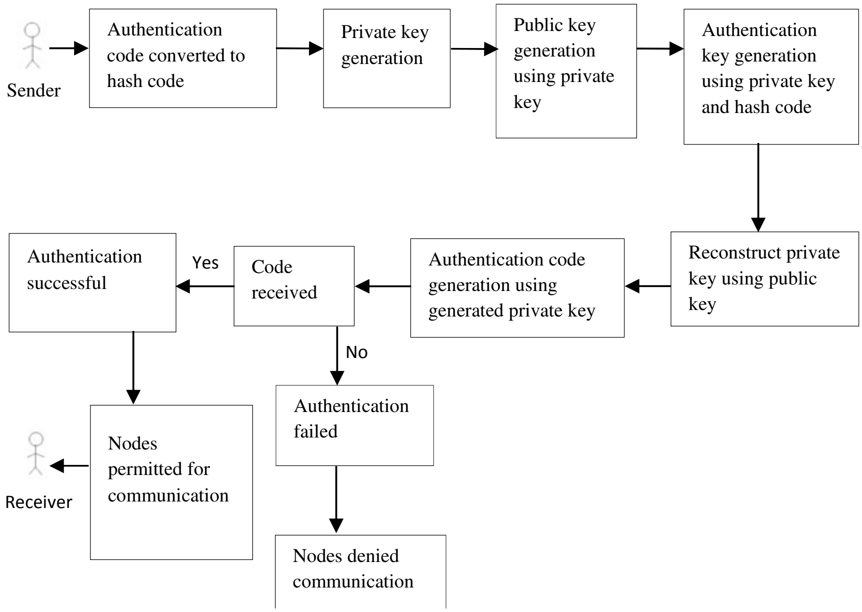

Sender
Authentication code converted to hash code
Private key generation
Public key generation using private key
Authentication key generation using private key and hash code
Reconstruct private key using public key
Authentication code generation using generated private key
Code received
Yes
Authentication successful
No
Authentication failed
Nodes permitted for communication
Receiver
Nodes denied communication

Fig. 2 Proposed secured framework

4 Result Analysis

In the proposed model DDOS attack is implemented in the wireless network. The implementation is performed in Network Simulator platform. There are certain parameters used for the implementation of attack and they are shown in table 1. Network Simulator simulates the proposed algorithm (NS-2). An open-source network simulation tool is called NS-2. İt will provide sufficient support for the simulation of different protocols in the wired and wireless networks. It will also provide sufficient support for different networking protocols, elements and routing types.

The proposed design topology uses 20 normal nodes and 5 attacker nodes are present. The entire mechanism makes use of 25 number of nodes. AODV is the protocol in use. To compare the findings, different parameters are taken into account. In NS-2, the 802.11 MAC layer is implemented for simulation. Link uses a wireless channel type with an omni-directional antenna type. The MAC type used in this simulation is 802.11. The simulation time is 10 s, and the routing protocol is the AODV protocol. For simulation, the two-ray ground propagation model is employed. Priority queueing is the sort of interface queue that is employed.

First phase of this model is to implement distributed denial of service attacker nodes in the network. the red colour nodes indicate the attacker nodes and the green colour nodes indicate the normal nodes. The attacker nodes are implemented continuously by sending repeated route requests to the network and make the node malicious. From the Fig. 3 as shown the nodes 4, 5, 14, 17, and 24 is set as the attacker nodes and the remaining nodes act as normal nodes. The dotted lines show the transmission of data during that particular time period and the circle shows the network coverage of that particular nodes.

The detected attacker nodes are displayed in the output terminal. The attacker nodes which cross the threshold value will be detected as an attacker. Then the detected attacker nodes will be excluded from the network. Moreover, there is a probability of adding the attacker nodes to the routing table due to shortest path

Table 1 Simulation parameters

Parameters	Value
Simulator	NS2 (Ver 2.35)
Normal nodes	20
Routing protocol	AODV
MAC	IEEE 802.11
Simulation time	10 s
Channel type	Wireless channel
Propagation	Two ray ground
Antenna type	Omnidirectional
Interface Queue Type	Priority queuing
No. of routing packets	12,000

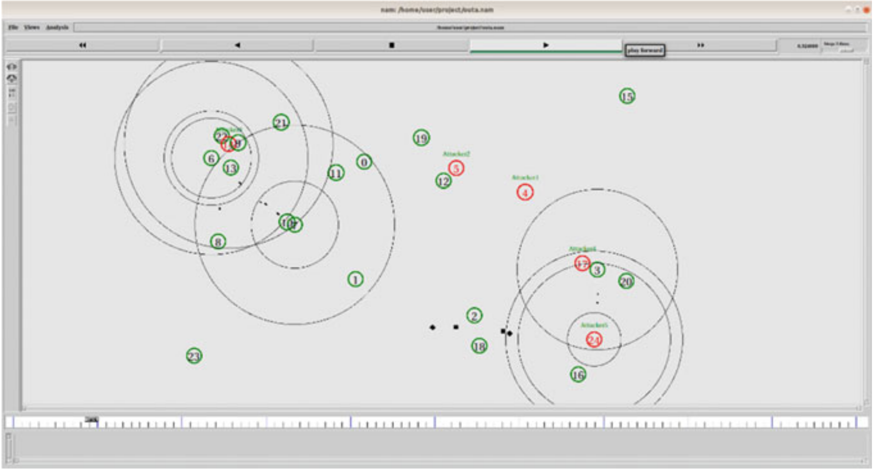


Fig. 3 Implementation of attacker nodes

finding property of AODV protocol. Due to this reason, the system will check the routing table again to verify whether any attacker nodes are added. Finally, the identified attacker nodes from the last iteration will be eliminated from the network. The hash code, private key, public key and the authentication key will be displayed in the output terminal. The hashing used in this proposed model is rotational hashing. The system will generate different hashing codes for each step and ensure more security to the network. Each node will be having its own private and public keys. The public key can be shared among others while the private key should be kept secretly. To analyze proper functioning of the proposed model certain parameters like packet loss, throughput, energy and end to end delay is evaluated:

- Total packet drop
- Total Throughput
- Average energy
- End to End delay.

The attacker mode will drop a large number of packets when compared to the attacker mode. Throughput is the maximum amount of successful data packet that can be received in the destination. Thus, the attacker throughput will be low as compared to the normal mode and controlled mode. Similarly, a large amount of energy will be consumed by the attacker node when compared to the normal and controlled mode. The end to end delay also will be less for the secured framework. The performance analysis is shown in Table 2.

Throughput performance of the system with normal, attacker mode and controlled mode is shown in the Fig. 4. In the normal mode i.e. without attacker, max number of packets will be received in the destination. In the attacker mode, the number of packets received will be less. While in the controlled mode the secured framework

Table 2 Performance analysis

S. No.	Parameters	Normal mode	Attacker mode	Controlled mode
1	Total packet drop	9397	10,110	5383
2	Total Throughput	2069	1297	2280
3	Average energy	1.89 J	10.04 J	1.54 J
4	End to end delay	0.38 s	0.63 s	0.19 s

will increase the rate of successful packet delivery. From the Fig. 4, it is clear that the designed framework will give a better performance compared with the other 2 modes. *x*-axis is plotted as time and *y*-axis is the number of throughput. From this it is clear that the system performance is low in the attacker mode.

Performance of the system in the normal, attacker and controlled mode based on the packet drop is shown in the Fig. 5. Packet drop will be more in the attacker mode. While the drop in the normal and contolled mode will be less compared to the attacker node. The system will detect the attacker node when performance of the system misbehaves from the normal mode.

Average delay for the normal, attacker and controlled mode is shown in Fig. 6. As the effect of attacks in the framework increases, the delay will also increases. Thus from Fig. 6, it is clear that the attacker effects in the framework is more copared to the normal and the controlled mode.

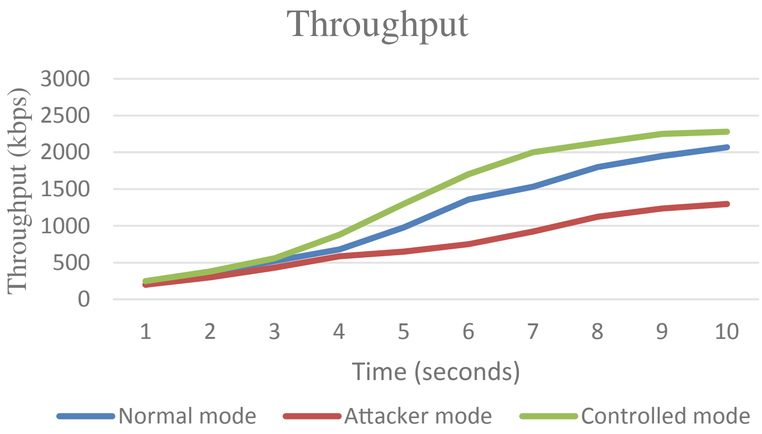


Fig. 4 Throughput graph for normal, attacker and controlled mode

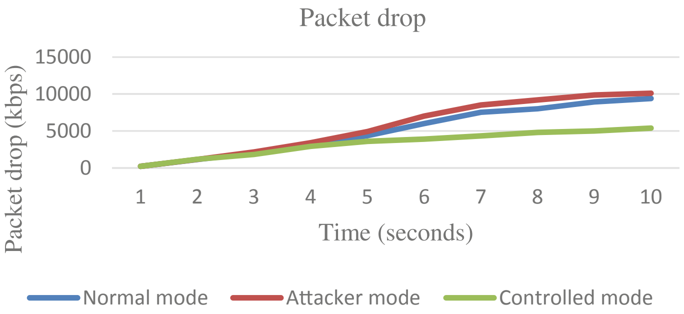


Fig. 5 Packet drop graph for normal, attacker and controlled mode

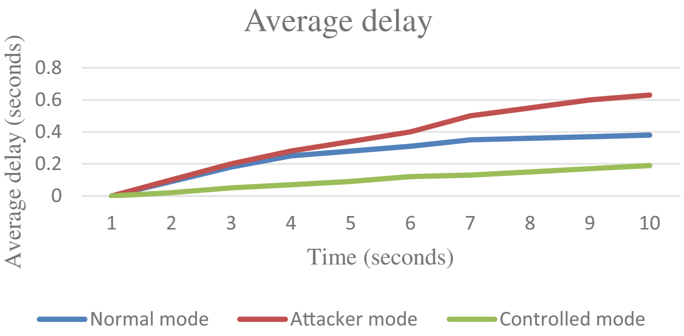


Fig. 6 Average delay graph for normal, attacker and controlled mode

5 Conclusion

Widespread deployments of numerous wireless networks of varying sizes, including wireless personal area networks (WPANs), local area networks (WLANs), metropolitan area networks (WMANs), and wide area networks, demonstrate the rapid progress that wireless networking has experienced (WWANs). These wireless networks can have many organizational structures, including cellular, ad hoc, and mesh networks. They can also be specialized networks, such sensor and vehicle communication networks. However, because the underlying communications are carried out via electromagnetic radiations in open space, wireless networks lack physical security. For those who work in computer and network security, wireless networks present a special difficulty. Data that is transmitted through a network during communication is never secure since various attackers have access to the data. The data are encrypted to transform them into an unreadable form since the data must be secured from the attack.

In future, this work can be extended from wireless networks to wireless sensor networks. The designed attacker topology can be implemented in other trust-based frameworks and can analyse the performance. Similarly, the designed secured framework can be implemented for multiple attacks.

References

1. Nazir R, Kumar K, David S, Ali M (2021) Survey on wireless network. Arch Comput Methods Eng 1–20
2. Bendale SP, Prasad JR (2018) Security threats and challenges in future mobile wireless networks. In: 2018 IEEE global conference on wireless computing and networking (GCWCN), pp 146–150
3. Balarengadurai C, Saraswathi S (2018) Comparative analysis of detection of DDoS attacks in IEEE 802.15. 4 low-rate wireless personal area network. Proc Eng 38:3855–3863
4. Sharma S, Mishra R, Singh K (2017) A review on wireless network security. İn: International conference on heterogeneous networking for quality, reliability, security and robustness, pp 668–681
5. Poongodi T, Khan MS, Patan R, Gandomi AH, Balusamy B (2019) Robust defense scheme against selective drop attack in wireless ad hoc networks. IEEE Access 7:18409–18419
6. Dhuria S, Sachdeva M (2018) Detection and prevention of DDoS attacks in wireless sensor networks. In: Networking communication and data knowledge engineering, pp 3–13
7. Qazi R, Qureshi KN, Bashir F, Islam NU, Iqbal S, Arshad A (2021) Security protocol using elliptic curve cryptography algorithm for wireless sensor networks. J Ambient Intell Humanized Comput 12(1):547–566 (2021)
8. Ali S, Humaria A (2020) An efficient cryptographic technique using modified Diffie–Hellman in wireless sensor networks. Int J Distrib Sens Netw 16(6):1550147720925772
9. Kshirsagar VH, Kanthe AM, Simunic D (2018) Trust based detection and elimination of packet drop attack in the mobile ad-hoc networks. Wireless Pers Commun 100(2):311–320
10. Cetinkaya A, Ishii H, Hayakawa T (2019) An overview on denial-of-service attacks in control systems: attack models and security analyses. Entropy 21(2):210 (2019)
11. Kolandaisamy R, Noor RM, Kolandaisamy I, Ahmedy IB, Kiah ML, Tamil EM, Nandy T (2021) A stream position performance analysis model based on DDoS attack detection for cluster-based routing in VANET. J Ambient Intell Humanized Comput 1–14
12. Riya soni1, Rajneesh Pachouri, Anurag Jain (2021) DDoS attack detection and prevention on wireless sensor network by using TBT method. Int Res J Eng Technol. p-ISSN: 2395-0072
13. Ghannam R, Sharevski F, Chung A (2018) User-targeted denial-of-service attacks in LTE mobile networks. In: 2018 14th International conference on wireless and mobile computing, networking and communications (WiMob), pp 1–8. IEEE
14. Bashirpour H, Bashirpour S, Shamshirband S, Chronopoulos AT (2018) An improved digital signature protocol to multi-user broadcast authentication based on elliptic curve cryptography in wireless sensor networks (WSNs). Math Computat Appl 23(2):17
15. Amrish R, Bavapriyan K, Gopinaath V, Jawahar A, Kumar CV (2022) DDoS detection using machine learning techniques. J IoT Soc Mob Analytics Cloud 4(1):24–32
16. Smys S (2019) DDOS attack detection in telecommunication network using machine learning. J Ubiquit Comput Commun Technol (UCCT) 1(01):33–44

Anomaly Based Intrusion Detection System Using Rule Based Genetic Algorithm

Shraddha R. Khonde

Abstract In emergent field of networks everyone is able to access data as required. Huge amount of data transmission is done on internet; so data security, integrity and confidentiality become important. Data Security is improved by use of intrusion detection system (IDS).This system allows administrator to monitor network to keep it secure from vulnerabilities. Most intrusions happen in the network are due to an attack. A model based on hybrid structure for intrusion detection system using rule based genetic algorithm for anomaly detection for new identified attacks is presented here. Model makes use of various algorithms of machine learning such as naïve bayes, support vector machine and random forest into ensemble. Ensemble approach helps in performance improvement. Behavior based detection is proposed using rule based genetic algorithm. Model is trained using UNSW NB15 dataset. This allow model to detect all types of modern attacks. Model overall performance is observed with 98.5% accuracy and 0.11% for false alarm.

Keywords Hybrid intrusion detection system · Ensemble · Random forest · Genetic algorithm · Naïve bayes · Support vector machine

1 Introduction

In era of virtual world an exponential growth of internet is marked. This is possible due to the reliability and availability provided by various internet providers now days. From last some decades a technical shift from traditional networks to multimedia networks is observed. Manan et al. [1] specifies the shift form text transmission to multimedia transmission. All the networks should be capable of transmission text, audio, video or any type of multimedia data. Due to increase in transmission of data security of data becomes utmost important issue in network security. Networks are transmitting data from one location to another, while this transmission data has to

S. R. Khonde (✉)
Department of Computer Science and Engineering, Modern Education Society's College of Engineering, SPPU, Pune, India
e-mail: khonde.shraddha@gmail.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*, Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_56

pass through various networks. All networks have their security system to protect data. These systems can be firewalls or intrusion detection systems (IDS). Every organization has its own security policies that are applicable to data transmitting through that network. But sometimes if network is not secure or vulnerable data transmitting through it can be misled or modified. To upgrade towards new data transmission era there is a need of system to provide security to data. This is the main reason why IDS is an important part of any network now days. An IDS is a system which monitors both inside and outside packets. IDS analyze all the communication inside the network to avoid all malicious activities. It also analyzes each packet entering into the system. IDS generate alarm and store the information about that malicious packet. This alarm is send to administrator so that further action can be taken to secure network.

IDS are mostly use for attack detection in networks. Nadiammai and Hemalatha [2] explain various data mining algorithms for attack detection. Almseidin et al. [3] provides evaluation of various algorithms of machine learning used for detection. Machine learning algorithms improve accuracy and detection rate of IDS. It helps to make IDS intelligent such that actions after attack detection can be taken by IDS. Deep learning can be considered as an emerging area in attack detection using IDS as proposed by Vinayakumar et al. [4]. IDS can perform attack detection using two important methods as explained by Butun [5]. Slight deviation in the usual behavior outline of the traffic can be deliberated as an anomaly attack. Alazab et al. [6] states that this type of IDS used to detect any type of novel attack even the web application or web services attacks.

2 Literature Survey

For identifying intrusions in several networks intrusion detection system is used. Most of the IDS make use of standard datasets and various algorithms from machine learning to complete their task. Some of the IDS also use data mining techniques to complete work of detection. Most of the researcher provides approaches to be used like distributed or collaborative IDS. As various types of attacks are known researchers use different algorithms to improve performance such that detection of all types of attacks can be done. To increase enactment of anomaly based IDS Aburomman and Reaz [7] elaborated various ML algorithms. Hybrid and ensemble approaches are used by IDS to improve detection performance. Many ML and DM algorithms used for IDS with comparison are elaborated. Discussion and review of algorithms in machine learning is elaborated by Buczak and Guven [8]. They further elaborate parameters which can be used like time complexity, algorithm complexity and accuracy. This will help to select efficient technique for IDS implementation. Feature selection role and methods are explained by Qassim et al. [9] such that most suitable subset can be found to increase efficiency of the IDS. Packet header based method is introduced to analyze and classify traffic. This method selects features subset using feature selection technique for monitoring and analyzing anomaly intrusions. Vimala et al. [10] elaborates various ML algorithms.

Ahmed et al. [11] provides survey of various anomaly detection techniques like statistical, clustering, classification and information theory. Feature reduction is done using combination of principal components analysis (PCA) and information gain (IG). This combination provides better result as compared to individual technique. This features are used with ensemble of SVM classifier, instance based learning algorithms and multilayer perceptron. Performance is tested on Kyoto2006 + along with other standard dataset as ISCX 2012 and NSL-KDD. Feng et al. [12] proposed a novel framework by combining SVM with ant-colony network. Kdd99 dataset is used for implementation in two parts as testing and training. Comparison of ensemble approach with individual classifier is elaborated. Results shows that ensemble approach gives better detection rate with reduce execution time. Li et al. [13] discussed a new hybrid approach to avoid bias detection of intrusion. Classifiers used are KNN clustering algorithm with binary classifiers. Binary classifiers are used to classify all known attacks. Remaining novel attacks are consider as outlier for cluster and handled by KNN clustering algorithm. Experiment performed on KDD dataset. Liu, et al. provides insight on ensemble approach [14].

Various approaches can be used for attack detection like deep learning [15–17], ensemble and cloud [18]. Neural network approach for detection of intrusions in network is mentioned by Wu et al. [19]. Various neural network algorithms are used by researchers now days for IDS. Authors make use of Computation Neural Network (CNN) for attack detection. IDS show better efficiency and good performance in terms of detection rate. Similar approach is use by Xiao, et al. [20]. They propose a method to convert traffic into image format. This will help to reduce computational cost of the system. Feature reduction is done using combination of encoding and principal component analysis. KDD99 dataset is use to perform experiments which results in high detection rate with reduced computational cost.

Li et al. [21] propose a novel auto-encoder IDS with random forest. Random forest machine learning algorithm provides good detection accuracy but fails when improper dataset is used. Proposed method use deep learning auto encoding technique to improve performance of RF and reduce training and testing time. This method use feature selection and feature grouping to reduce number of features require for training. Results shows less training time of proposed method as compare to traditional method.

IDS do not use single classifier for intrusion detection instead it make use of ensemble techniques to do so. In [22, 23] author makes use of feature selection technique with ensemble of classifier to improve performance of classifier. Experiment in both papers show that ensemble provides better performance as compare to individual classifier. Mukherjee and Sharma [24] explain importance of feature selection. Feature selection is done using information gain, correlation analysis and gain ratio. These three are compared with new novel method feature vitality based reduction.

From literature survey we can observe that various approaches and frameworks helps is performance improvement of intrusion detection. Most of the ensemble techniques are use with feature selection for reducing computation time. Detection accuracy of IDS is depends on dataset used by it. Various datasets and its features are available in market. Various datasets are reviewed by Ring et al. [25]; it makes use of various dataset properties to analyze it. Moustafa et al. [26] gives detailed knowledge of various datasets used in intrusion detection.

3 Methodology of Proposed System

The proposed model consists of an intrusion detection system capable of detecting all types of modern attacks. Modern IDS provides efficient performance and good detection rate. As most of the IDS works on signature based approach they were not able to detect modern attacks. Anomaly based detection for detection of unknown attacks is implemented in model. Anomaly detection is based on behavior checking of packets. Behavior analysis is done using genetic algorithm so that normal and malicious behavior of the system can be detected easily. If behavior is normal it is consider as a normal traffic and passed into the network. Otherwise if the behavior is malicious then packet is rejected by considering it as new attack. To check realism of model multiple classifiers are trained and tested individually along with ensemble approach. As per observations Ensemble approach gives good accuracy as compared to individual classifiers.

3.1 Ensemble Classifier

To test model for modern attack detection various techniques from machine learning are used. These techniques are basically used to provide better accuracy and detection rate. Machine leaning classifier is divided into multiple classes as unsupervised, semi-supervised and last is supervised. This classification is based on the leaning methods classifiers use to train themselves. Model makes use of supervised learning classifiers to test dataset on the real time traffic. Ensemble approach helps to gain accuracy and performance of detection. All the three classifiers are tested in ensemble approach. Research of many researchers proves that ensembling of classifiers shows good performance of system as compare to individual classifier. Ensembling is done with majority voting algorithm in model. In this algorithm votes from each classifier is taken and final decision is made accordingly. Votes are nothing but predictions generated by every classifier, once packets are analyzed for attack. Classifiers used are supervised learning classifier, which makes use of labeled data to train and test classifier. Classifiers used are SVM, naïve Bayes and RF.

3.2 *Anomaly Detection: Genetic Algorithm*

This algorithm does attack detection on the basis of behavior. System behavior is described using rules. Behavioral rules are demarcated using genetic algorithms. Genetic algorithm use behavior patterns to define rules in four step process. It will first generate initial population followed by chromosome designing. For every chromosome, fitness value is calculated to design genetic operator. Rule based dataset is formed using process of genetic algorithm. Output of this process is rule based dataset. Input used for this is standard dataset UNSW NB15 for intrusion detection. UNSW NB15 dataset is used to find the optimal features which are used to find final rules. Each feature set is used to calculate fitness value. The one having strong fitness value will be converted into rules to create dataset. Further classifiers are trained sing this dataset. Naïve Bayes, random forest and SVM classifiers are used. Genetic algorithm process to define rule based dataset is explained below.

3.2.1 Rule Based Dataset

Each packet entering in network will be observed and checked according to behavior rules used to train classifiers. Normal behavior of data is defined using rules. Any deviation of rules is observed as abnormal behavior. Behavior match with rule will be considered as normal behavior. To create a rule based database from the standard benchmark datasets genetic algorithm follows four steps as explained below.

Process basically starts with random selection of population from chromosomes. Chromosome is a problem need to be solved. According to various features of the problem, chromosomes are represented using bits, characters or numbers. These are represented at different position as required. These positions are randomly changed within a range of values and mostly known as genes. During a stage of process a set of chromosomes used is called population. As number of populations can be generated by randomly picking up genes from chromosomes an evaluation function is used to calculate its fitness.

Goodness of population is obtained using fitness value which shows the importance of that population. Crossover and Mutation helps in generating more specific rules. Genetic algorithm process is described in Fig. 1. Initial formulation of chromosome is followed by fitness function evaluation. More specific chromosome is obtained using crossover and mutation. Rules are formed out of this every chromosome to generate database. Abnormal activity can be detected using this dataset.

Process explained in Fig. 1 is a general process followed by genetic algorithms. The size of population, folds used for crossover and the rate of mutation based on benchmark datasets. This process helps in describing specific rules to define normal behavior of data. These rules are applied and used for training all the classifiers. Ensemble approach is used to find final prediction. Final prediction of the anomaly

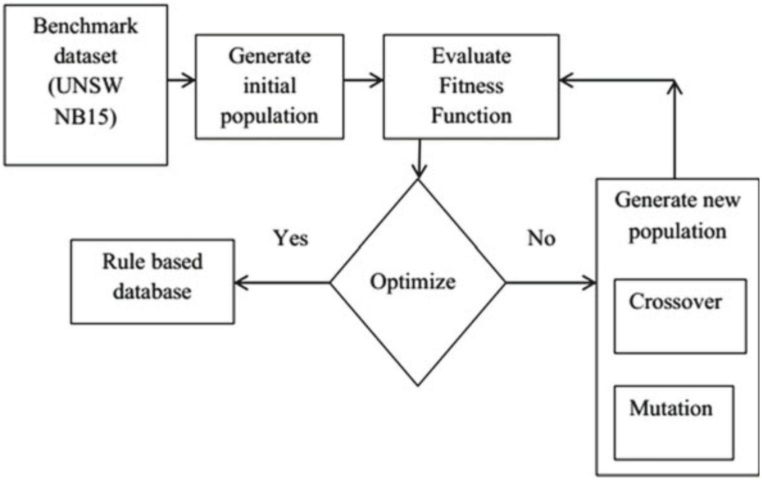

Benchmark dataset (UNSW NB15)
Generate initial population
Evaluate Fitness Function
Optimize
Yes
No
Rule based database
Generate new population
Crossover
Mutation

Fig. 1 Rule based database generation process using genetic algorithm

based detection engine will be given by majority voting. Rules from the database is represented as shown below.

If {condition} then {action} else {action}.

All rules are stored in above format in rule base dataset. Example of rule is given below.

If {connection consist of: source IP address 122.17.2.28; destination IP address: 138.22.106.45; destination port number: 15; connection time: 08 s} then {stop the connection} else {allow connection to continue}.

Explanation of rule is, if any connection request is coming in network from source IP address 122.17.2.28 having destination IP address as 138.22.106.45 with destination port number 15 and connection time required is 8 s then reject that transmission by stopping that connection. All the blocked IP addresses can be used in formation of rules. If above condition does not match then packets are consider as normal packets. Figure 2 shows the architecture of anomaly based detection engine.

For genetic algorithm crossover rate consider is 0.15 along with 0.35 as mutation rate. Fitness functions used is shown in Eq. 1 below which is used to calculate fitness value for each chromosome. Generations used are depending on types of attacks from dataset. So it can be 5, 7 or 10 in order to increase accuracy of detection. Reason behind number of generations is because number of datasets used for testing.

$$F = \frac{p}{P} - \frac{q}{Q} \tag{1}$$

where

p: count of attacks correctly classified

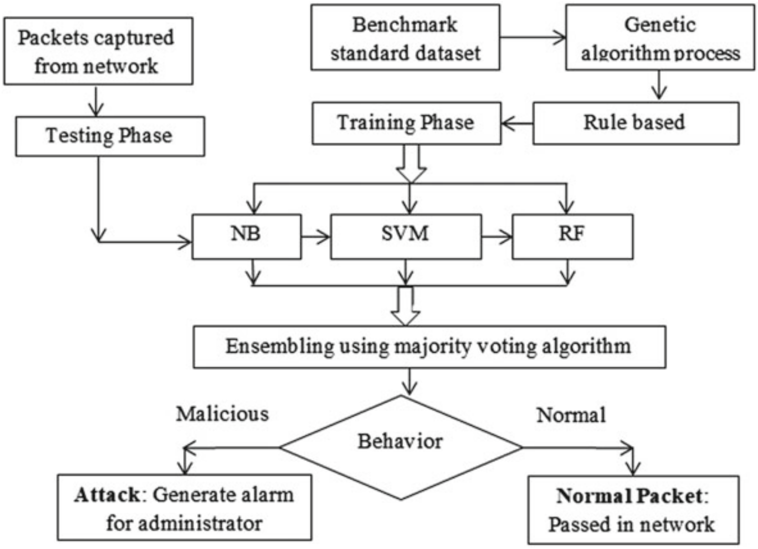

Packets captured from network
Testing Phase
Benchmark standard dataset
Genetic algorithm process
Rule based
Training Phase
NB
SVM
RF
Ensembling using majority voting algorithm
Malicious
Behavior
Normal
Attack: Generate alarm for administrator
Normal Packet: Passed in network

Fig. 2 Anomaly based detection engine

P: Count of attacks

q: normal packets correctly classified in count

Q: total number of packets in population.

Input passed is the standard dataset UNSW NB15 for rule generation. These dataset.csv files as passed as input to genetic algorithm to get rule based database. These rules are used to train classifiers in terms of behavior of data. Rules are used to define normal behavior of data. Any deviation will be observed as abnormal against the rule. This data is announced as attack. Upon detection of attack alert is generated for administrator. All packets observed as normal packets are transfer. An ensemble approach using majority voting is used to avoid biased prediction.

4 Results and Discussions

For experimental setup UNSW NB15 is used to train classifier. Dataset allows detecting many modern attacks. Classifiers are tested in real time where packets are captured using Wireshark network analyzer. All the packets from network are passed to handling and detecting attacks using anomaly based approach. IDS performance is evaluated using parameters as FAR, Accuracy and DR. All three classifiers are used to test dataset into real time environment. Various parameters used for testing is explained below.

Testing of all classifiers using UNSW NB15 dataset is done according to ensemble approach. Formula to calculate parameters accuracy, detection rate and false alarm rate is represented in Eqs. (2–4).

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{total}} \tag{2}$$

$$\text{DR} = \frac{\text{TP}}{\text{FN} + \text{TP}} \tag{3}$$

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \tag{4}$$

where

TP—Correct positive prediction.

FP—Incorrect positive prediction

TN—Correct negative prediction

FN—Incorrect negative prediction.

Model makes use of all parameters discuss above to evaluate performance of classifier. Classifier used is SVM, naïve bayes and RF. This all classifiers work as supervised learning classifier to do validation of dataset and evaluating performance of classifier. The classifier performances are compared according to all parameters and the best classifier is used for implementation of IDS. Performance of classifiers is check by training all with UNSW NB15 dataset to detect modern attacks. While testing for attacks confusion matrix is created such that parameters can be calculated according to dataset. Following section represents confusion matrix for dataset.

4.1 Performance Using UNSW NB15

This dataset is generated by cyber security research group at the Australian Centre for Cyber Security (ACCS) in 2015. This dataset consist of normal records with records of nine different type of attack. UNSW-NB15 is one of the most complex dataset as compare to other datasets. This dataset sets a benchmark in the area of NIDS for datasets. Total size of dataset is 100 GB. Pcap files of 100 MB are generated to provide reliable features in the dataset. Table 1 shows parameter evaluation over UNSW NB15 dataset. Graphical representation of results is shown in Fig. 3.

Table 1 Performance of model on accuracy, detection rate and false alarm rate

Attacks/parameters	Accuracy	Detection rate	False alarm rate
Normal	0.94	0.90	0.33
Generic	0.99	0.95	0.31
Exploits	0.89	0.80	0.38
Fuzzers	0.94	0.92	0.61
Reconnaissance	0.98	0.95	0.25
DoS	0.91	0.86	0.64
Analysis	0.96	0.92	0.91
Backdoor	0.97	0.91	0.60
Shellcode	0.92	0.85	0.27
Worms	0.98	0.95	0.20

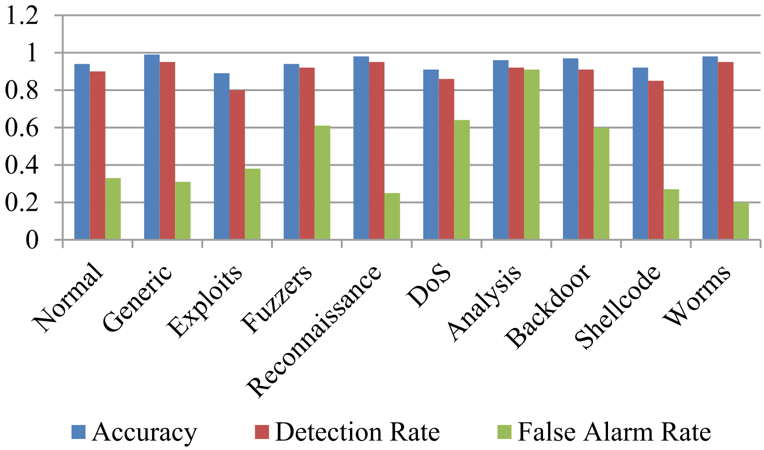


Fig. 3 Performance parameters for all classes over UNSW NB15 dataset

5 Conclusion

The presented model uses hybrid model of anomaly based detection technique and genetic algorithm for attack detection. Behavior based analysis is done using anomaly based detection based on genetic algorithm. Rules are generated using genetic algorithm during packet analysis. Prediction given by anomaly detection will be the final prediction. Model makes use of ensembling technique to increase accuracy of prediction and rate of detection along with reduce false rate. Model performance is checked using parameters as false positive rate, detection rate and accuracy. Testing of this model is done in real time environment. To detect modern era attacks model is tested using UNSW NB15 dataset. Overall performance of model is enhanced as compared to existing hybrid models. Limitation of this model is it will not detect attack hidden in the payload of protocol. In future work a new phase can be added to this model based on content analysis so that maximum type of attacks can be detected using this novel method. This model can also be provided as service to customers such that depend on network attack users can request any type of detection engine to used, signature, anomaly or content based detection engine.

References

1. MananJ, Ahmed A, Ullah I, Merghem-Boulahia L, Gaiti D (2019) Distributed intrusion detection scheme for next generation networks. J Netw Comput Appl 147
2. Nadiammai G, Hemalatha M (2014) Effective approach toward Intrusion detection system using data mining techniques. Egypt Inform J 15:37–50
3. Almseidin M, Alzudi M, Kovacs S, Alkasassbeh M (2017) Evaluation of machine learning algorithms for intrusion detection. In: 15th International symposium on intelligent systems and informatics, Subotica, Serbia, pp 14–16
4. Vinayakumar R, Alazab M, Soman K, Poornachandran P, Al-Nemrat A, Venkatraman S (2019) Deep learning approach for intelligent intrusion detection system. IEEE Access 7:14525–41550
5. Butun I, Morgera S, Sankar R (2014) A survey of intrusion detection systems in wireless sensor networks. IEEE Commun Surv Tutorials 16(1):266–282
6. Alazab A, Hobbs M, Abawajy J, Khraisat A, Alazab M (2014) Using response action with intelligent intrusion detection and prevention system against web application malware. Inf Manage Comput Secur, 22(5):431–449
7. Aburomman, Reaz M,"A survey of intrusion detection systems based on ensemble and hybrid classifiers. Comput Secur 65:135–152
8. Buczak, Guven E (2016) A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Commun Surv Tutorials 18(2):1153–1176
9. Qassim Q, Zin A, Aziz M (2016) Anomalies classification approach for network-based intrusion detection system. Int J Netw Secur 18(6):1159–1172
10. Vimala S, Khanaa V, Nalini C (2019) A study on supervised machine learning algorithm to improvise intrusion detection systems for mobile ad hoc networks. Clust Comput 22:4065–4074
11. Ahmed M, Mahmood AN, Hu J (2016) A survey of network anomaly detection techniques. J Netw Comput Appl 60:19–31
12. Feng W, Zhang Q, Hu G, Huang JX (2014) Mining network data for intrusion detection through combining svms with ant colony networks. Futur Gener Comput Syst 37:127–140
13. Li L, Yu Y, Bai S, Hou Y, Chen X (2017) An effective two-step intrusion detection approach based on binary classification and k-NN. IEEE Access 6:12060–12073
14. Liu J, He J, Zhang W, Ma T, Tang Z, Niyoyita JP, Gui W (2019) ANID-SEoKELM: adaptive network intrusion detection based on selective ensemble of kernel ELMs with random features. Knowl Based Syst 177:104–116
15. Khonde SR, Ulagamuthalvi V (2022) Blockchain: secured solution for signature transfer in distributed intrusion detection system. Comput Syst Sci Eng 40(1):37–51
16. Khonde SR, Ulagamuthalvi V (2022) Hybrid intrusion detection system using blockchain framework. Eurasip J Wirel Commun Netw 58
17. Ferrag MA, Maglaras L, Moschoyiannis S, Janicke H (2020) Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study. J Inf Secur Appl 50:102–419
18. Garg S, Kaur K, Batra S, Aujla GS, Morgan G, Kumar N, Zomaya AY, Ranjan R En-abc: an ensemble artificial bee colony based anomaly detection scheme for cloud environment. J Parallel Distrib Comput 135:219–233
19. Wu K, Chen Z, Li W (2018) A novel intrusion detection model for a massive network using convolutional neural networks. IEEE Access 6:50850–50859

20. Xiao Y, Xing C, Zhang T, Zhao Z (2019) An intrusion detection model based on feature reduction and convolutional neural networks. IEEE Access 7:42210–42219
21. Li X, Chen W, Zhang Q, Wu L (2020) Building auto-encoder intrusion detection system based on random forest feature selection. Comput Secur 95
22. Zhou Y, Cheng G, Jiang S, Dai M (2020) Building an efficient intrusion detection system based on feature selection and ensemble classifier. Comput Netw 174
23. Rajadurai H, Gandhi UD (2020) A stacked ensemble learning model for intrusion detection in wireless network. Neural Comput Appl
24. Mukherjee S, Sharma N (2012) Intrusion detection using naïve bayes classifier with feature reduction. Procedia Technol 4:119–128
25. Ring M, Wunderlich S, Scheuring D, Landes D, Hotho A (2019) A survey of network-based intrusion detection data sets. Crypt Secur
26. Moustafa N, Saly J (2015) UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network dataset). In: Military communications and information systems conference, pp 10–12

Hybrid Learning Approach for E-mail Spam Detection and Classification

Rimitha Shajahan and P. L. Lekshmy

Abstract Email is the most common channel through which cyber attackers commit crimes and initiate spamming attacks. Spamming is a popular method of sending unsolicited messages in order to distribute malware. The disadvantages of spamming include system slowdown, time consumption, and the presence of viruses. The main goal of this work is to identify spamming text, url, and message id and reduce the rate of spamming on e-mail. To reduce and control spam in e-mail, the text content, links, and header information in the corresponding e-mail can be analyzed. In the above scenarios, one of the main driving forces is the use of natural language processing to distinguish between spam and non-spam occurrences. The major goal of the proposed study is to consider three features—text, url, and message id—rather than deploying a single feature. The deep learning models, which play a significant role in predicting the spamming texts than any other traditional machine learning models. The machine learning models also play an important role in predicting the spamming urls and message id. Three standard datasets—Enron for text content, Phishtank for urls and SpamAssassin for Message Id are collected and processed. The aggregation of machine learning and deep learning models are an added advantage in this work. This identifies the spamming e-mail on the real-world datasets by using LSTM, Random Forest, Multinomial Naive Bayes models, which then evaluates the performance. Further, the output from the three models namely text, url, message id are integrated using weighted fusion approach and finally the output will be obtained as spam or ham.

Keywords Long short term memory · Random Forest · Term frequency-inverse document frequency · Bag of words

R. Shajahan (✉) · P. L. Lekshmy
LBS Institute of Technology for Women, Poojappura, Thiruvananthapuram, Kerala, India
e-mail: rimithashajahan@gmail.com

P. L. Lekshmy
e-mail: lekshmypl@lbsitw.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_57

1 Introduction

E-mail is a type of electronic communication in which data is stored on a computer and exchanged by establishing a telecommunication connection between two users. Email is a message, which include text, files, photos, or other attachments and is sent through a network to a single recipient or a group of recipients. Spamming is the way of transferring a large number of people several unwanted messages for the goal of commercial advertising, non-commercial proselytizing, any forbidden purpose, or simply delivering the same message to a huge number of people. Spam attacks are usually communicated via electronic communications includes texts and emails. Spamming has indeed demonstrated to be an effective attack, misleading a diverse group of people. Attackers frequently pose as popular social media sites, banks, IT managers, or popular commerce websites. These emails may persuade recipients to click on links that lead to malware downloads or to enter personal information on a malicious website that appears to be authentic. Spamming is a deceitful technique that uses social engineering and technology to get private information, including passwords and credit card details, by pretending to be a reliable individual or business in an electronic conversation. Spamming is the practise of sending fabricated emails that look real and come from reliable sources, including financial organisations or e-commerce websites, in order to convince recipients to visit shady websites by clicking on links included in the spam email.

Email spam has become a serious issue in recent years, and as internet users grow, so does the spam email population. They are being misused for dishonest and unethical activities like fraud and spamming. Sending spam emails with harmful links in them, which can compromise our system and access yours. Spammers can rapidly set up a bogus profile and an email account, and they can pose as real persons in their spam emails. The targets of spam are people who are not aware of the schemes. In addition to the annoyance and time lost sifting through unwanted emails, spam can seriously affect users' computers by infecting them with malicious software that can impair the system and steal personal information. It has the power to use up all available network resources. Millions of internet users are impacted by spam attacks, which are extremely expensive for businesses and spam recipients. Spam has grown to be a serious hazard to both individuals and businesses. In order to ensure the greatest possible level of human benefit and security, strive to detect spamming emails based on the text content, URL present in the email, and message-Id of the email.

2 Related Works

The review of literature focus on E-mail spam detection and classification using Machine Learning and Deep Learning techniques based on the three features—text content, url and message id of the e-mail.

2.1 Machine Learning Techniques

Joshi et al. [1] proposed an ensemble machine learning in order to detect fraudulent URLs communicated in emails. Blacklists are a popular and established approach of identifying dangerous URLs. The classification of URLs has been studied using machine learning and deep learning models that integrate static data such as lexical characteristics from the URL string, host information. It shows how to categorise URLs using only static lexical information from the URL string. In this investigation, the Crisp-DM Methodology was employed for analyse the presence of spam and ham urls. The classification using Random Forest model's findings with 92% accuracy outperforms than Naive Bayes, SVM, and Logistic Regression, ensemble classifiers like AdaBoost and Gradient Boost.

Sultana et al. [2] investigated that e-mail has been the most significant mode of communication in recent years and spam is a serious problem that is not only inconvenient for recipients. Kaggle's SMS spam was used as the dataset. The methodology presented in this work not only detects the spam term but also the system's IP address, enabling the spam message to be immediately recognised as banned the next time it is transmitted from that system based on the IP address.

Siddique et al. [3] focused on as the social communication has progressed that an e-mail has remained the most usual methods of formal and informal communication. Urdu, a South Asian language spoken largely in Pakistan, is gaining traction as a communication medium on social media platforms, websites, and emails. As the emails usage has grown, so has the volume and various types of spam messages in Urdu is observed. The python script Google-trans, which leverages the Google Translate Ajax API, was used to collect 5000 emails from the web resource 'kaggle' and translate them into Urdu. This work proposes leveraging existing machine learning techniques such as Naive Bayes, CNN, SVM, and LSTM to recognise and categorise e-mail content. LSTM model surpasses other models with accuracy score of 98.4%.

Patgiri et al. [4] proposed to do a detailed analysis into the detection of malicious URLs using Machine Learning methods. To extract these lexical properties, the URL is broken down into a series of words or tokens depending on the delimiters in the URL. The length of the URL is obtained by reading the URL as a string and instantly calculating the length of the string as an input to the Feature Extraction(URL) method. Host-based features are the URL's host-name properties are used to acquire host-based functionality. Since identifying malicious URLs is a binary classification problem, several machine learning algorithms, including Random Forests, SVMs, and Naive Bayes, are used to the training dataset. Furthermore, it has been found that for the given problem, the Random Forest classifier performs better than the SVM classifier.

Washha et al. [5] present as long as spammers continue to develop new ways and tactics for defeating and confusing email spam filtering systems. The header session messages, which are based on publicly available datasets, are utilised to demonstrate a useful and powerful email header feature. To confirm the effectiveness of the acquired header features in filtering spam and ham broadcasts, the impact of various machine learning-based classifiers on the retrieved header characteristics is again assessed and compared. Using the following classifiers: Random Forest (RF), C4.5 Decision Tree (J48), Voting Feature Intervals (VFI), Random Tree (RT), REPTree (REPT), Bayesian Network (BN), and Naive Bayes. Random forest outperforms than other classifiers and filter out the spam and ham header information.

2.2 *Deep Learning Techniques*

AbdulNabi and Yaseen [6] aims the work is to put state-of-the-art models to the test with the goal of detecting spam emails. Spambase data set from the UCI machine learning repository and Kaggle's open source Spam filter dataset are used. The best model, with an accuracy of 98.67% and an F1 score of 98.66%, is the bert-base-cased transformer model because it uses attention layers to take context into consideration. When compared to Keras word embedding, which assigns a different number to each word, Bert contextual word embedding enhances the capacity to identify spam emails. It was used in the BiLSTM model, which has an F1 score of 96% and accuracy of 96.43%. The results versus unknown data show the robustness and endurance of the fully fitted models.

Jain et al. [7] investigated that spam classification is a prominent topic in natural language processing, especially using the internet for social networking grows as the number of people increases. LSTM can learn abstract traits, unlike conventional classifiers that need manually created features. Before being input into the LSTM for classification, the text is transformed into semantic word vectors using word2vec, WordNet, and ConceptNet. The classification's outcomes are contrasted against benchmark classifiers including SVM, Naive Bayes, Artificial Neural Network, k-nearest Neighbor, and Random Forest. Two datasets—the SMS Spam Collection dataset and the Twitter dataset—are used to compare the outcomes. Results are assessed using the accuracy and F-measure. The outcomes show that LSTM can significantly outperform current machine learning methods for spam identification.

Bhuvaneshwari et al. [8] suggested that consumer reviews are a valuable source of information in the world of e-commerce. In this study, a deep learning (DL)-based novel framework for learning document level representation for recognising spam reviews as a replacement for ML-based detection is described. Using the Self Attention-based CNN Bi-LSTM (ACB) mechanism, the approach assesses the significance of each word in the phrase and scans the text for signs of spamming. The model then use a convolution neural network to learn phrase representation and extract higher-level n-gram features (CNN). In order to identify spam reviews with context, sentence vectors are concatenated as document feature vectors using Bi-directional LSTM (Bi-LSTM).

Fariska [9] proposed the study of the effect of using various combinations of pre-processing steps on some spam detection algorithms. They selected two spam detection classifiers that represented different approaches: Naïve Bayes and Support

Vector Machine. Pre-processing methods are differentiated in this study are: noise removal, stemming, lemmatization, and term frequency (TF-IDF). Each classifier's accuracy will increase if the proper pre-processing techniques are applied (or not applied) to it. For the Naive Bayes classifier, the stop words removal and stemming combination performs better than other combinations. The pre-processing stage, on the other hand, does not always result in improved classification results for Support Vector Machine (SVM) classifiers. Word shapes and the existence of stop words are sensitive to the probabilistic classifier—Naive Bayes classifier. SVM, on the other hand, does not require nearly all pre-processing methods because it is a non-probabilistic classifier.

3 Proposed Work

Detecting spam email is a difficult task because it is skillfully made to appear as real. Aside from attachments, an email primarily consists of a header, text body and links. In the proposed system, three features are used to detect and classify whether it is a spamming or a legitimate email. The proposed system makes use of Aggregated Model for E-mail Spam Detection and Classification. In this work, Enron dataset for text content, Phishtank dataset for URLs and SpamAssassin dataset for message id are utilized. First of all, read the data and then it will undergo the data pre-processing that contains removing noisy data, replacing missing data and removal of stopwords, punctuations etc. After that, feature extraction method is used for generating the relevant features and the input is given to the aggregated model. Thus detect and classify the email as spam or ham (Fig. 1).

3.1 Detecting Spam E-mail Based on Text Content

The proposed model for E-mail Spam Text Content Analysis contains the steps as follows data collection, pre-processing, feature extraction and classification and finally the spam text content indicated as 1 and ham text content indicated as 0:

Text Content Dataset: The data is collected from Enron Corpus dataset that is available in Kaggle. The dataset contains 5728 text contents that is about 4360 ham and 1360 spam text contents. About 500,000 emails generated by Enron Corporation employees were included in Enron email dataset. Real email is included in the dataset, which is openly accessible and can be used to improve the present email tools.

Dataset Preprocessing: After collecting the dataset, the preprocessing is performed. The available data is converted into lower case and punctuations are removed. The dataset includes the symbols (hash tags), @ (user mention), and rt (re-tweets), as well as the URLs http and https. After removing these symbols, word tokenization is carried out. Sentences are broken down into tokens through the process of tokenization.

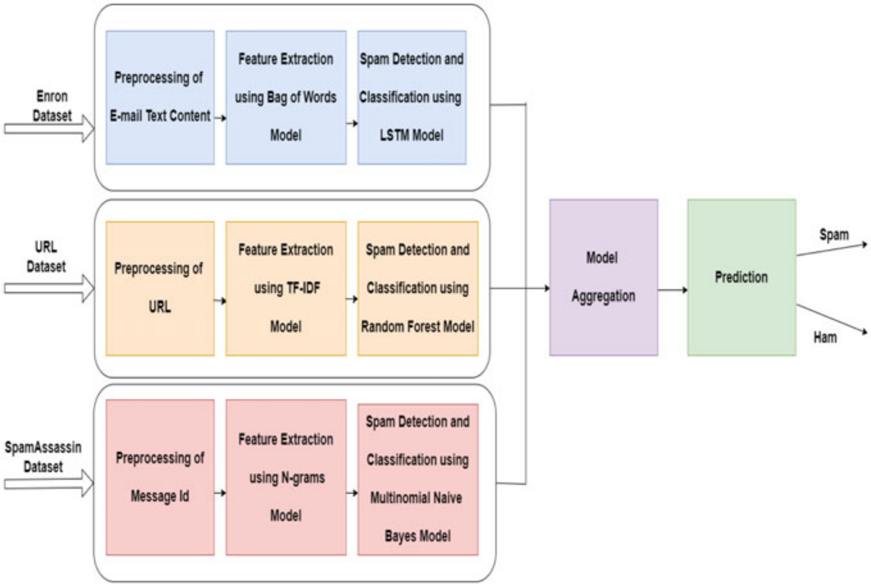

Enron Dataset
Preprocessing of E-mail Text Content
Feature Extraction using Bag of Words Model
Spam Detection and Classification using LSTM Model
URL Dataset
Preprocessing of URL
Feature Extraction using TF-IDF Model
Spam Detection and Classification using Random Forest Model
SpamAssassin Dataset
Preprocessing of Message Id
Feature Extraction using N-grams Model
Spam Detection and Classification using Multinomial Naive Bayes Model
Model Aggregation
Prediction
Spam
Ham

Fig. 1 Hybrid learning approach for spam e-mail detection and classification

Here, a considerable sample of text is divided into words using word tokenization. Stop words are finally eliminated and these are phrases that have no influence on the detection of email spam. Stop words removal is the procedure of getting rid of these stop words. The Natural Language Tool Kit has 179 stop words in English that is used in this work.

Feature Extraction using Bag of Words Model: The bag-of-words model is used to convert the text into numerical form that is available in natural language processing and information retrieval. Depict a sentence as a vector of words, much like the term itself. The most frequent words are obtained from the text and use a dictionary to hold the bag of words. Tokenize the text to words and then every word in text check whether the word exists in declared dictionary. If the word is present, increase the count by one and else hold the word in dictionary and set the count as 1. For building the bag of words, make a vector that will check a word in every text is a frequent word or not. If the word is a frequent word, mark it as 1 otherwise mark it as 0.

Classification using LSTM Model: After the feature extraction step, the input is fed into the deep learning model architecture for e-mail spam analysis on Enron dataset. The dataset is split into training for 70% and testing for 30%. LSTMs perform considerably better since they are skilled at remembering particular patterns. The important information is saved and all the irrelevant information is deleted in every single cell when the LSTM goes through each hidden layer, just like every other NN. The model consists of 2 simple rnn layer, two dropout layers, LSTM layer, 2 dense layer and elu layer and dropout layer. The Input layer takes sequence of inputs. Each word in the input post is transformed into a word embedding before being tokenized. The following layer receives these embedding vectors as input. All of the words in the training dataset will have an embedding learned by the Embedding layer, which starts out with random weights. The words are transformed into word embeddings, which are then transformed into lower-dimensional vectors [10].

Word embeddings are best used for deep learning and NLP applications since they are quick and effective and can capture context similarity due to their smaller dimensionality. Each word in the phrase has its word embeddings concatenated and supplied to the hidden layer. A predetermined size sequence of words is processed by the embedding layer. One drop out layer is given after the LSTM layer to reduce the overfitting. The dropout rate for the dropout layer is set as 0.2. Here the LSTM has 25 layers. From the input vectors, features are extracted. LSTM extracts features from embedding vectors. The network's dense layer 1 is the top layer, and it flattens and integrates the high-level information that the other layers have learned. The regularisation dropout layer is also a part of this layer. The Dense layer2 depends on number of classes here 2, spam and ham. Finally, the Activation function SoftMax is applied to take probability. The SoftMax output layer receives this layer's output for prediction that is spam text content indicated as 1 and ham text content indicated as 0.

Adam optimizer is one of the most efficient optimization algorithms used in the proposed model. It optimizes model performance mostly by reducing the cost function associated with the model. Adam optimizer is easy to implement, requires little memory space, performs well with huge data sets and large parameters, as well as situations with noisy or sparse gradients.

The lstm model is to choose from a wide range of LSTM parameters, including learning rates and input and output biases. Thus, there is no need for precise modifications. With LSTMs, updating each weight is simpler than with Back Propagation Through Time (BPTT), reducing complexity to O(1). Tried out with Bi-lstm model for the classification, it gives less accuracy and false positive rate is high and thus selected the lstm model for classification with high accuracy, less loss rate and false positive rate.

3.2 Detecting Spam E-mail Based on URL

The proposed model for E-mail Spam URL Analysis contains the steps as follows data collection, pre-processing, feature extraction and classification and finally the spam url indicated as 1 and ham url indicated as 0:

URL Dataset: Url dataset is collected from phishtank dataset and other datasets from github. The labeled dataset contains 10,000 urls that is 5000 for spam urls and 5000 for ham urls.

Data Preprocessing: In the data preprocessing step, remove the reduntant data and eliminate the missing data.

URL features: URL is splitted into protocol and domain address part. To extract the url features, the URL is broken down into a series of words or tokens depending on the delimiters in the URL and check whether the url having the below mentioned features:

1. **Presence of IP address**: Users can be sure that someone is attempting to steal their personal information if an IP address is used in the URL instead of the domain name, such as "http://125.98.3.123/fake.html."
2. **URL length**: Spammers can hide the suspicious portion of a URL in the address bar by using a lengthy URL.
   Shortening Service: On the "World Wide Web," a URL can be significantly shortened while still directing to the desired webpage by using this technique. This is done by using a "HTTP Redirect" on a short domain name that links to the full URL of the webpage. The URL "http://portal.hud.ac.uk/" can be condensed to "bit.ly/19DXSk4", for instance.
3. **Presence of @ symbol**: Using the @ sign in a URL causes the browser to ignore everything before the @ symbol, and the actual address frequently comes after the @ symbol.
4. **Double slash redirecting**: The visitor will be moved to another website if the URL path contains the character "//."
5. **Prefix-suffix**: Legitimate URLs rarely employ the dash symbol. In order to give users the impression that they are visiting a trustworthy website, spammers frequently append prefixes or suffixes to the domain name, separated by (–).

Feature Extraction using TF-IDF Model: TF-IDF method is used to identify the important term in the corpus and focus on that word for detecting the presence of spam and ham. This method makes tokens after splitting by double slash redirecting (//), dash (–), dot (.), remove .com and reduntant tokens.

A numerical measure called term frequency-inverse document frequency aims to show how significant a term is to a given document in a collection or corpus. A term's weight in a document is merely proportionate to how often it appears.

$$tf(t, d) = count\ of\ t\ in\ dnumber\ of\ words\ in\ d \tag{1}$$

Document Frequency, or df, is the quantity of times the term t appears in the document collection N.

$$df(t) = occurrence\ of\ t\ in\ documents \tag{2}$$

Inverse Document Frequency (IDF) is the dataset's size N divided by frequency of the text df(t).

$$idf(t) = log(N/df(t)) \tag{3}$$

tf-idf(t, d) indicates how important the term t in the whole document d, tf(t, d) indicates the number of times the term t occurs in the document d and idf(t) indicates the size of the corpus to the frequency of the text t.

$$tf - idf(t, d) = tf(t, d) * idf(t) \tag{4}$$

The TF-IDF value in this scenario is to range above 0.5 which considers the word that is more needed and below that range can be discarded. After the feature extraction using TF-IDF, get the feature vector and that will be passed as input for the classification using Random Forest. In order to account for the fact that some words are used more frequently than others, the relevance of the TF-IDF value rises according to the number of times a term appears in the text and is offset by the number of documents in the corpus that contain the word.

Classification using Random Forest: The dataset is divided into training for 70% and testing for 30%. It also marks the presence of each feature in the Url using 1 or 0 and TF-IDF value that will be given as input to the random forest model. Random forest is the model in which features having its own decision trees and using majority voting mechanism final output predicts as spam (1) or ham (0) [11, 12].

3.3 Detecting Spam E-mail Based on Message Id

The proposed model for E-mail Spam Message Id Analysis contains the steps as follows data collection, pre-processing, feature extraction and classification and finally the spam message id indicated as 1 and ham message id indicated as 0:

Message Id Dataset: SpamAssassin dataset is collected from Kaggle. The dataset contains 4008 message ids that is extracted from easy ham, hard ham and spam text files and contains 1399 spam and 2609 ham message ids. Spam message id is denoted using 1 and non-spam message id is denoted using 0.

Preprocessing of Message Id: Spam assasin dataset has easy ham, hard ham and spam dataset of message ID. From this easy ham and hard ham message ID are taken. LHS and RHS were separated using @ and data is labelled using 1 or 0. Two folders were created, one containing spamming message id and other containing non spamming message id (easy ham and hard ham).

Feature Extraction using N-gram Frequency: N-gram frequency (initially bi-gram) is done on the data using n-gram library from nltk package. It is a string or text that contains n characters in a row. For instance, in the word abc123, the character sequences for the 1st gram would be a, c, 2, etc. Similar to 2-grams, 2-grams are sequences of 2 characters that overlap, such as ab, bc, c1, 12, and 23. In a similar way, this concept can also be used to higher order n-grams [13].

Classification using Multinomial Naive Bayes: The dataset is divided into training for 70% and testing for 30%. Email filtering with naive Bayes classifiers is a common statistical technique. To determine if an email is spam or not, the Bayes theorem is applied after correlating token usage. The filter will modify the likelihoods that each word will appear in spam or valid email in its database for all terms in each training email. These word probabilities are used to determine whether an email that contains a specific set of words falls within one of the two categories—spam or ham message ids.

P(W/S) is the frequency of messages containing a specific term in the messages labelled as spam message id.

$$P(W|S) = P(W) * P(S|W)/P(S) \tag{5}$$

P(W/H) is the frequency of messages containing a specific term in the messages labelled as ham message id.

$$P(W|H) = P(W) * P(H|W)/P(H) \tag{6}$$

Each word in the email affects its likelihood of being spam, or merely the most intriguing words. The Bayes theorem is used to calculate this contribution, which is known as the posterior probability. The spam likelihood of the email is then calculated over all of its terms, and if the sum is higher than a predetermined threshold, the filter will classify the email as spam otherwise ham.

3.4 Model Aggregation

For the aggregated model, the three features—text content, url and message id corresponding datasets are combined into single dataset of about 4008 data and each feature uses its own data pre-processing and feature extraction steps. The dataset is split into training and testing datasets. The training is about 70% and testing about 30% of the datasets. Further it will detect and classify the email as spam or ham using machine learning and deep learning techniques.

The text content model is utilized the Enron dataset. When the e-mail text content is passed as an input to the text model. First, it will perform data pre-processing and then extracted the features using Bag of Words model. Finally, detection and classification is done using LSTM model and predicts the output as spam or ham text content. If text probability and predict is 1, the output predicted as spam text content else output as ham if text content text probability and predict is 0.

The URL model is utilized the Phishtank dataset. When the e-mail attached URL is passed as an input to the URL model. First, it will perform data pre-processing and then extracted the features using TF-IDF model. Finally, detection and classification is done using Random Forest model and predicts the output as spam or ham URLs. If URL probability and predict is 1, the output predicted as spam URL else output as ham URL if URL probability and predict is 0.

The Message Id is utilized the SpamAssassin dataset. When the e-mail header information—Message Id is passed as an input to the Message Id model. First, it will perform data pre-processing and then extracted the features using N-grams model.

Finally, detection and classification is done using Multinomial Naive Bayes model and predicts the output as spam or ham Message Ids. If Message Id probability and predict is 1, the output predicted as spam Message Id else output as ham Message Id if Message Id probability and predict is 0.

The three models gives an output that will be either spam (1) or ham (0). Finally the aggregated model is done on the corresponding result of the three features using weighted fusion model and thus final output is obtained. Weighted fusion model is firstly find the prediction probability and the weight of each feature. Then the probability of the aggregated model is calculated as

$$pb = \sum w_i * pb_i \tag{7}$$

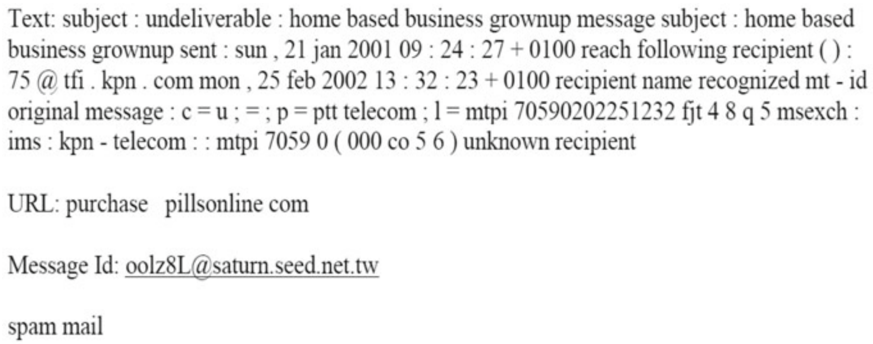
pb is the probability distribution of the model and wi is the average weight of the model and pbi is the probability distribution individual model. The probability distribution of the aggregated model will be generated and if the prediction is 1 (e-mail is spam) else the prediction is 0 (e-mail is ham).

4 Result Analysis

For the aggregated model, the three datasets corresponding to text content, url and message id into single dataset and each feature has its own data preprocessing and feature extraction steps. Further it will detect and classify the email as spam or ham using machine learning and deep learning techniques. The three models gives an output that will be either spam (1) or ham (0). Finally the aggregated model is done using weighted fusion model and final output is obtained.

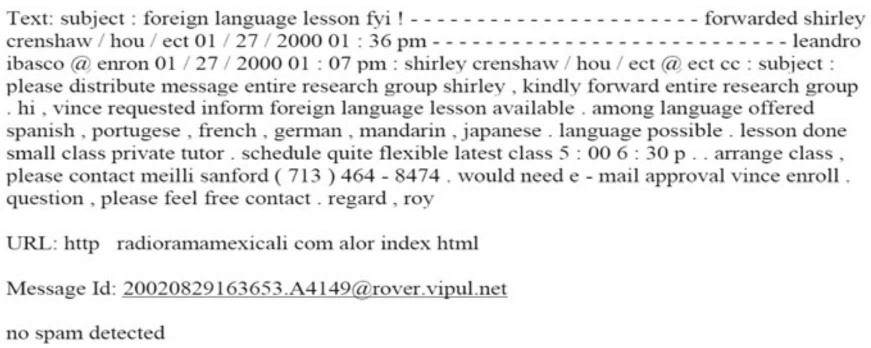
The sample output for spam aggregated model is shown in Fig. 2.

The sample output for ham aggregated model is shown in Fig. 3.



```
Text: subject : undeliverable : home based business grownup message subject : home based
business grownup sent : sun , 21 jan 2001 09 : 24 : 27 + 0100 reach following recipient ( ) :
75 @ tfi . kpn . com mon , 25 feb 2002 13 : 32 : 23 + 0100 recipient name recognized mt - id
original message : c = u ; = ; p = ptt telecom ; l = mtpi 70590202251232 fjt 4 8 q 5 msexch :
ims : kpn - telecom : : mtpi 7059 0 ( 000 co 5 6 ) unknown recipient

URL: purchase   pillsonline com

Message Id: oolz8L@saturn.seed.net.tw

spam mail
```

Fig. 2 Sample output of aggregated model indicates spam



```
Text: subject : foreign language lesson fyi ! - - - - - - - - - - - - - - - - - - - - - - forwarded shirley
crenshaw / hou / ect 01 / 27 / 2000 01 : 36 pm - - - - - - - - - - - - - - - - - - - - - - - - - - - - leandro
ibasco @ enron 01 / 27 / 2000 01 : 07 pm : shirley crenshaw / hou / ect @ ect cc : subject :
please distribute message entire research group shirley , kindly forward entire research group
. hi , vince requested inform foreign language lesson available . among language offered
spanish , portugese , french , german , mandarin , japanese . language possible . lesson done
small class private tutor . schedule quite flexible latest class 5 : 00 6 : 30 p . . arrange class ,
please contact meilli sanford ( 713 ) 464 - 8474 . would need e - mail approval vince enroll .
question , please feel free contact . regard , roy

URL: http   radioramamexicali com alor index html

Message Id: 20020829163653.A4149@rover.vipul.net

no spam detected
```

Fig. 3 Sample output of aggregated model indicates ham

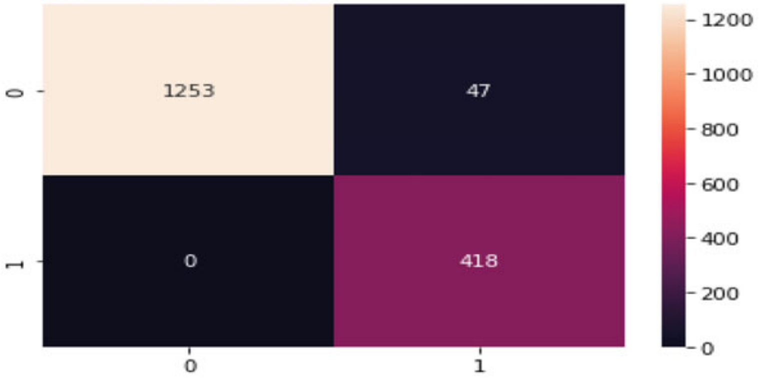
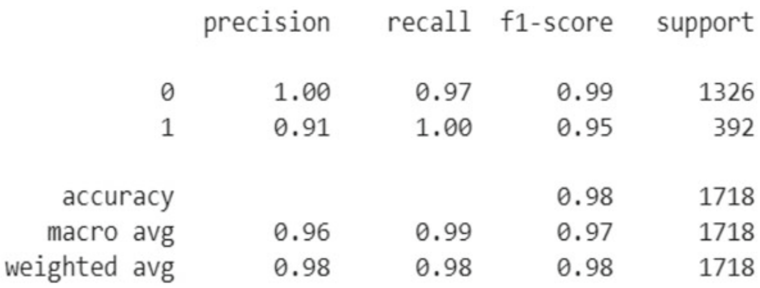


Fig. 4 Confusion matrix for aggregated model

A confusion matrix is used to evaluate the performance of a classification algorithm. The true positive is 1253 that means the output is correctly predicted ham. Here the true negative is 418 that means the output is correctly predicted spam. A false negative is an outcome where the model falsely predicts the ham class. Here the false negative is 47 that means the output is falsely predicted spam (Fig. 4).

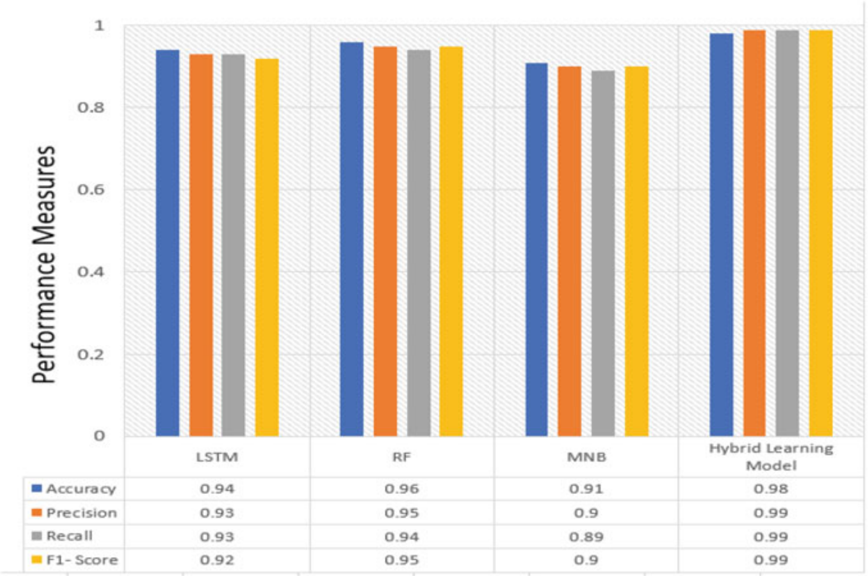
A Classification report is used to evaluate the rate of predictions from a classification algorithm. After classification, a classification report was generated to analyze the precision, recall, f1-score and accuracy of the model (Fig. 5).

The performance measures like precision, recall, f1-score and accuracy is evaluated for each individual models and aggregated model. The aggregated gives the better result with less false positive rate and correct prediction. Accuracy of the aggregated model is calculated as the total number of actual positives to the overall predictions. Precision of the aggregated model is calculated as the total number of actual positives to the total number of positives and number of negatives. Hybrid learning approach gives better accuracy as compared to the individual models for the prediction and evaluate the spam e-mail and ham e-mail (Fig. 6).



```
              precision    recall  f1-score   support

           0       1.00      0.97      0.99      1326
           1       0.91      1.00      0.95       392

    accuracy                           0.98      1718
   macro avg       0.96      0.99      0.97      1718
weighted avg       0.98      0.98      0.98      1718
```

Fig. 5 Classification report for aggregated model



	LSTM	RF	MNB	Hybrid Learning Model
■ Accuracy	0.94	0.96	0.91	0.98
■ Precision	0.93	0.95	0.9	0.99
■ Recall	0.93	0.94	0.89	0.99
■ F1- Score	0.92	0.95	0.9	0.99

Fig. 6 Comparison of individual models with aggregated model

5 Conclusion

Millions of internet users are affected by spam attacks, which are extremely expensive for businesses and spam recipients. Spam has recently been a major source of worry on social, economic, political and organizational levels, as it diminishes employee productivity and increases network traffic congestion. The various spamming operational modes serve as a reminder to focus on certain elements that could effectively identify spamming attempts. In order to solve the issue of spamming via email, a framework for successfully detecting spam has been suggested. It makes use of features that have been shown to be effective in the literature and produces high accuracy utilizing machine learning and deep learning approaches. By alerting users about emails that are associated with spamming, this framework assists users to avoid being spammed via emails. The aim to resolve this problem by proposing a system for the detection of spamming emails by considering three features for classifying an email as spam or non spam. The system focus on message id, message content and attached URL for detecting a spamming email using machine learning and deep learning techniques. The model is built for each of the three features and then integrated it for the final classification. Since the spammers are so skilled that they can easily fool the users by sending spam emails in a real way, the aggregated model helps the user to know whether the email is spam or not. Here False positive rate is reduced, accuracy is increased and chances of mis-classification of email is very rare.

References

1. Joshi A, Lloyd L, Westin P, Seethapathy S (2019) Using lexical features for malicious URL detection—a machine learning approach. ArXiv http://orcid.org/abs/1910.06277
2. Sultana T, Sapnaz KA, Sana F, Najath J (2020) E-mail based spam detection. Int J Eng Res Technol (IJERT) 9(06)
3. Siddique ZB, Khan MA, Din IU, Almogren A, Mohiuddin I, Nazir S (2021) Machine learning-based detection of spam e-mails. Article ID 6508784. Hindawi. https://doi.org/10.1155/2021/6508784
4. Patgiri R, Katari H, Kumar R, Sharma D (2019) Empirical study on malicious URL detection using machine learning. In: International conference on distributed computing and internet technology ICDCIT 2019: distributed computing and internet technology, pp 380–388
5. Washaha M, Khater IM, Qaroush A (2012) Identifying spam e-mail based-on statistical header features and sender behavior. In: International information technology conference and exhibition (CUBE), September 2012, Pune, India
6. AbdulNabi I, Yaseen Q (2021) Spam detection using deep learning techniques. In: The 2nd international workshop on data-driven security (DDSW 2021), March 23–26, Warsaw, Poland
7. Jain G, Sharma M, Agarwal B (2018) Optimizing semantic LSTM for spam detection. Int J Inf Technol 11(3)
8. Bhuvaneshwari P, Rao AN, Robinson YH (2021) Spam review detection using self attention based CNN and bi-directional LSTM. Multimed Tools Appl 80:18107–18124 . https://doi.org/10.1007/s11042-021-10602-y
9. Fariska R (2019) Study on the effect of preprocessing methods for spam email detection. Indonesian J Comput (Indo-JC)
10. Rahman SE, Ullah S (2020) Email spam detection using bidirectional long short-term memory with convolutional neural network. In: 2020 IEEE region 10 symposium (TENSYMP). IEEE, pp 1307–1311
11. Govil N, Agarwal K, Bansal A, Varshney A (2020) A machine learning based spam detection mechanism. In: 2020 Fourth international conference on computing methodologies and communication (ICCMC). IEEE, pp 954–957
12. Chen J-Y, Wang Y-J (2022) Semi-supervised fake reviews detection based on AspamGAN. J Artif Intell 4(1):17–36
13. Balasubramaniam V (2021) Design of associate content based classifier for malicious URL prediction by rule generation algorithm. J Inf Technol Digit World 3(1):44–56

Smart Solid Waste Management System Using IoT Technology: Comparative Analysis, Gaps, and Challenges

Meenakshi Shruti Pal and Munish Bhatia

Abstract With the consistent spike in the world's population and continued expansion of urban cities, waste is a highly visible city problem for every prominent stakeholder including the respective State Government of the territory. Waste management in the cities has big expenditures based on the method of waste collection and disposable systems that are opted by a particular city. There are several ways to enhance the waste collection mechanism, however, one of the most appropriate systems with appropriate use of technology is key to getting the waste management system right and making it commercially viable. The latest entrant to smart city waste management is the Internet of Things (IoT). The smart bins comprising IoT sensors and route optimization improve traditional waste management processes in the most efficient manner. Therefore, this paper reviews various IoT-based solutions used for waste collection and route optimization methods for garbage collector vehicles in the past decade. The objective of the paper is to study and analyze the state-of-the-art techniques used for the waste management system with smart IoT-enabled bins. In this paper, analyses of research papers are described in the literature. Research gaps from an existing work have been concluded based on the results of the study. Further, this paper also describes the various challenges and issues of the smart waste management system. This thus, calls for further improvement and innovation.

Keywords Waste management · Internet of Things · Smart bins · Route optimization

M. S. Pal (✉) · M. Bhatia
Lovely Professional University, Phagwara, Punjab, India
e-mail: er.meenakshi23@mail.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_58

1 Introduction

In the digital space, the Internet of Things (IoT) has witnessed a major shift in the last couple of years and it would continue to evolve itself. The IoT is the next frontier with wireless networks, smart sensors and through-going computing capabilities. The idea of IoT is just not about connecting the equipment to the internet altogether but also about sensing the "things" in a smart and real way. The basic principle of IoT is anything, anytime, anywhere, and any network [1, 2]. There is huge growth in the IoT sector and its noteworthy impact on our day-to-day life. Waste management is a core area that is addressed through the IoT in smart cities [3]. According to the studies done globally, it is estimated that the waste produced in the cities will be around 3.5 billion tons by the year 2050. Also, the cost of management of this kind of waste will be around 640 billion dollars which is a huge amount as per the current scenario and the economic situation in this world. There is no doubt that this problem is not going to solve itself and it is a big worry for modern cities all around the world many innovative ways need to be designed to ensure that this problem is solved. Traditional and unscientific ways of Municipal Solid Waste Management (MSWM) impact the environment and hence causes various health hazards to inhabitants [4, 5]. An inadequate solution to waste collection and poor transportation is liable for the accretion of municipal solid waste at every corner and junction [6]. To counter the problems, only the solutions which are related to it in the management domain are not enough and technological solutions for efficient systems are needed, which can enhance the overall efficiency of the waste management system in cities [7, 8]. Since it is necessary to monitor the fill level of waste inside the garbage bins to avoid overflow of bins which creates an unhygienic environment and hence adverse impact on human health. Therefore, smart bins equipped with sensors are being used to monitor the level of garbage inside the smart bins. In a smart city, for the solid waste management system (SWMS) such IoT-enabled smart bins are placed in various places to monitor the real-time status of the fill level of bins. By using a wireless network IoT enabled bins shall capable of transmitting real-time alerts to the command and control center as and when smart bins will be reaching at threshold or full level. The fill level of the bin is detected using ultrasonic sensing. By monitoring the amount of time that passes between delivering and receiving an ultrasonic pulse, the sensor can calculate the distance to a target. When an object is present, the sensor receives a bounced-back ultrasonic pulse that was originally sent out into the air. The distance is calculated using the travel time and sound speed. The garbage collector vehicle will collect garbage from smart bins for treatment/disposal at the dumping site. Figure 1 depicts an end-to-end solution for an SWMS. Without enhancing the routing of garbage collection trucks, we cannot assume the effectiveness of the SWMS [9]. The Optimized and shortest route for collecting waste is predicted by using various algorithms such as greedy algorithms, Digikshitra algorithms, and traveling salesman problems. The Optimized route is based on various parameters such as the number of bins, filling rate, current status, and distance, which depends upon the city to city. Further, IoT-enabled solutions such as Lower-Power Wide Area Network (LPWAN)

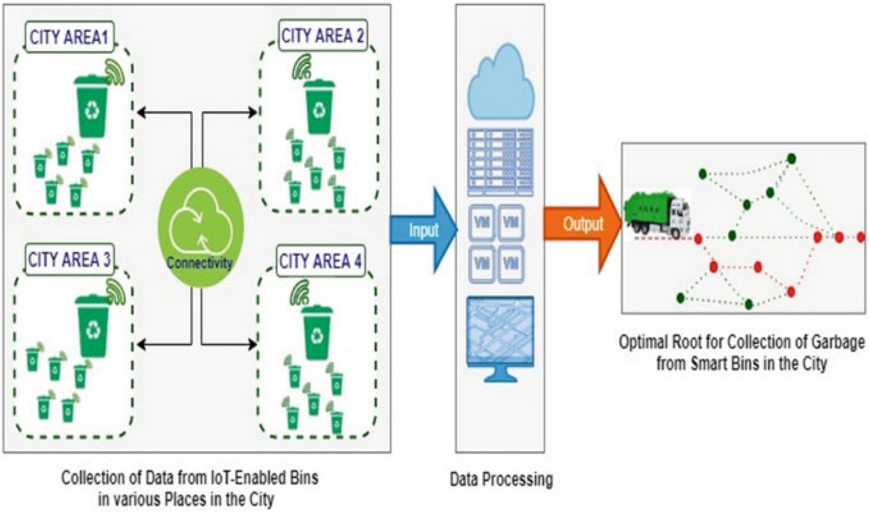

CITY AREA1
CITY AREA 2
CITY AREA 3
CITY AREA 4
Connectivity
Input
VM
Output
Optimal Root for Collection of Garbage from Smart Bins in the City
Collection of Data from IoT-Enabled Bins in various Places in the City
Data Processing

Fig. 1 End-to-end solid waste management system

leveraged low power means long battery life for the device, low bandwidth, and long range could excel as SWMS in smart cities [10]. Hence, this paper brings together the different techniques used for the SWMS with smart IoT-enabled bins by performing reviews on existing research that catalysis end-to-end solutions for the SWMS.

1.1 Paper Organization

The paper initiated a basic introduction to IoT and the SWMS. The rest of the paper is outlined in various sections. Section 2, the paper provides a brief overview of the relative and the previous work followed by the comparison of various techniques and algorithms. Section 3, describes various challenges and issues for the Smart Waste Management system. Section 4, covers critical discussion. Finally, Sect. 5 concludes all the review-based research that has been carried out and, the future direction thereof.

2 Analysis of Previous Works

As IoT proves a trustworthy one, for SWMS by resolving the issues like accumulating data, processing it, and outputting the result by effectively using certain communication protocols. Data is collected from the real-time sensors. In the data processing phase, data mining and classification techniques are used to process the data into useful information. The output unit is used to display the results. This section deals with previous works that emphasize IoT-enabled solutions with various routing techniques used for the smart waste management system. The series of recent studies have been categorized in the following three subsections.

2.1 IoT-Enabled Smart Bin

The authors in [11], proposed a solution by implementing the sensor-based bins in which the ultrasonic sensors are used to measure the emptiness of the bin and use the LoRaWAN protocol for data collection. The conducted study was a pilot study with outdoor bins and the results show that the existing technology needs improvement to implement this kind of system with inexpensive equipment. This system was not an actual implementation but a prototype to explore the options of placing smart IoT-enabled bins at the university campus. This proposed system did not use any particular algorithm but was just set up to test this system in a real-time environment. The authors presented [12] a system that assists the user in the correct way of sorting and disposing of waste and also conveyed information about the content of each bin. This model has two phases: In the first phase, a smart bin automatically computes the weight of garbage, changes it into reward points, and credits those points into the card. The second phase focus to present the gift to peoples, who abandon their garbage correctly. The gathered points can be exchanged for an item or even withdrawn by the banking system in form of virtual currency. A solution is suggested by the authors [13] to the problem of garbage containers that are not cleaned timely and resultant overflow. The system is embedded with an alarm that triggers when a garbage container is nearby to fill. Further bins are associated with the NIR (near-infrared spectroscopy), a screening system, that enables the identification of the different five types of plastic resin, which are non-biodegradable. Using NIR reflectance spectroscopy to identify and classify plastic resin has many benefits. There are particular NIR wavelengths used to identify the different Plastic resins. Hence, biodegradable waste is easily classified and sent to the production of biogas which is one of the best methods to dispose of the waste.

This paper [14] shows that every time information is collected on the use of garbage bins it helps the administrator to detect and decide if a particular place needs extra dumps or to move them to the other required areas. The contractor/bin provider could plan in a better way from daily updated information that when and where need to send a garbage collector vehicle for the emptiness of the filled bins. A “Smart bin” solution is suggested where smart bins are distributed in different places with a unique ID [15]. When bins are about to fill the bin ID and location are communicated through GSM (global system for mobile communication) to the concerned person who is responsible to clean that area.

2.2 Smart Waste Management in the Context of Smart Cities

The research paper [16], focused on the smart management of waste using the IoT and an optimized architecture to manage the waste in the perspective of smart cities where the particular situation of access waste is prevailing. The particular study delved into the quality of typology of the sensor node, based on the cost and the power consumption has to be minimum. The sensing node used in the system consists of a small microcontroller, a sensor used to check the filling up of the status of the bin, and a long-range transmission device based on the LoRa LPWAN technology to look after the overall waste collection process. LoRa LPWAN is a sort of wireless wide-area network created to enable long-distance communications at a low bit rate among the connected things such as sensors operated on a battery. The paper especially analyzed the architecture of the nodes, energy savings during the transmission, and the policies to increase the network longevity. The main purpose was to increase the battery life because the sensors being wireless, need to be energy conserved through hardware and software optimization to increase long working time. The study presented in the paper [17], focused on the Governments and corporations which are looking for solutions to increase the efficiency of garbage collection in cities through the latest technologies. The things used are smart sensors with the IoT and other cloud platforms etc. The proposed method is used to turn the automated garbage collection mechanism into a smart system so that it can be easily included in any smart city setup. Intelligent monitoring is proposed [18] in two phases: the first phase is to monitor the level of waste in the smart bin regularly and in the second phase transportation of the waste through the optimal route. All simulation work is done by the smart-M2 platform, which is an extension of cross-domain search, and through this, it is possible to interoperate applications from different areas. This paper [19] described a cloud-based solution for the waste management system, in which the current status of dustbins is sent to the cloud so that the stockholders may fetch relevant information according to their interests. This solution provides a smart way of handling and disposal waste. There are various intelligent approaches for waste collection and disposal [20]; which can be applied to smart waste management and hence beneficial in terms of cost and environmentally [21, 22].

2.3 Smart Garbage Monitoring and Route Optimizing System

Various routing algorithms are used for route optimization of the garbage collection vehicles such as the Mathematical approach, GIS-based, and combined optimized approach. Authors in their work [23], proposed an IoT-based smart garbage management system for the city where the daily task of collecting the generated garbage has to be done. Huge resources are committed to this mechanism but the results are still too vague. Recently we are moving for hybrid approaches with a minimum cost of IoT architecture which presents a better outcome. This paper presented an effective way for waste management by predicting the filled level of bins. The optimized and shortest route for collecting the waste is predicted by using machine learning and graph theory. The proposed system examined the data transfer on long-range transmission sensors such as the LoRa module and was implemented at Ton Duc Thang University (Vietnam), the results show the system saves lots of time by finding the shortest or best path for waste collection. This paper [24] presented a smart bin for monitoring the garbage and dynamic scheduling for a vehicle to collect the garbage in a smart city. In this approach status of the bin along with the location is sent to the cloud platform. The proposed method also leverages the facility on mobile for the driver of the garbage collector vehicle to follow the shortest path from his present location to the filled bin. The visualization of the gathered data is shown on the ThinkSpeak platform. Data can be collected, visualise, and analyse in real-time data streams in the cloud with ThingSpeak. It is simple to set up devices to communicate with ThingSpeak using well-liked IoT protocols. The author suggested making this system eco-friendly by installing a solar panel for energy to devices which is a truly renewable energy source. The author [25], focused on the retrieved nature of the senses and intentionally focused on the possibility collection algorithm works in the direction of optimizing the waste collection process to automatically retrieve bins with the help of the IoT and the optimized approach for monitoring. Various comprehensive simulations during the study prove that the algorithm proposed is better in terms of quality and quantity criteria which are adapted to analyze the strength and weaknesses of the particular algorithm that is proposed in this work. This research paper [26], addressed that waste management is a multidisciplinary activity that contains the generation, collection, storage, and transportation of waste. The latest technologies and diffusion of the internet together with the advent of compact hardware has allowed the development of efficient waste management system. This study moves towards the coupling of waste collection with the help of GIS web-oriented systems.

This article [27], described a different mechanism developed to calculate the efficiency and viability of the system implementation for waste management. A prototype is proposed for wireless nodes using LoRa-WAN to sense the temperature, weight, and level of the bins. The gathered information is analyzed and generates the dynamic route. The result of the case study based on the region of Salamanca shows that the optimized route provides minimum cost, time, and workforce as compared to a static waste collection route. This study shows a developed node has a great operational lifespan, cost, and long coverage with the installation of antennas in the specified region. Another inline procedure in the paper [28] utilizes the Particle Swarm Optimisation which is incorporated with the capacitated vehicle routing problem to determine the best possible route for the waste collection and also make sure that the efficiency is above 75% all the time. Authors in the paper [29] proposed an innovative system that efficiently collects the waste, detects the fire in the waste, and also predicts waste generation. The IoT-enabled devices perform the monitoring of the smart bin and the wireless devices are used to send the status of the filling level of the bins along with its location to the cloud for processing. Two predictive analytic techniques such as neural networks and decision trees are used to predict the future generated waste. The proposed system is also enabled for fire detection which is very useful to save human life and economic loss. This study focused on the cost and efficiency of the system. The authors [30] explained modern traceability devices such as volumetric sensors, RFID (Radio Frequency Identification) system, GPRS (General Packet Radio Service), and GPS (Global Positioning System) technology are used to receive the real-time data that is basic to develop an effective routing model for garbage collection. The proposed work was initially implemented in the Italian City which has a hundred thousand inhabitants. This model has been tested and validated by using a simulation tool, which shows the model is economically feasible. The authors in this paper [31], consider a real-life waste collection problem that can be seen as a variant of the vehicle routing problem with time windows. In this approach, a definite time window has been allotted for serving of client, disposal facility, depot, lunch break for driver, and so forth. The study offers an algorithm based on the Ant Colony Meta-heuristics for real-life waste collection problems with better results. The power protocols 6LoWPAN (Pv6 over Low-Power Wireless Personal Area Networks) and RPL (Routing Protocol for Low-Power and Lossy Networks) are used for large geographical areas. If the default routes become unreachable, RPL is very adaptive to changing network circumstances by offering alternate routes. RPL performs better than 6LoWPAN when the node density is larger [32].

2.4 Comparative Analysis and Research Gaps

A Series of literature have been examined in this paper specially focused on the physical infrastructure of waste bins using IoT technology for an efficient waste management system. The comparative evaluation of this study is depicted in Table 1. In light of the physical infrastructure of IoT enabled bins for smart waste management system has been analysed based on various parameters such as type of waste supported; location of bins (outside or inside); pneumatic pipes tube that automatically compresses garbage to decrease volume; recycling points and processing points for discarded waste to convert in new items or re-processing for correct disposal; types of sensors; GPS and automatic actuators for preventing excessive deposit. Although plenty of techniques have been reported in the literature for handling waste management, however, a closer look reveals that few of the literature considered important features like pneumatic pipes, recycling, and processing point for organic waste, which are the major challenges to dispose of the waste efficiently. After a detailed study, several techniques [33, 34] for waste collection include technology like the IoT and real-time monitoring of smart bins using wireless sensors [35, 36]. It is important to understand that a balance has to be maintained between the energy consumption and the efficiency of the system because the sensors which are working on the bins are wireless and battery-based [37–40]. The energy must be used optimally, otherwise, the sensors are going to die very soon because they will run out of their batteries [41]. Further, the capabilities of a wide number of sensors distributed in large geographic areas with extended coverage using less power consumption must be taken into consideration. On the other hand, the route selection mechanism must be working most efficiently to ensure that the route selection is the best possible through multiple ways of optimization techniques. A comparative analysis of the smart waste management system is depicted in Table 2. In this review, we have seen a variety of architectures that are being used in which some have real-time data and others have real-time updates through a prototype. In an ideal scenario, the real-time data has to be maintained and interpreted in the prototype, to check the universal approach of the system. Along with it, the route selection has to be efficient through machine learning or with other optimization techniques that can ensure that the locations of the bins are also taken into account for route selection in an optimized manner. The length of the route is optimized to ensure the cost and fuel consumption is minimized. Another ideal aspect that is needed for the smooth mechanism is the long-range transmission option within the nodes along with low power consumption to save energy.

3 Challenges and Issues for Smart Waste Management System

Major challenges and issues of the waste management system are described as under.

3.1 System Segregation at the Collection Level

While implementing an SWMS, the householders need to augment recycling by themself, which will not only reduce the quantity of waste but also better the performance of the SWMS. The collection and segregation could be done at the source level by rag-pickers, so that reuse or reprocessing could be done efficiently. Thus this practice will reduce the required landfill area and also economically better.

3.2 Optimal Collection Route

Out of the total cost of SWMS, 80% of the cost spend on the collection of garbage. The garbage vehicle moves daily to cities for collecting the garbage discarded by the civilians. The main challenge is to provide the shortest and congestion-free path to the garbage vehicle which leads to saving time and money. Hence there is a need to optimize the route for garbage collecting vehicles to collect the filled bins from the different places in the city. The objective of this parameter is to achieve an effective path for garbage collector vehicles in terms of the shortest route, less fuel as well as less time-consuming.

Table 1 Comparative analysis of smart waste management system

Refs.	Type of waste	Bin location	Sensors	Pneumatic pipes	Recycling points	Processing points	GPS	Actuators
[12]	Glass, plastic, paper, metal	Outside	Capacity	No	Yes	No	No	No
[13]	Plastic	Outside	Capacity, weight	No	Yes	No	Yes	No
[14]	General waste	Outside	Capacity	No	No	No	Yes	No
[15]	General waste	Outside	Capacity	No	Yes	No	No	No
[18]	Glass, plastic, paper, general waste	Outside	Capacity, weight	No	No	No	No	No
[19]	Organic, glass, plastic, paper, metal	Underground	Capacity	No	Yes	No	No	No
[20]	General waste	Outside	Capacity	Yes	No	No	No	No
[21]	General waste	Outside	Capacity	No	No	No	No	No
[22]	Glass, plastic, paper, metal	Outside	Capacity	No	Yes	No	No	No
[33]	Organic, glass, plastic, paper, metal, toxic	Outside	Temperature, humidity, chemical, pressure	Yes	Yes	Yes	Yes	Yes

3.3 Communication Technology

Various communication technologies and protocols are being used in the IoT applications for sending/receiving the signals/data from the sensor to the network and vice-versa. Some of the prominent technologies are Bluetooth, wi-fi, zig-bee so on so forth. Everyone has its pros and cons. The major challenge with these techniques is short-range, maximum power consumption, and privacy. Hence there is a need to focus on improving their short-range because there are many IoT applications that need the long-range transmission of data/signals. Further, there is also a need

Table 2 Literature analysis of IoT-enabled technology used for smart waste management system

Refs.	Sensors enabled in smart bins	Algorithm	Use of GIS tool	Simulation tool	Architecture	Route selection	Communication technology	Limitations	Published year
[16]	Ultrasound sensor	Neighborhood search algorithm	–	–	Real-time	No	Lora LP-WAN	To improve the efficiency of the nodes	2018
[23]	Volume sensor, weight sensor	Different heuristics and local search method	GIS	–	Real-data	Yes	Lora WAN	To optimize the techniques to increase the life of the network	2018
[24]	HC-SR04 Ultrasonic sensor	Google Map API	GPS	ThinkSpeak	Real-time	Yes	Wi-Fi	Wi-fi is a small range of technology	2018
[28]	Ultrasonic sensor, load sensor	Particle swarm optimization	–	MATLAB	Real-data	–	Zigbee	Zigbee is prone to attack from unauthorized persons	2020
[29]	Ultrasonic, flame, weight, and temperature sensor	–	–	MATLAB	Real-data	No	GPRS	Unauthorized persons can access the network	2020
[33]	Ultrasonic sensor	–	–	ThinkSpeak	Real-time	No	Wi-Fi	Wi-fi is a small range of technology	2019

(continued)

Table 2 (continued)

Refs.	Sensors enabled in smart bins	Algorithm	Use of GIS tool	Simulation tool	Architecture	Route selection	Communication technology	Limitations	Published year
[34]	Ultrasonic sensor, IR sensor	–	Ethernet	–	Real-Time	No	Machine learning concept	Bins notification, do not have locations details and ID	2017
[35]	Ultrasonic sensor	Meta-heuristic: backtracking search algorithm	–	MATLAB	Real-data	Yes	–	Algorithms do not work efficiently when more constraints are added	2017
[36]	Ultrasonic sensor	Integer linear programming	GIS used	Net2Plan	Real-data	Yes	GPRS	Unauthorized persons can access the network	2019
[37]	Weight, ultrasonic sensor	–	GIS server	–	–	–	RFID	Needs a person to always have the RFID card upon using the bins for notification	2017
[38]	Load sensor, IR sensor	–	–	–	–	No	Wi-Fi	WiFi is short-range communication technology	2020

(continued)

Table 2 (continued)

Refs.	Sensors enabled in smart bins	Algorithm	Use of GIS tool	Simulation tool	Architecture	Route selection	Communication technology	Limitations	Published year
[39]	Ultrasonic sensor, weight sensor, gas sensor	Heuristic	–	–	Real-data	Yes	SIM900 GPRS, Wi-Fi	Unauthorized persons can access the network	2020
[40]	Capacity sensor	ROS and RRT	GPS	–	Real-time	Yes	Lora	Easily accessible by unauthorized persons	2020
[41]	Weight, temperature, humidity and, Ul	–	GPS model (modelNeo-6M)	–	Real-time	–	SIM900 GSM/GPRS module	Anybody can have access the system because the new user easily register himself	2020

to emphasize innovation on less power consumption as well as security for a better system.

3.4 User-Friendly System

The waste management system must be people's centric where they can easily interact with the system through a city App on a mobile platform. Which should have the capacity to receive complaints from various modes like Facebook, Twitter, etc., and dispose of them on the citizen satisfaction, to increase the happiness index of the citizens of the particular city. The city App should also be capable to upload an image of the discarded garbage thrown in the open area, to closely monitor and interact with the general public.

3.5 Lifetime Maximization of IoT Devices

An absolute necessity at this point of COVID-19 Lockdown is the automation of legacy the solid management system with the use of the latest entrant technology. The use of the IoT in the SWMS has changed the traditional methods with the latest digital technology and is also helpful in reducing costs. All the IoT-enabled bins in the city are being monitored centrally. But there is also the added benefit of using such technology which leverages end-to-end solutions from collecting to disposal of SWMS. Though IoT-enabled devices play a vital role to improve efficiency such devices are battery-enabled and have limited energy. We can control power transmission, find energy leaks, reduce power during peak hours, and do other things with a power management system. It is affordable to install an energy management system. Therefore need for the hour is to design and develop a holistic technique for IoT-enabled devices which consume less power, to maximize network lifetime in IoT applications for an efficient system. IoT sensors are battery-powered and have limited energy, which reduces the effectiveness of IoT networks as a whole. According to the adaptive sampling and sleep and wake-up technique, each IoT sensor selects an operational mode—such as transmission, sleep, or listening; during each cycle of the garbage collection path based on the filling rate and present condition of the bins. As a result, nodes are built to conserve as much energy as possible in order to extend the lifespan of IoT networks.

4 Critical Discussion

The objective of the paper is to study and analyze the state-of-the-art techniques used for the waste management system with smart IoT-enabled bins. This study is a comprehensive review of various techniques being used for the SWMS by using the IoT. Based on literature analysis, detailed information is typically presented in Table 2. In this table, we have studied mainly two categories of papers is IoT enabled bins and additionally, IoT-enabled bins along with routing algorithms used for the garbage collector vehicle. It has been found there is no significant difference in the overall performance of the system while using IoT-enabled bins without using the optimizing routing algorithm. By and large, the communication tools used to send the status of the bin to the receiving end are short-ranged and sometimes not able to send the full information of the smart bin. Further, there is also a need to emphasize the physical infrastructure of waste bins using IoT for an efficient waste management system. The article makes no mention of waste policies to be improved in order to have a sustainable garbage collection system in the future.

5 Conclusion

It has been observed that SWMS is a prominent area that has to be sorted out efficiently in every city. In day to day delightful acceptance of IoT, most of the connected edge devices around us have penetrated more or less every aspect of life. The absence of IoT platforms in the SWMS has manifestly been formidable in terms of everything. Hence, IoT is redefining the way SWMS was looked at initially. This paper analyzed the smart waste management system using the IoT and also route optimization methods used by the garbage vehicle. Apart from this, the paper also highlighted the various challenges and issues of the smart waste management system. Hence, it is significant that invention and enrichment be geared to manage the waste in our smart cities to ensure and sustain the foremost quality of life with a healthy ecosystem.

As a future scope, it is suggested that improvement in communication technology which leads to increased efficiency and extended coverage using less power consumption and reliability may be used with IoT-enabled solutions for the effectiveness of SWMS. Future research could focus on extending the solar-powered IoT sensor to enhance the life expectancy of sensing nodes and use unmanned aerial vehicles, or “drones,” for garbage collection and monitoring. Future research challenges may focus further on vehicle routing for garbage collection with territorial-specific parameters. Furthermore, the route optimization techniques used in solid management need further enrichment for fuel efficiency and crucial cost. The physical infrastructure of garbage bins using IoT which supports various parameters may be further improved for a better management system. Thus, using all the above features will certainly enhance the overall SWMS.

References

1. Gubbi J, Buyya R, Marusic S, Pal Aniswami M (2013) Internet of things (iot): a vision, architectural elements, and future directions. Future Gener Comput Syst 29(7):1645–1660
2. Bhatia M, Sood SK (2018) Internet of things based activity surveillance of defence personnel. J Ambient Intell Humanized Comput 9(6):2061–2076
3. Srinidhi NN, Dilip Kumar SM, Venugopal KR (2019) Network optimizations in the internet of things: a review. Eng Sci Technol Int J 22(1):1–21
4. Minghua Z, Xiumin F, Rovetta A, Qichang H, Vicentini F, Bingkai L, Giusti A, Yi L (2009) Municipal solid waste management in Pudong new area, China. Waste Manage 29(3):1227–1233
5. Rathi S (2006) Alternative approaches for better municipal solid waste management in Mumbai, India. Waste Manage 26(10):1192–1200
6. Kaushal RK, Varghese GK, Chabukdhara M (2012) Municipal solid waste management in India-current state and future challenges: a review. Int J Eng Sci Technol 4(4):1473–1489
7. Schaffers H, Komninos N, Pallot M, Trousse B, Nilsson M, Oliveira A (2011) Smart cities and the future internet: towards cooperation framework for open innovation. In: The future internet assembly. Springer, Berlin, Heidelberg, pp 431–446
8. Pavithra D, Balakrishnan R (2015) Iot based monitoring and control system for home automation. In: 2015 global conference on communication technologies (GCCT). IEEE, pp 169–173
9. Tirkolaee EB, Mahdavi I, Esfahani MMS, Weber G-W (2020) A hybrid augmented ant colony optimization for the multi-trip capacitated arc routing problem under fuzzy demands for urban solid waste management. Waste Manage Res 38(2):156–172
10. Mdukaza S, Isong B, Dladlu N, Abu-Mahfouz AM (2018) Analysis of iot-enabled solutions in smart waste management. In: IECON 2018–44th annual conference of the IEEE industrial electronics society. IEEE, pp 4639–4644
11. Lundin AC, Ozkil AG, Schuldt-Jensen J (2017) Smart cities: a case study in waste monitoring and management. In: Proceedings of the 50th Hawaii international conference on system sciences
12. Abd Wahab MH, Kadir AA, Tomari MR, Jabbar MH (2014) Smart recycle bin: a conceptual approach of smart waste management with integrated web based system. In: 2014 international conference on IT convergence and security (ICITCS). IEEE, pp 1–4
13. Thakker S, Narayanamoorthi R (2015) Smart and wireless waste management. In 2015 international conference on innovations in information, embedded and communication systems (ICIIECS). IEEE, pp 1–4
14. Folianto F, Low YS, Yeow WL (2015) Smartbin: smart waste management system. In: 2015 IEEE tenth international conference on intelligent sensors, sensor networks and information processing (ISSNIP). IEEE, pp 1–2
15. Ramya E, Sasikumar R (2017) A survey of smart environment conservation and protection for waste management. In: 2017 third international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB). IEEE, pp 242–245
16. Cerchecci M, Luti F, Mecocci A, Parrino S, Peruzzi G, Pozzebon A (2018) A low power iot sensor node architecture for waste management within smart cities context. Sensors 18(4):1282
17. Popa CL, Carutasu G, Cotet CE, Carutasu NL, Dobrescu T (2017) Smart city platform development for an automated waste collection system. Sustainability 9(11):2064
18. Catania V, Ventura D (2014) An approach for monitoring and smart planning of urban solid waste management using smart-m3 platform. In: Proceedings of 15th conference of open innovations association FRUCT. IEEE, pp 24–31
19. Aazam M, St-Hilaire M, Lung C-H, Lambadaris I (2016) Cloud-based smart waste management for smart cities. In: 2016 IEEE 21st international workshop on computer aided modelling and design of communication links and networks (CAMAD). IEEE, pp 188–193

20. Saha HN, Auddy S, Pal S, Kumar S, Pandey S, Singh R, Singh AK, Banerjee S, Ghosh D, Saha S (2017) Waste management using internet of things (iot). In: 2017 8th annual industrial automation and electromechanical engineering conference (IEMECON). IEEE, pp 359–363
21. Chowdhury B, Chowdhury MU (2007) Rfidbased real-time smart waste management system. In: 2007 Australasian telecommunication networks and applications conference. IEEE, pp 175–180
22. Lu J-W, Chang N-B, Liao L, Liao M-Y (2015) Smart and green urban solid waste collection systems: advances, challenges, and perspectives. IEEE Syst J 11(4):2804–2817
23. Khoa TA, Phuc CH, Lam PD, Bao Nhu LM, Trong NM, Thi N, Phuong H, Van Dung N, Tan-Y N, Nguyen HN, Minh Duc DN (2020) Waste management system using iot-based machine learning in university. Wireless Commun Mob Comput
24. Chaudhari SS, Bhole VY (2018) Solid waste collection as a service using iot-solution for smart cities. In: 2018 international conference on smart city and emerging technology (ICSCET). IEEE, pp 1–5
25. Anagnostopoulos T, Kolomvatsos K, Anagnostopoulos C, Zaslavsky A, Hadjiefthymiades S (2015) Assessing dynamic models for high priority waste collection in smart cities. J Syst Softw 110:178–192
26. Rada EC, Grigoriu M, Ragazzi M, Fedrizzi P (2010) Web oriented technologies and equipments for msw collection. Proceedings of the International Conference on Risk Management, Assessment and Mitigation-RIMA 10:150–153
27. Lozano A, Caridad J, De Paz JF, Gonzalez GV, Bajo J (2018) Smart waste collection system with low consumption Lorawan nodes and route optimization. Sensors 18(5):1465
28. Hannan MA, Akhtar M, Begum RA, Basri H, Hussain A, Scavino E (2018) Capacitated vehicle routing problem model for scheduled solid waste collection and route optimization using pso algorithm. Waste Manage 71:31–41
29. Jim AAJ, Rafiul Kadir M, Mamun AA, Abdullah-Al Nahid M, Ali et al (2019) A noble proposal for internet of garbage bins (iogb). Smart Cities 2(2):214–229
30. Faccio M, Persona A, Zanin G (2011) Waste collection multi objective model with real time traceability data. Waste Manage 31(12):2391–2405
31. Islam R, Sohel Rahman M (2012) An ant colony optimization algorithm for waste collection vehicle routing with time windows, driver rest period and multiple disposal facilities. In: 2012 international conference on informatics, electronics & vision (ICIEV). IEEE, pp 774–779
32. Hariharakrishnan J, Bhalaji N (2021) Adaptability analysis of 6LoWPAN and RPL for healthcare applications of ınternet-of-things. J ISMAC 3(02):69–81. https://doi.org/10.36548/jismac.2021.2.001
33. Atayero AA, Williams R, Badejo JA, Popoola SI (2019) Cloud based iot-enabled solid waste monitoring system for smart and connected communities. Int J Civil Eng Technol 10(2):2308–2315
34. Baby CJ, Singh H, Srivastava A, Dhawan R, Mahalakshmi P (2017) Smart bin: an intelligent waste alert and prediction system using machine learning approach. In: 2017 international conference on wireless communications, signal processing and networking (WiSPNET). IEEE, pp 771–774
35. Akhtar M, Hannan MA, Begum RA, Basri H, Scavino E (2017) Backtracking search algorithm in cvrp models for efficient solid waste collection and route optimization. Waste Manag 61:117–128
36. Bueno-Delgado M-V, Romero-Gázquez J-L, Jiménez P, Pavón-Marino P (2019) Optimal path planning for selective waste collection in smart cities. Sensors 19(9):1973
37. Al-Jabi M, Diab M (2017) Iot-enabled citizen attractive waste management system. In: 2017 2nd international conference on the applications of information technology in developing renewable energy processes & systems (IT-DREPS). IEEE, pp 1–5
38. Ali T, Irfan M, Alwadie AS, Glowacz A (2020) Iot-based smart waste bin monitoring and municipal solid waste management system for smart cities. Arab J Sci Eng 45:10185–10198
39. Johnson M, Emilin Shyni C Smart garbage bin with efficient routing and management system 14:80–89

40. Zhang Q, Li H, Wan X, Skitmore M, Sun H (2020) An intelligent waste removal system for smarter communities. Sustainability 12(17):6829
41. Anagnostopoulos T, Zaslavsky A, Kolomvatsos K, Medvedev A, Amirian P, Morley J, Hadjieftymiades S (2017) Challenges and opportunities of waste management in iot-enabled smart cities: a survey. IEEE Trans Sustain Comput 2(3):275–289

HLWEA-IOT: Hybrid Lightweight Encryption Algorithm Based Secure Data Transmission in IoT-MQTT Networks

S. Hariprasad, T. Deepa, and N. Bharathiraja

Abstract Internet of things (IoT) devices can store and manage the real-time data created by many restricted Internet-connected devices. If one of the nodes were compromised due to Man-in-the-Middle (MITM) attack, the network might suffer significant damage. Due to the limited resources of constrained devices, it is difficult to incorporate appropriate cryptographic capabilities. Hence lightweight cryptography strives to meet the security needs of situations with few resource-constrained devices. In this paper, the framework is constructed using the smart aircraft environment monitoring system (SAEMS) and created with the help of nodes and the message queuing telemetry transport (MQTT) protocol for communicating the sensor data. A hybrid lightweight encryption algorithm (HLWEA) is proposed to mitigate the MITM attack on IoT devices. The HLWEA comprises (i) Key generation and (ii) encryption and decryption. The proposed method achieves an encryption time of 0.0309 ms; encryption bandwidth is 19.02 kbps, decryption time of 0.029 ms and decryption bandwidth of 19.36 kpbs. The proposed implementation is a smaller key size, minimal time complexity, and enhanced real-time cryptography-capable security.

Keywords Internet of things (IoT) · Man-In-The-Middle attack · Lightweight cryptography · Key generation · Light encryption device

S. Hariprasad (✉) · T. Deepa
SRM Institute of Science and Technology, Kattankulathur, Tamilnadu, India
e-mail: hs6512@srmist.edu.in

T. Deepa
e-mail: deepat@srmist.edu.in

N. Bharathiraja
Dept. of Bio Medical Engineering, Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_59

1 Introduction

Internet of Things (IoT) devices are becoming more prominent in many domains which acquire and exchange the data connected to the internet [20]. IoT components are everything: any device, anybody, any service, company, path, network, anytime, and link. According to Statista [25], 30.9 billion IoT devices will be linked worldwide in 2025. Connectivity across infrastructure and services is not possible now without the development of IoT. The three levels of security aspects in IoT such as (1) Design security, (2) Hardware security (3) Data security. Data security has a lot of problems, like dealing with huge amounts of data and keeping track of all the IoT devices. For more extensive data, more time-consuming security algorithms like Advanced Encryption Standards (AES) [18] and Data Encryption Standards (DES) [21] are needed to protect the data in IoT networks.

Therefore, lightweight cryptography can support a smaller number of bits in the maximum of 32 or 64 bits of data with smaller keys up to 64 bits. There are other limitations on lightweight algorithms, which need to be looked into during the first phase of the standardization process. Lightweight cryptography is classified into block ciphers, hash functions, message authentication codes, and stream ciphers as shown in Fig. 1. Through a series of rounds, a block cipher encrypts the data bit by bit. These block cipher structures are classified into further types, such as substitution permutation network (SPN) and Feistel Network. In-Stream cipher encrypts one or two bits of data at a time. A hash function is a function that randomly encrypts the data according to the length of the data. The block cipher-based lightweight cryptographic primitives have performance advantages over other cryptographic standards based on power, energy consumption, latency and throughput. Therefore lightweight cryptography can support a smaller number of bits maximum of 32 or 64 bits of data with smaller keys up to 64 bits. The detailed algorithms/methods of all block cipher based lightweight cryptography and types of structure of block ciphers were discussed in the Table 1.

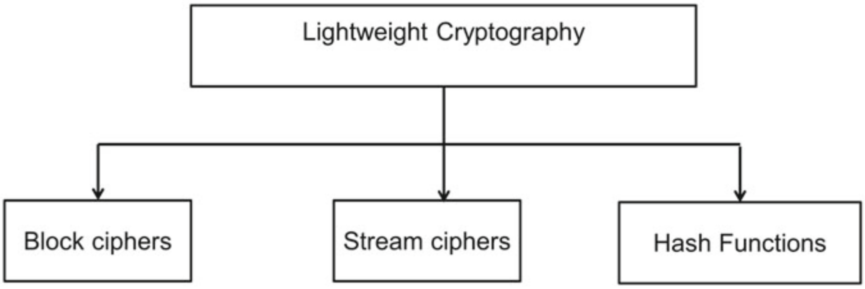


Fig. 1 Different types of lightweight cryptography methods

Table 1 Works related to lightweight cryptography

References	Method	Features
[4]	PRESENT (symmetric)	Ultra-lightweight
[29]	RECTANGLE (symmetric)	Fast implementation
[14]	HIGHT (symmetric)	Ultra-lightweight
[27]	TWINE (symmetric)	Energy efficient
[3]	SIMON (symmetric), SPECK (symmetric)	Key schedule
[10]	LED (symmetric)	Used in RFID tags
[6]	KTANTAN32 (symmetric)	Algebric equation
[17]	RSA (asymmetric)	More secured
[23]	ECC (asymmetric)	High speed
[5]	PRINCE (symmetric)	Low delay

The main research contribution is described as follows.

- A novel framework of simulation setup for a smart aviation environment monitoring system (SAEMS) is created with ten nodes.
- A novel HLWEA is formed using (i) Key generation and (ii) Encryption and decryption.
- A HLWEA is applied to protect the heterogeneous sensor data for secure transfer.
- Finally, an assessment is done to show the performance analysis of the proposed scheme, and it is compared with the existing system metrics such as computational time complexity.

2 Related Works

The IoT devices are resource-constrained to reduce the communication overhead. Lightweight protocol called message queuing telemetry transport (MQTT) is used for publishing/subscribing messaging to client-server communication [8]. It is designed as lightweight and easy to implement. These protocols can be used in restricted communications environments in the machine to machine (M2M), and less network bandwidth is needed. The various security aspects, application-related and compared protocols with MQTT are discussed below [19]. An IoT-based decision device for intelligent irrigation systems was developed using two protocols such as MQTT and HTTP. Compared to MQTT, other works of literature used a heavy computation protocol such as HTTP to transfer the sensor data to the cloud.

Studies have shown that using IoT with MQTT protocol has a lot of benefits [16]. Smart grids can now transmit and interpret data in real-time. The multi-tier edge computing model was developed [28] using two MQTT as a remote broker in fog and cloud. As part of [22]. MQTT messages are exchanged between several nodes, including MQTT Control Packets. There are three principal parts in an MQTT control packet: the fixed header, the variable header, and the payload. In recent years, edge computing has gained popularity among cloud service providers, such as Amazon, Google, and Microsoft [1], due to its low latency IoT connectivity and large data processing capacity. The architecture of the MQTT protocol is open source and easily vulnerable to attacks. The study [2] eventually led to a more robust protocol development for IoT-related applications. Secure-MQTT [12] uses fuzzy logic to detect malicious MQTT broker node activities. MQTT publisher traffic patterns are taken into consideration while selecting traffic features [9]. The attack simulation results in a network with 10 to 50 % malicious nodes, resulting in many malicious nodes.

The usage of the MQTT protocol in IoT with the working of encrypting the message transfers in applications is demonstrated in [11]. Fiestal network structure converts 64-bit input into two 32-bit output. RPP decodes data using logical XOR, encrypts and decrypts data via bit swapping and recursive positional substitution on prime and nonprime of cluster results in 64-bit cipher [7]. Data from many heterogeneous IoT devices with attack detection modelling named the SENMQTT-SET has been discussed [24]. A study of alternative media transport mechanisms in IoT networks is conducted [13] using constrained application protocol (CoAP) and MQTT-SN for media propagation in low power lossy networks (LLNs). Order messages and re-send lost messages are essential components of IoT's reliable message communication system [15]. An algorithm known as Improved Artificial Bee Colony (IABC) [26] is used to determine when the key should be upgraded for maximum effectiveness.

2.1 Problem Statement

The nodes in an aviation monitoring system are resource-constrained devices on the heavy computation cryptography algorithm, leading to computational complexity and time decay. The proposed Hybrid Lightweight Encryption Algorithm (HLWEA) method addresses this issue. This HLWEA proves that eavesdropping has been a complication and cannot hijack the nodes.

3 Methodology

The proposed methodology comprises two parts (a) creating a simulation setup framework for SAEMS and (b) securing the data using HLWEA.

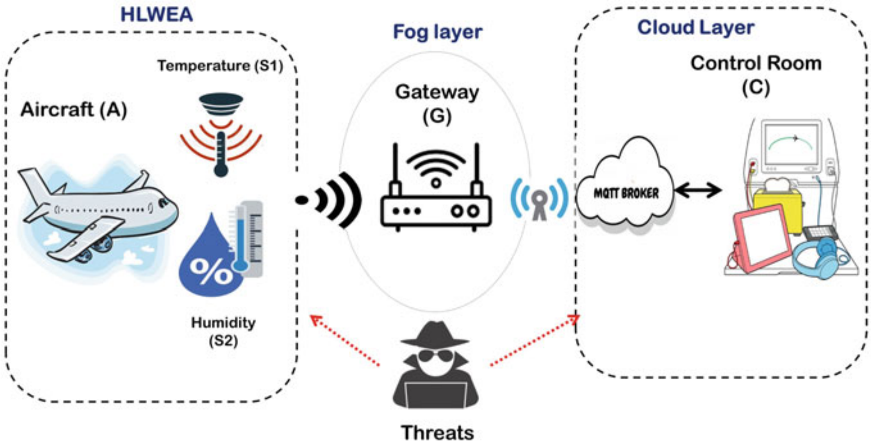


Fig. 2 Framework for smart aviation environment monitoring system (SAEMS)

3.1 *Framework of SAEMS*

The proposed framework consists of four modules such as (i) Sensors, (ii) Threat analysis, (iii) Cloud and (iv) Graphical user interface (Control room), as illustrated in Fig. 2.

Sensors The SAEMS framework is constructed by using ten nodes by temperature and humidity sensors. The temperature sensor is used to sense the engine's temperature, and humidity is used to monitor the environment at an instance of time. These two parameters are vital in preventing any hazards to the aircraft nodes. The sensed data of all nodes is fetched and communicated to the server using the MQTT protocol. The collected sensor information is encrypted using the proposed HLWEA transferred through the MQTT protocol. The secure sensor data is like a publisher, and the data analytics section is like a subscriber. The cloud works as a broker MQTT, checking the publisher and subscriber identity using topics and keys generated.

Threats There are more chances to prevail against more threats in the fog layer for the data. These threats could be controlled and precise from the forging of collected data. Sometimes these threats can destroy the devices by node traps like MITM. To overcome the existing threats, HLWEA is proposed for more real-time security without modifying any existing service architecture.

Cloud MQTT broker will act as a cloud. It has two other main components, such as publisher and subscriber. The node of the publishing packet needs to be encrypted. The sensed data of all nodes is fetched and communicated to the MQTT broker with a specific topic. The collected sensor information is encrypted and decrypted using the proposed HLWEA.

Graphical User Interface(GUI) The collected information from the cloud server is given to the control room for data analytics. The received encrypted data will be decrypted at actuators using a micro python script with the help of the key of each node. The GUI part was created with the help of node-red and MQTT brokers. The collected sensor data from the publisher node is encrypted using the lightweight proposed cryptography method using the HLWEA algorithm. The encrypted data will be decrypted at the GUI admin using python script and can be able to witness the node environment details at every instance.

3.2 Hybrid Lightweight Encryption Algorithm (HLWEA)

The HLWEA is used to secure the sensor data of the nodes and transmit it to the receiver for data analytics. The proposed HLWEA system consists of a lightweight cypher. It consists of (i) Key generation and (ii) Encryption/Decryption. The proposed HLWEA is unbreakable and compatible with lightweight devices. The embedded system needs more hardware components, cost, area and power consumption; hence, a new replacement for hardware components is a software aspect called lightweight cryptography.This ciphers are solutions for hardware and more secured transmission. Figure 3 illustrates the proposed flow diagram HLWEA and the steps involved in the proposed method. Initially, the temperature sensor data of the aircraft nodes is considered as plain bits and the key is generated using the SPECK key scheduling algorithm using the MQTT topic, which is a public key. Later, using encryption for encrypting, the plain bits are changed to cipher bits and sent to the end user using the MQTT protocol. In the GUI, the encrypted bits are decrypted with the of a public key of the MQTT topic by using a decryption algorithm.

Key generation SPECK Key generation is based on two steps (i) round functions and (ii) Key schedule. In the round function, the sensor data is first added bitwise of the key taken from the MQTT topic. Later by addition modulo and key schedule of left and right circular shift. Finally generated, the key for encryption and decryption to be performed.

Encryption and Decryption After the SPECK key schedule with add of sensor data is further encrypted using four steps and 8, 12 rounds. The four steps are (i) Add constant and Shift columns, (ii) Substitute S box, (iii) Shift rows and (iv) Mix Columns.

Add Constant and Shift columns The constants are added to the consecutive bits of data received after adding the key. The columns are shifted.

Substitute S box For substitute S box, the RECTANGLE method is used. It is also a lightweight block cipher and works fast in implementation. RECTANGLE has a four-bit box substitution. RECTANGLE consumes low power than any other box and is a more secure way of transmission and prevents linear attacks.

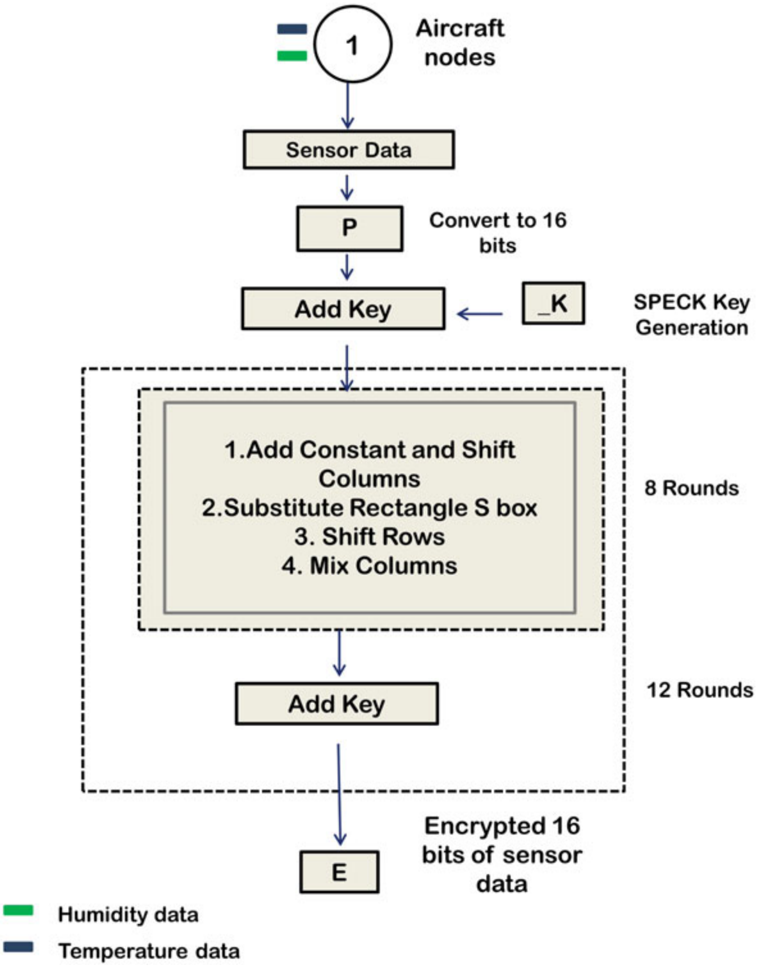


Fig. 3 Proposed flow diagram for ELWEA

Shift rows The rows are shifted for encrypting using the linear feedback shift keying method to shift the bits.

Mix Columns At last, The column is mixed once again for shuffling the bits to get stronger encrypted code. Each constant value is mixed using a linear feedback shift register and updated with a new value.

4 Results and Discussions

The experimental setup of 10 aircraft nodes has been constructed with the help of python 3.6. Two vital parameters of aircraft nodes, such as temperature and humidity data, are generated randomly at an instance of time. Figure 4 illustrates the attack design framework of the proposed model. The proposed HLWEA protects the data from aircraft nodes to the control room. If an attacker gets an MQTT packet, they can not decrypt the data without the key.

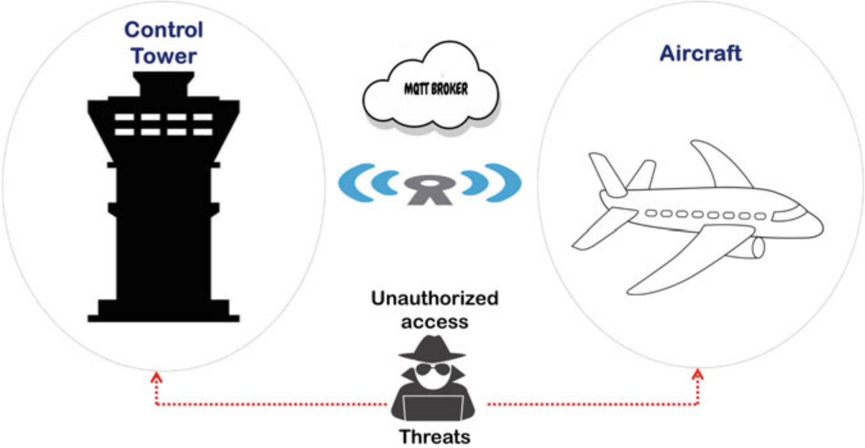
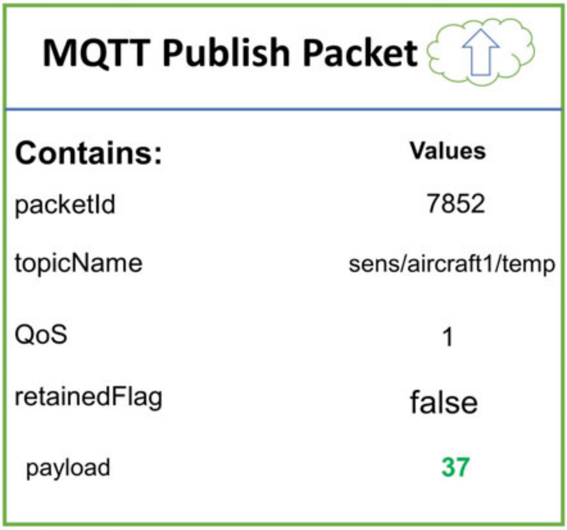


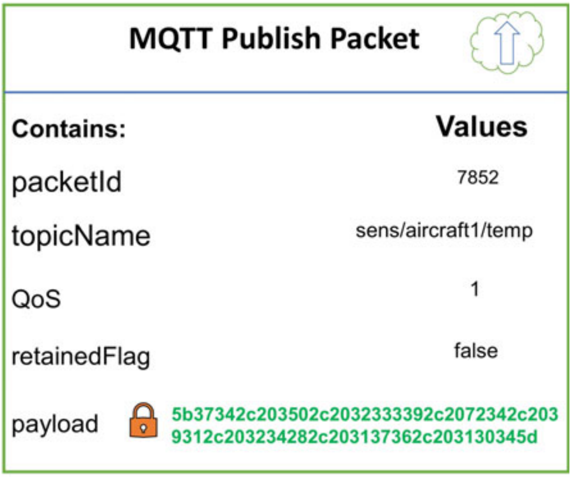
Fig. 4 Attack design framework

Fig. 5 Message received to the control room without encryption



MQTT Publish Packet

Contains:	Values
packetId	7852
topicName	sens/aircraft1/temp
QoS	1
retainedFlag	false
payload	37



MQTT Publish Packet

Contains:	Values
packetId	7852
topicName	sens/aircraft1/temp
QoS	1
retainedFlag	false
payload	5b37342c203502c2032333392c2072342c2039312c203234282c203137362c203130345d

Fig. 6 Message received to the control room with proposed HLWEA

Table 2 Proposed HLWEA Performance metrics

Method	Encryption	Decryption	Encryption	Decryption
	time	time	Throughput	Throughput
	(ms)	(ms)	(kbps)	(kbps)
Proposed	0.0309	0.029	19.02	19.36

Figures 5 and 6 shows the data transmitted and received between aircraft nodes to the data centre without encryption and the proposed HLWEA algorithm. The proposed encryption algorithm has been deployed in all ten aircraft nodes. Sensor nodes will act as a publisher, MQTT broker servers as a cloud, and the end data centre is called a subscriber. During this time, the publisher sends with and without encryption parts are reflect the subscriber system. Table 2 shows the various comparison like block size and key in terms of bits and time in seconds compared to the proposed method. To calculate the node lifetime of the proposed model, the voltage of each node is considered as 3.3 V DC, and the current is ten microamps. The total lifetime of the nodes is around five years.

5 Conclusion

This study describes end-to-end payload MQTT-based IoT devices. Ten nodes using the MQTT protocol make up SAEMS infrastructure to protect data against spoofing, and HLWEA was proposed. Initially the ten nodes are considered without encryption, and then ten nodes are considered with HLWEA. Compared to the other literature, the proposed method achieved less time complexity of encryption time of 0.0309 ms. Encryption bandwidth is 19.02 kbps, decryption time of 0.029 ms and decryption bandwidth of 19.36 kpbs. In future work, the same lightweight encryption can be implemented other protocols.

References

1. Ahmad T, Morelli U, Ranise S, Zannone N (2022) Extending access control in AWS IoT through event-driven functions: an experimental evaluation using a smart lock system. Int J Inf Secur 21(2):379–408
2. Akhtar S, Zahoor E (2021) Formal specification and verification of MQTT protocol in pluscal-2. Wireless Pers Commun 119(2):1589–1606
3. Beaulieu R, Shors D, Smith J, Treatman-Clark S, Weeks B, Wingers L (2015) The simon and speck lightweight block ciphers. In: Proceedings of the 52nd annual design automation conference, pp 1–6
4. Bogdanov A, Knudsen LR, Leander G, Paar C, Poschmann A, Robshaw MJ, Seurin Y, Vikkelsoe C (2007) Present: an ultra-lightweight block cipher. In: International workshop on cryptographic hardware and embedded systems, Springer, pp 450–466
5. Borghoff J, Canteaut A, Güneysu T, Kavun EB, Knezevic M, Knudsen LR, Leander G, Nikov V, Paar C, Rechberger C et al. (2012) Prince–a low-latency block cipher for pervasive computing applications. In: International conference on the theory and application of cryptology and information security, Springer, pp 208–225
6. Cannière CD, Dunkelman O, Knežević M (2009) Katan and ktantan-a family of small and efficient hardware-oriented block ciphers. In: International workshop on cryptographic hardware and embedded systems, Springer, pp 272–288
7. Chatterjee R, Chakraborty R, Mondal J (2019) Design of lightweight cryptographic model for end-to-end encryption in IoT domain. IRO J Sustain Wirel Syst 1(4):215–224
8. Edited by Andrew Banks and Rahul Gupta: MQTT Version 3.1.1. (2014). http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html
9. González-Zapata AM, Tlelo-Cuautle E, Cruz-Vega I, León-Salas WD (2021) Synchronization of chaotic artificial neurons and its application to secure image transmission under MQTT for IoT protocol. Nonlinear Dyn 104(4):4581–4600
10. Guo J, Peyrin T, Poschmann A, Robshaw M (2011) The led block cipher. In: International workshop on cryptographic hardware and embedded systems, Springer, pp 326–341
11. Gupta V, Khera S, Turk N (2021) MQTT protocol employing IoT based home safety system with ABE encryption. Multimedia Tools Appl 80(2):2931–2949
12. Haripriya A, Kulothungan K (2019) Secure-MQTT: an efficient fuzzy logic-based approach to detect dos attack in MQTT protocol for internet of things. EURASIP J Wirel Commun Netw 2019(1):1–15
13. Herrero R (2020) MQTT-SN, COAP, and RTP in wireless IoT real-time communications. Multimedia Syst 26(6):643–654
14. Hong D, Sung J, Hong S, Lim J, Lee S, Koo BS, Lee C, Chang D, Lee J, Jeong K et al (2006) Hight: A new block cipher suitable for low-resource device. In: International workshop on cryptographic hardware and embedded systems, Springer, pp 46–59
15. Hwang HC, Park J, Shon JG (2016) Design and implementation of a reliable message transmission system based on MQTT protocol in IoT. Wireless Pers Commun 91(4):1765–1777
16. Kondoro A, Dhaou IB, Tenhunen H, Mvungi N (2021) Real time performance analysis of secure IoT protocols for microgrid communication. Futur Gener Comput Syst 116:1–12
17. Mahto D, Khan DA, Yadav DK (2016) Security analysis of elliptic curve cryptography and RSA. In: Proceedings of the world congress on engineering, vol 1, pp 419–422

18. Moradi A, Poschmann A, Ling S, Paar C, Wang H (2011) Pushing the limits: a very compact and a threshold implementation of AES. In: Annual international conference on the theory and applications of cryptographic techniques, Springer, pp 69–88
19. Nawandar NK, Satpute VR (2019) IoT based low cost and intelligent module for smart irrigation system. Comput Electron Agric 162:979–990
20. Ray PP (2018) A survey on internet of things architectures. J King Saud Univ Comput Inf Sci 30(3):291–319
21. Satoh A, Morioka S (2003) Hardware-focused performance comparison for the standard block ciphers AES, camellia, and triple-des. In: International conference on information security, Springer, pp 252–266
22. Seoane V, Garcia-Rubio C, Almenares F, Campo C (2021) Performance evaluation of COAP and MQTT with security support for IoT environments. Comput Netw 197:108338
23. Sharma S, Chopra V (2017) Data encryption using advanced encryption standard with key generation by elliptic curve diffie-hellman. Int J Sec Appl 11(3):17–28
24. Siddharthan H, Deepa T, Chandhar P (2022) SenMOTT-set: an intelligent intrusion detection in IoT-MQTT networks using ensemble multi cascade features. IEEE Access 10:33095–33110
25. Statista Research Department: Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (2016). https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/
26. Suma V, Haoxiang W (2020) Optimal key handover management for enhancing security in mobile network. J Trends Comput Sci Smart Technol (TCSST) 2(04):181–187
27. Suzaki T, Minematsu K, Morioka S, Kobayashi E (2011) Twine: a lightweight, versatile block cipher. In: ECRYPT workshop on lightweight cryptography, vol 2011
28. Veeramanikandan M, Sankaranarayanan S (2019) Publish/subscribe based multi-tier edge computational model in internet of things for latency reduction. J Parallel Distrib Comput 127:18–27
29. Zhang W, Bao Z, Lin D, Rijmen V, Yang B, Verbauwhede I (2015) Rectangle: a bit-slice lightweight block cipher suitable for multiple platforms. Sci China Inf Sci 58(12):1–15

A Practical Approach for Crop Insect Classification and Detection Using Machine Learning

Ravindra Yadav and Anita Seth

Abstract Insect identification is one of the most pressing difficulties for Indian farmers, as numerous insect species harm a vast number of crops and hence diminish the quality of harvests, resulting in financial losses for both farmers and the country. However, in agriculture, the combination of IoT and machine learning (ML) allows for ease and innovation, allowing farmers all over the world to better their farming operations. On the other hand, in India, a very little amount of farmers is aware of smart farming and its benefits. Various study on research paper shows that the proper use of IoT devices embedded with the machine learning algorithm can reduce the task of farmer at very early stage of the plant life and thus saving the crops from being degraded, also included the survey of various research done across the globe and identified the potential methods which must be included for the current era farmers in order to minimize the insect effect on the crops. The aim of our experiment is to involve ML and IoT technology to sense the crop conditions in terms of quality and whether it is affected by insect or not for this a experimental study with the help of image processing has been performed thus calculation of results done accordingly. There are various sensors, which are equipped with ML technology like computer vision algorithm, which make the sensor powerful, and images being captured by these sensor can be analysed automatically and thus trigger the automated pesticide treatment systems using a ML-based decision support model. In this paper, study of Convolution Neural Network (CNN), Long Short Term Memory (LSTM), Support Vector Machine (SVM), Grid search based SVM (Grid-SVM), and K-nearest Neighbour classifier has been done Among them based on the required performance, nd the CNN-based model is much accurate for predicting the required treatment. The CNN has achieved up to 88% of accurate classification. Further, the model has been extending by incorporating the regression analysis, which enables the system to

R. Yadav (✉) · A. Seth
IET DAVV, Indore, Madhya Pradesh, India
e-mail: ryadav@ietdavv.edu.in

A. Seth
e-mail: aseth@ietdavv.edu.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_60

recommend the quantity of the required treatment. In this context, study on the KNN regression and Support Vector Regression (SVR) model has been made, among them the KNN regression provides up to 99.8% accurate prediction for treatment quantity prediction.

Keywords Machine learning · Internet of Things · Pest control · Supervised learning · Insect class analysis

1 Introduction

Many obstacles can develop in terms of data limitation, uneven data count, and background noise when it comes to automatically identifying the insects harming the crops. For improved performance, they must all be overcome [1]. Mixing CNNs based on numerous topologies (EfficientNetB0, ResNet50, GoogleNet, ShuffleNet, tMobileNetv2, and DenseNet201) and various Adam optimization techniques performed well for pest identification. Two novel Adam algorithms based on the Adam variant DGrad for deep network optimization are proposed, which include a scaling component in the learning rate that is applied to the absolute difference term. On three benchmark bug data sets, CNNs with different data augmentation to assure variety or different types of Adam optimization were trained. Three evaluation measures were used to compare and evaluate fusions. The best performing ensemble, made up of CNN ensembles and the additional variations introduced here, is demonstrated to outperform the literature on all three tests [2]. The highly accurate method for the detection of bug and pest in the farm inspire the farmers for adoption of new technologies to work in their field [3]. When it comes to automatically detecting the insects causing crop damage, many challenges can arise due to data limitations, unequal data counts, and background noise. They must all be conquered in order to increase performance [1]. For pest detection, a combination of CNNs based on a variety of topologies (EfficientNetB0, ResNet50, GoogleNet, ShuffleNet, tMobileNetv2, and DenseNet201) with several Adam tuning approaches performed well. This paper shows, the highlights of recently carried out our work in this area [4–8]. In our recent contribution [2nd paper reference], the IoT-enabled and ML-based technique employment in agriculture has been considered. In this paper, an overview of these contributions hase been discussed in the paper Further discusses the objectives, which were established previously. Further, the experimental study-based identified ML algorithm has used to design a predictive model, which recommends the relevant treatment for the plants. In this context, different deep learning models has been modelled, which will learn on the IoT sensor-based plant health conditions traces and predict the requirement of the pest control [9–12]. The work has also investigated the regression analysis techniques to predict the required quantity of the treatment. Therefore the paper includes the implementation and results analysis of two different architectures of smart farming module. Among them, first one is a low cost solution and can be deploy only with the treatment prediction. The second model is a fully featured model, use of classification as well as regression technique has been made for providing an exclusive system for automating pesticide treatment [12, 13].

This section provides the overview of the proposed work involved in this paper, the next sections provide the summary of recent contributions, further, In this paper discussion on an experimental study and their consequences has been shown, and then proposed the required smart farming models for the prediction of water, fertilizer and pesticide treatment of the crops. Finally, by using the experimentations, the results have discussed and the future direction of research work has been proposed.

2 Literature Review

The recent technologies, methods, and applications in agriculture have taken into consideration to collect 50 research articles from different journals using Google Scholar. These contributions has summarized in this section.

2.1 Noteworthy Contribution

The source has given the wide specification and classifies the insects based on their nature, and provides a wide spectrum to understand the kind of diseases associated with a specific type of insects. The below table provide the details.

Wu et al. [14] has addressed the problems with agricultural product output caused by the presence of insects in the field, emphasizing the importance of having an accurate dataset for insect identification so that timely preventative measures can be done to avoid economic losses. They gathered a large scale dataset for insect pest recognition called IP102. There are almost 75,000 photos in the dataset, divided into 102 categories.

Thenmozhi and Srinivasulu Reddy [15] while identifying the issue of pest detection, the writers of this research took into account a variety of crops. They described how to use a Convolution neural network (CNN) in conjunction with deep architecture to solve the problem of identifying different insects that have very little difference in terms of shape and size among them. The CNN performs automatic feature extraction and learns complex high-level features in image classification applications.

Caballero et al. [16] use hyper spectral imaging (HSI) and multispectral imaging (MSI) to address crop health, water or fertiliser use, and probable sickness. Writer has provided a description that gives an overview of some of the pertinent scientific literature on the use of HSI and MSI on agriculture fields. Some of the applications include the detection of contaminants and heavy metals, as well as the management and assessment of water.

Kulkarni et al. [17] proposed to generate messages to notify farmers. That will assist farmers by getting data from the land to take necessary steps to do. Jankielsohn [18] discusses the different varieties of the insect which are friend of the farmer and evaluate a methodology how to increase the growth of good insects, what ae the different favourable condition for them to grow and sustain Insects have achieved enormous In terms of species diversity and abundance, it has been a success. Insects are the most abundant group of organisms on the planet, accounting for roughly 66% of all animal species. They can be found almost anywhere, and because they are excellent dispersers and exploiters of virtually all types of organic matter, they are an important part of every ecosystem and provide valuable ecosystem services. Insects have long been regarded as competitors in the fight for survival. Herbivorous insects cause 18% of global agricultural output to be lost. Despite this, pests make up less than 0.5% of the total number of bug species known. Humans controlling the environment choose crops for their larger growth and higher output. Jaiganesh is a type of insect pest. Sowmyashree and Srinivas [19] looked at the role of IOT in agriculture. It allows for the creation of yields, estimation, composts, diseases details for cure, and development proposals. Taneja et al. [20] is offering an irrigation system to help reduce water usage. They compute the amount of water by measuring various characteristics. The system is low-cost and energy-efficient. Sensors were employed to manage the irrigation valve, and a smart phone was used to monitor the situation remotely. Høye and Johanna [21] has introduced about the Computer vision and deep learning advances which may bring innovative answers to this global problem of accurate insect identification,. Entomological observations can be made effectively, continuously, and noninvasively using cameras and other sensors throughout the diurnal and seasonal cycles. In the lab, automated imaging can also capture the physical appearance of specimens. When trained on these data, many deep learning models can guess the nearby estimates of insect, biomass, and diversity. Deep learning models can also assess variation in phenotypic traits, behaviours, and interactions. To upgrade irrigation, Ratnayake et al. [22] Crop development, reduced water use, and proper water use are the goals. They calculated the required water quantity using humidity, temperature, and soil moisture. Ilyas et al. [23] Using computer vision and a deep learning network, the author attempts to observe insect behaviour in their natural habitat. They have developed a technique for tracking insect behaviour using image-based tracking. a new hybrid detection and tracking method for outdoor monitoring of unmarked insects they developed a software and this software can detect an insect, discern when a tracked insect is blocked from view and when it re-emerges, and integrate a succession of insect locations into a coherent route. Nandyal et al. [24] has performed systematic literature review (SLR). To analyse and evaluate primary research of image-based insect identification and species classification algorithms, the author has deducted that 980 research published between 2010 and 2020 and chose 69 relevant studies using specified inclusion/exclusion criteria from among them. In this SLR, they examined the dataset properties (i.e. bug species targeted, crops, geographical locations, image capturing methods) and insect classification methodologies used in the primary studies.

Jha et al. [25] plan a WSN for estimation of soil conditions. The kind of soil has evaluated and bases on soil composition the crop and fertilizers have prescribed. Thinking about the changing pace of soil, a methodology had applied to build a harmony between energy utilization and the exactness of phosphorus. Manoukis et al. [26] for entomologists, the author covers the principles of applied computer vision, including as image capture, data extraction, and analysis. They go over some of the most cutting-edge imaging gear and cameras, as well as lighting, software, and basic data collection scenarios, as well as specific examples. Quantification of behavioural events will become more widespread in insect research, that computer vision techniques for quantification will become more widely used, and that the application of these tools and approaches will yield new insights and answers to entomology challenges'. Alzu'b et al. [27] Using a case study, focuses on CloudIoT for solutions in various sectors. The MQTT protocol was used to create an irrigation system that is 22% more energy efficient and 15% faster. Tripathi et al. [28] aims at classifying soils based on characteristics and recommending the best crop using IoT and ML. Rehman et al. [29], used IoT to build a smart greenhouse. They propose voice control, which replaces the conventional interface, lowers the barrier of entry to science and technology, inspires a range of applications, and improves the production. Venkat et al. [30] proposes a geo-fencing and livestock tracking solution. Create a safe zone using IoT and GPRS, with dedicated sensors for the livestock. Cattle can be monitored and controlled remotely using data on their location, well-being, and health. Nagaveni et al. [31], a system has developed to watch the growth of crops development parameters. The system consists of a drone, with a camera to record images. It includes various crops decisions based on image. From images, they analysis the amount of green in leaf, moisture content etc. Sanches et al. [32] The Internet of Things (IoT), Wireless Communications (WiFi), Machine Learning (ML), and Artificial Intelligence (AI) are all explored. Crop diseases, storage management, insecticides, weed management, irrigation challenges, and water management are all problems in agriculture. Soil production and fertility have both been found to improve with automation.

According to, Mas et al. [33] To meet problems, the agribusiness is armed itself with tools like sustainable farming. IoT is helpful in all aspects of farming, crop monitoring, water level monitoring, pest and animal control, and soil richness data. It has been determined to employ remote sensing methods. The situation was evaluated using remote cameras. Bjerge and co. The author demonstrates a portable computer vision system that can draw in and find live insects. The results suggest that live animals should be photographed and utilised to identify and categorise species when they are attracted to a light trap. An Automated Moth Trap (AMT) was constructed to draw and keep track of live insects during the evening and night. It has a camera and different light sources. Counting and Classifying Moths (MCC) a computer vision system based on deep learning image analysis, tracked and counted insects while also recognising moth species. 48 evenings, more than 250,000 people participated. An average of 5675 pictures were taken each night [34]. A special convolutional neural network was trained on 2000 labelled photos. Figueiredo et al. [35] The author of the study has suggested a smart trap with Internet of Things (IoT) capabilities that employs computer vision to recognise the desired insect. The answer consists of three parts: A web application that displays data using a programmable heat map, an embedded system with a camera, GPS sensor, and motor actuators, and IoT middleware that serves as a database service provider. The primary concern is addressed and the suggested cure is put into practise. Saoud [36] provide a novel method to assess the ease of identification of insects using their characteristic values (CV). They investigated two things in order to do this: (1) changes in SPIPOLL IESs, and (2) the connection between CV and IESs. The CV be applied to determine the IES of SPIPOLL insects, according to the findings.

3 Literature Summary

The farmers' world need to explore and test innovative technology to meet our food needs and create sustainable farming practises because of the expanding population, global warming, and quick changes in climatic conditions. On the other hand, farming is difficult and has a low income due to low production yields and uncontrollable variables. As a result, improving the processes involved in traditional farming is essential to improve crop output and quality. In this regard, IoT and machine learning-based methods significantly affect farming activities. Crop output and the management of agricultural resources are just two examples of how agriculture operations have improved. Numerous recent additions to agricultural activities have been examined, including monitoring, resource planning, irrigation (water supply), fertiliser, soil productivity, classification, disease identification, and others. These technologies enable farmers to obtain useful data for loss prevention while automating a variety of tasks. Application trends have also been found based on the literature. Figure 1 shows a variety of IoT and machine learning-based applications [37]. This increases the power of farmers and farming practises. According to current trends, irrigation and water management automation takes up the majority of work, with crop recommendation models and irrigation coming in second. Monitoring and examination of the soil are also important. Databased methodologies are beneficial [38]. The apps that also put a focus on animal welfare include the ones that develop management information systems (MIS), reviews, and surveys. Due to certain meteorological conditions, a particular form of pest is produced in the environment, which leads to a particular type of disease in the crop, lowering the quality of the crop thus result in degradation of the farmers as well as the country.

4 Objectives

The focus of the work is to survey various methods of pest detection and classification in the field of agriculture using machine-learning algorithms along with use of IoT and propose a method to be included in our future work for pest detection and

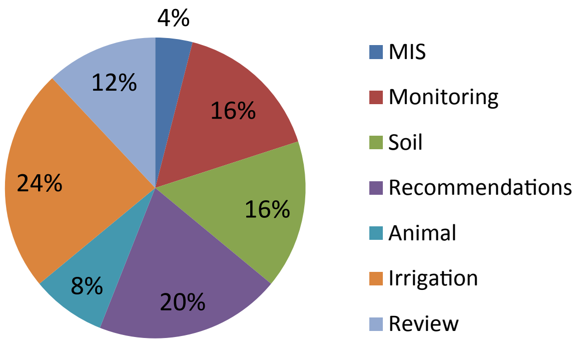


Fig. 1 Applications of ML and IoT in agriculture [39]

classification based on shape and size features using machine learning and IoT. In this context, based on known facts some key objectives has established as.

4.1 To Review and Understand the Morphology of the Insects

To study the particular insect verities on the basis of kind of defect made by the insect whether the insect is a leaf eater or stem attacker or the insect attacks on the root also the different classes of the insects which grows in a particular whether system with the help of available literature and the data obtained from the field directly also to identify the specific disease associate with a specific kind of insect.

4.2 To Detect and Classify the Pest Based on Shape and Size Features, Using ML Techniques

The aim is detect and classify pest using shape and size features data model that usage the data in the image format to train and predict the class of pest and soil dryness to provide instructions to cultivate the crops or other relevant treatment like fertilizer and pest control. This will help to enhance the involved agricultural processes.

4.3 To Study Various Machine Learning and Deep Learning Algorithm for Insect Classification and Identification

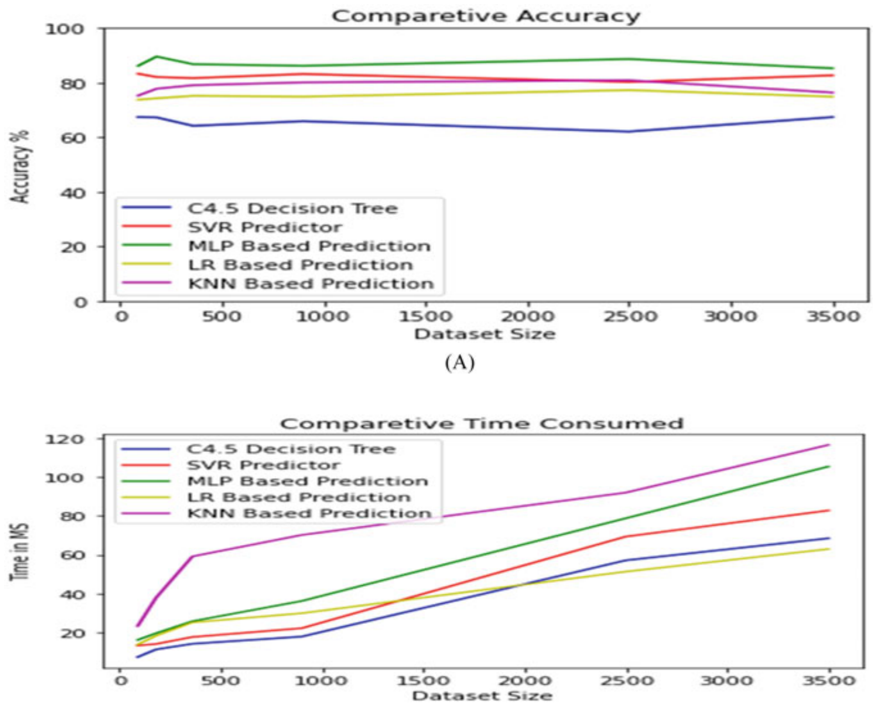
To study about the various literatures available on the machine-learning algorithm for the detection and classification of insect.

4.4 To Develop the Hybrid Machine-Learning Algorithm for the Classification and Detection of Insects

In order to get the optimized results we will develop the hybrid version of the machine-learning algorithm.

5 Recent Study

Recently, investigated some effective ML method of pest detection and classification. Thus, proposed a new machine-learning model for pest detection and classification in the crop fields with insect database, which are publically available [40, 41]. The established objectives are working on ML based techniques. In this context, an experimental study has carried out for identifying the efficient and accurate ML algorithms. In this study, algorithms namely, C4.5 decision tree, Support Vector Regression (SVR), Multi-layer Perceptron (MLP), Linear Regression (LR) and k-Nearest Neighbour (KNN) classifier was use. We have taken a dataset of different sizes from UCI machine learning repository. Additionally, we have considered two parameters i.e. accuracy and time to compare the models. Based on a set of experiments the obtained results are demonstrating in Fig. 2. The performance of the model in terms of accuracy has given in Fig. 2a. Figure 2b demonstrates the training time in MS. According to experimental observation, we found that the MLP, SVR, and KNN are producing higher accurate prediction [39]. On the other hand, in terms of time consumption the SVR, C4.5 decision tree, and linear regression is providing low time consumption. However, based on the performance of both the parameters we found MLP produces higher accuracy and less time consumption. Thus, we have proposed to be using the suitable variants of artificial neural network (ANN). Along with this we have studied the **CNN model** [42]. The CNN model is a type of neural network that lets us extract higher-level representations from picture input. Unlike classical image recognition, which requires the user to create image characteristics, CNN takes the raw pixel data from the image, trains the model, and then extracts the features for better categorization. And **YOLO** [43] Many applications, such as self-driving cars, necessitate great precision and real-time inference speed. As a result, selecting an Object Detector that satisfies both speed and accuracy requirements are critical. YOLO (You Only Look Once) is a single-stage object detector that accomplishes both of these objectives (i.e., speed and accuracy). To offer you a complete view of the YOLO family, there are various YOLO versions (for example, YOLOv1, YOLOv2, YOLOX, YOLOR).



(A)

(B)

Fig. 2 The comparison of machine learning algorithms in terms of **a** accuracy (%) and **b** training time [39]

6 Proposed Work

Based on the conducted study and observed facts, we have established some objectives. The process work flow is shown in Fig. 3. Among them, we have discussed the objective to design a ML framework for support the agriculture process by suggesting the treatment for the crops by classification of different type of pest. The Hardware (Graphical processing Unit) for experiment to be carried out is Nvidia RTX 4000 with 1 TB of SSD.

6.1 Dataset and Methods

For the future perspective we have performed our partial work in this paper for that at the very first stage we have collected the samples of insect images from IP102 dataset [44]. Dataset Exhibits natural Long tailed Distribution, It has hierarchical Taxonomy and the insect pests.

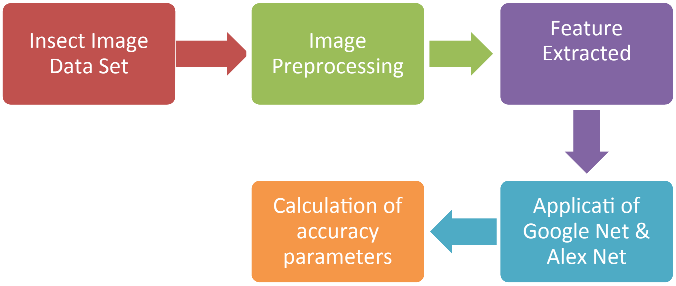

Insect Image Data Set
Image Preprocessing
Feature Extracted
Applicati of Google Net & Alex Net
Calculation of accuracy parameters

Fig. 3 The work flow

Which mainly affect one specific agricultural product is grouped in the same upper level category. Our proposed work consists of following data summary. The percentage of Train image is 79% and the Test image is about 21%.

Train Image Count: 15,178
Test Image Count: 3798
Total Image Count: 18,976 ≅19,000.

X-axis: Class Labels 0–102
Y-axis: total number of samples the Y-axis values are not shown clearly as the amount of data is very large.

The sample images related to the datasets are shown in Fig. 4.

6.2 Image Preprocessing

In order to get the exact features of the images we have applied following steps given below. All the images were randomly applied preprocessing, the preprocessing functions were applied in the following sequential way.

1. Grayscale image with 3 channel output.
2. Canny Edge Detection method applied on images.
3. Cropping the image to the bounding boxes.
4. Random Horizontal Flip with probability = 0.5.
5. Random Rotation with degree of rotation = [−20, 20].
6. Random vertical Flip with probability = 0.3.
7. Converting to tensor to be fed into Dataset Classes.

Sample Images

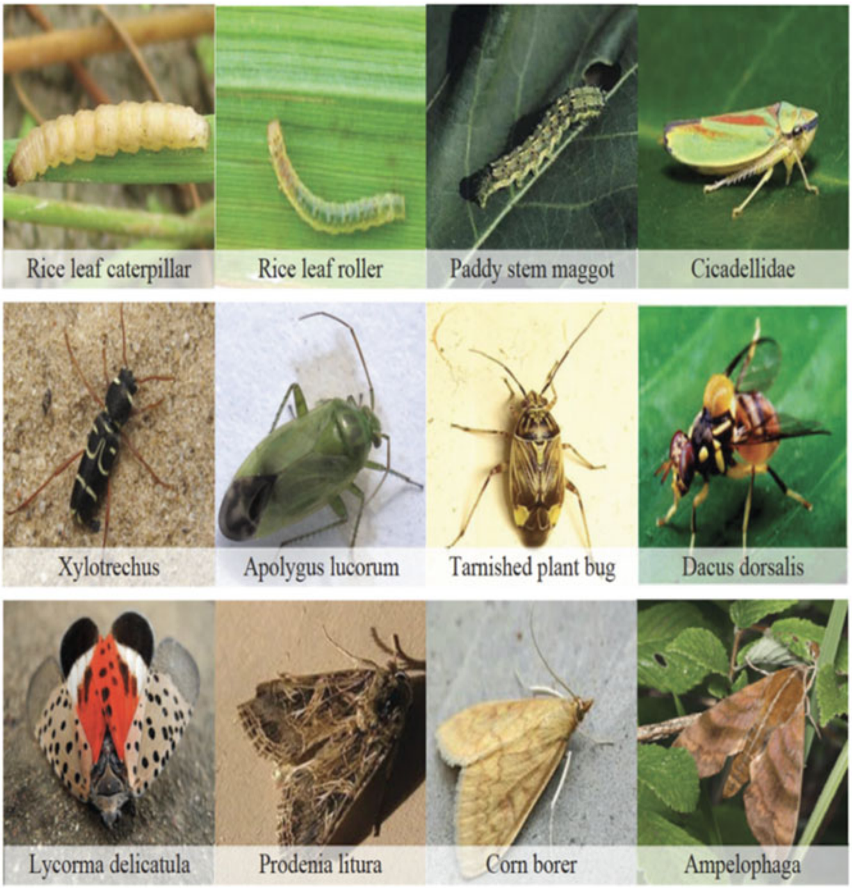


Fig. 4 Sample images

The Noise reduction helped us in the correct identification of the edges of the insects images. For the Noise reduction, The above 7 steps has been applied on image data.

X-axis: Class Labels.
Y-axis: Total No of Samples.
The training and the testing datasets are shown in Figs. 5 and 6.

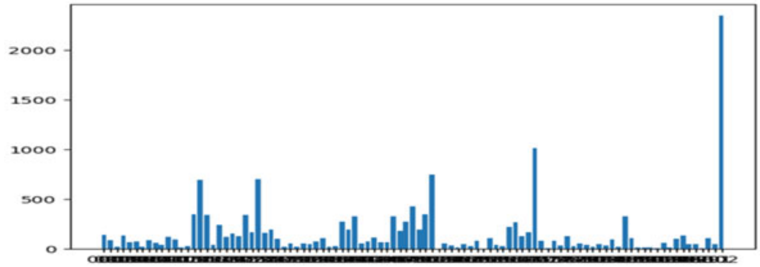


Fig. 5 Training data set

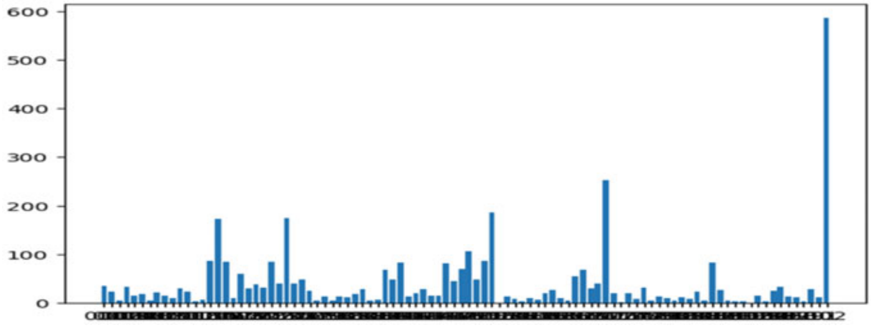


Fig. 6 Testing dataset

6.3 Features Extracted and Method of Extraction

Dataset used is unstructured, and thus the number of features extracted is a black box in nature. Although the features extracted are directly correlated to the number of learnable params in each of the algorithms used. Initially Cross Validation was not applied while splitting the dataset; Dataset was split from the start. Till now what we observed is In Vision Based learning, moreover it increases variance in CNN networks after applying Cross Validation.

6.4 Methodology

The methodology framework are shown in Figs. 7 and 8.

Rescaled Size of image:

$$H * W * D$$

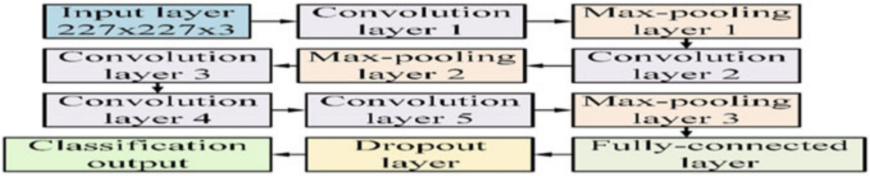


Fig. 7 Alex Net [45]

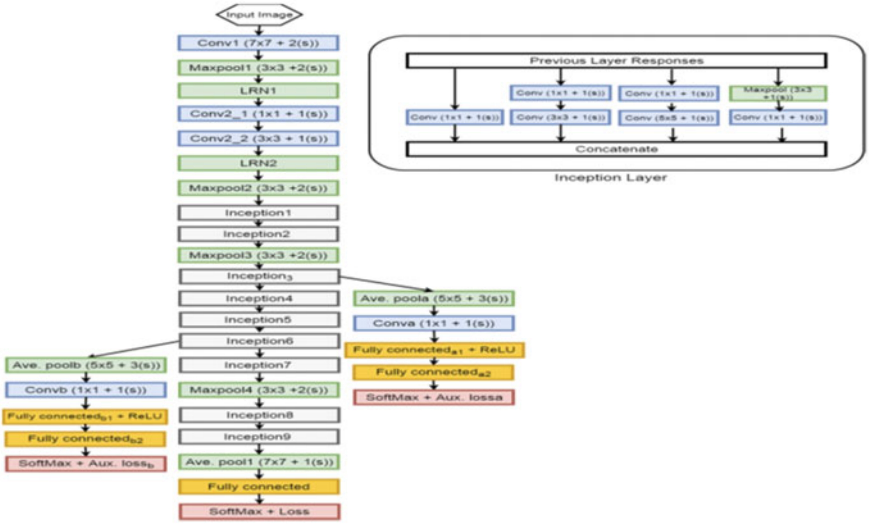


Fig. 8 GoogleNet [46]

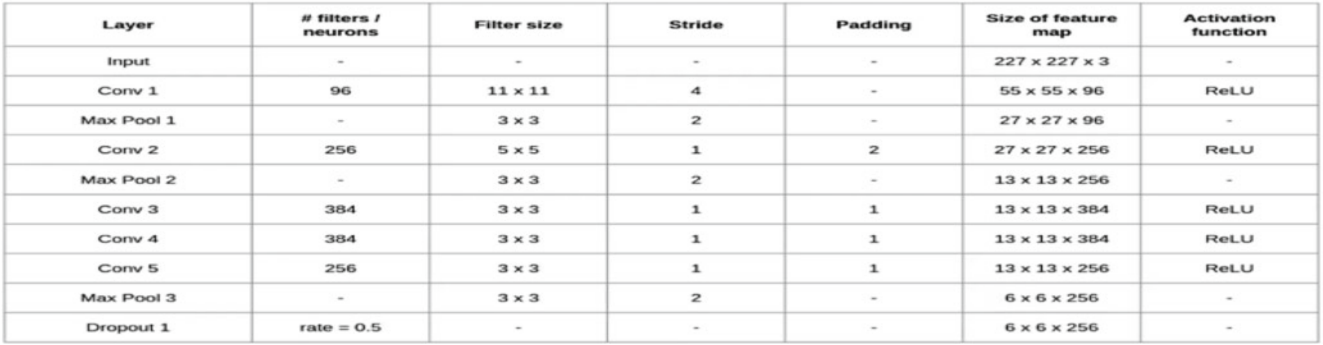
H	Height: 227
W	Width: 227
D	Image channels: 3 (all are same as the image is grayed out after transformations).

6.5 Filter Size for Each Layer

Filter size are shown in Figs. 9 and 10.

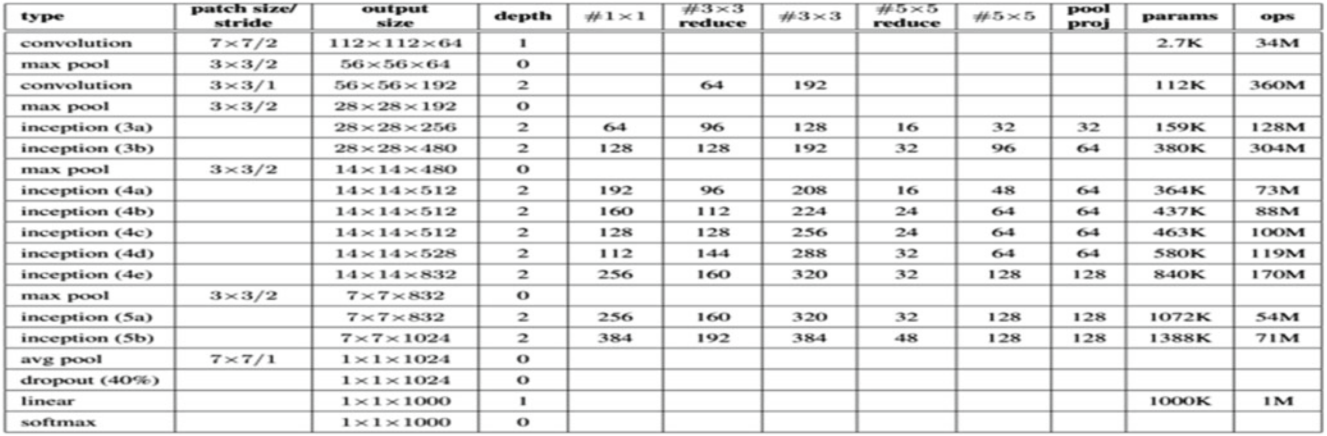
6.6 Classification Accuracy

Classification accuracy for Alex net and GoogleNet are shown in Table 1.



Layer	# filters / neurons	Filter size	Stride	Padding	Size of feature map	Activation function
Input	-	-	-	-	227 x 227 x 3	-
Conv 1	96	11 x 11	4	-	55 x 55 x 96	ReLU
Max Pool 1	-	3 x 3	2	-	27 x 27 x 96	-
Conv 2	256	5 x 5	1	2	27 x 27 x 256	ReLU
Max Pool 2	-	3 x 3	2	-	13 x 13 x 256	-
Conv 3	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 4	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 5	256	3 x 3	1	1	13 x 13 x 256	ReLU
Max Pool 3	-	3 x 3	2	-	6 x 6 x 256	-
Dropout 1	rate = 0.5	-	-	-	6 x 6 x 256	-

Fig. 9 Filter size for Alex Net [45]



type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Fig. 10 Filter size for Google Net [46]

Table 1 Classification accuracy for Alexnet and GoogleNet

Model	Best val loss	Best accuracy	Precision	Recall	F1 score
Alexnet	48.1019	0.3431	1	1	0.5
GoogleNet	8.5205	0.3532	1	1	0.5

Table 2 Correct identification of insect class

Predicted class	Class name	Actual class	Class name
4	Asiatic rice bowler	4	Asiatic rice bowler

Table 3 Wrong identification of insect class

Predicted class	Class name	Actual class	Class name
52	Blister beetle	51	Legume blister beetle

6.7 Computational Time

Both the classifiers were trained with zero-shot learning and thus training time was relatively fast.

Alexnet: ~15 s for 10 epochs
GoogleNet: ~16 s for 10 epochs
Highest Classification Accuracy
Alex net = 0.3431
GoogleNet = 0.3532 (Table 2).

GoogleNet showed better classification Accuracy (Table 3).

Figures 11, 12 and 13 are the validated output results. Classifiers are trained to give an array of probabilities with 102 elements signifying the distinct 102 classes of the dataset. The achieved Confidence AlexNet and Google Net are 98.34 and 99.34 respectively. The Loss value is 48.10 and 8.520 respectively.

7 Conclusion

In this paper Study on number of research paper has been done about the insect classification and detection using the deep learning model and the existing dataset which has been collected by the no of different resources like IoT, Google, manual etc. This paper shows an experiment which lead us towards the bigger solution for the problem of insect identification and classification specially for the Indian farmers. Further work can be done using several other deep learning models like Yolo6 from where further results can be deduced for insect identification and classification.

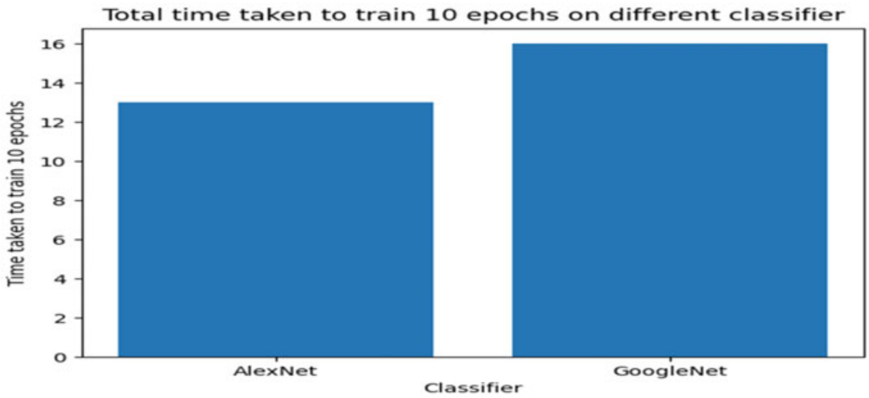


Fig. 11 Training time graph

Fig. 12 Confidence 0.98345

Fig. 13 Confidence 0.99342

References

1. Abeywardhana DL, Dangalle CD, Nugaliyadde A et al (2022) An ultra-specific image dataset for automated insect identification. Multimedia Tools Appl 81:3223–3251
2. Nanni L, Manfè A, Maguolo G, Lumini A, Brahnam A (2022) High performing ensemble of convolutional neural networks for insect pest image detection. Ecol Inf 67:101515. ISSN: 1574-9541
3. Bhattacharya P, Neamtiu I, Shelton CR (2012) Automated, highly-accurate, bug assignment using machine learning and tossing graphs. J Syst Softw 85(10):2275–2292. ISSN: 0164-1212
4. Rajasekaran T, Anandamurugan S (2019) Challenges and applications of wireless sensor networks in smart farming—a survey. Adv Bi Da Cl Comp Adv Inte Sys Comp 750
5. Maduranga MWP, Abeysekera R (2020) Machine learning applications in IOT based agriculture and smart farming: a review. Int J Engg App Sci Tech 4(12):24–27
6. Sowmiya M, Prabavathi S (2019) Smart agriculture using Iot and cloud computing. Int J Rec Tech Engg 7(6S3)
7. Abd El-GhanyShadia NM, Abd El-AzizShadia E, Abd El-AzizShahira E, Marei S (2020) Environ Sci Pollut Res 27(6)
8. Utkin KY, Marenych M, Galych O, Sliusar I (2020) Main aspects of the creation of managing information system at the implementation of precise farming. In: 11th IEEE international conference on depe. system, service and technology, Kyiv, Ukraine
9. Gimeno CR, Voort Mvd, Niemid JK, Lauwers L, Kristensen AR, Wauters E (2019) Assessment of the value of information of precision livestock farming: a conceptual framework. NJAS –Wage. J Li. Sci 90–91:100311
10. Farooq MS, Riaz S, Abid A, Abid K, Naeem MA (2019) A survey on the role of IoT in agriculture for the implementation of smart farming, vol 7. IEEE, spec. sect. on new technology for Sm. Far. 4.0: research challenge & opportunity
11. Doshi J, Patel T, Bharti Sk (2019) Smart farming using IoT, a solution for optimally monitoring farming conditions. Proc Comput Sci 160:746–751
12. Rekha P, Ramesh MV, Rangan VP, Nibi KV (2017) High yield groundnut agronomy: an IoT based precision farming framework. 978-1-5090-6046-7/17/$31.00 ©2017. IEEE
13. Jaiganesh S, Gunaseelan K, Ellappan V (2017) IOT agriculture to improve food and farming technology. In: Proceedings of IEEE conference on emerging devices and smart systems, Mahe. Engg. Coll., Tamilnadu, India. 978-1-5090-5555-5/17/$31.00. IEEE
14. Wu X, Zhan C, Lai Y-K, Cheng M-M, Yang J (2019) Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 8787–8796
15. Thenmozhi K, Srinivasulu Reddy U (2019) Crop pest classification based on deep convolutional neural network and transfer learning. Comput Electron Agric 164:104906. ISSN 0168-1699
16. Caballero D, Calvini R, Manuel Amigo J (2020) Hyperspectral imaging in crop fields: precision agriculture. Amigo JM (ed) Data handling in science and technology, vol 32. Elsevier
17. Kulkarni A, Kulkarni P, Dandin P, Kumar S (2020) Smart and economic farming using IoT. © JUL 2020, IRE J 4(1)
18. Jankielsohn A (2018) The importance of insects in agricultural ecosystem. Adv Entomol 6(2)
19. Sowmyashree S, Srinivas C (2020) A survey on smart soil analysis and predicting the irrigation using IOT a literature survey and review paper. Int J Inn Sci Res Tech 5(2)
20. Taneja M, Jalodia N, Byabazaire J, Davy A, Olariu C (2019) SmartHerd management: a microservices-based fog computing–assisted IoT platform towards data-driven smart dairy farming. Softw: Pract Exper 49:1055–1078. published by John Wiley & Sons Ltd.
21. Høye T, Johanna A (2022) Deep learning and computer vision will transform entomology. In: Proceedings of the national academy of sciences. https://doi.org/10.1073/pnas.2002545117 2022/03/27
22. Ratnayake MN, Dyer AG, Dorin, Tracking individual honeybees among wildflower clusters with computer vision-facilitated pollinator monitoring. PLoS ONE 16(2):e0239504. https://doi.org/10.1371/journal.pone.0239504

23. Ilyas QM, Ahmad M (2020) Smart farming.: an enhanced pursuit of sustainable remote livestock tracking and geofencing using IoT and GPRS. Hind Wire Comm Mob Comp 2020, Art. ID 6660733:12
24. Nandyal S, Khamitkar PS, Joshi PS (2019) Agridrone: automation of agriculture using IoT. Int J Innov Sci Res Tech 4(6)
25. Jha K, Doshi A, Patel P, Shah M (2019) A comprehensive review on automation in agriculture using artificial intelligence. Artif Intel Agric 2:1–12
26. Manoukis NC, Collier TC (2019) Computer vision to enhance behavioral research on insects. Ann Entomol Soc Am 112(3):227–235. https://doi.org/10.1093/aesa/say062
27. AlZu'bi S, Hawashin B, Mujahed M, Jararweh Y, Gupta BB, An efficient employment of internet of multimedia things in smart and future agriculture. Mult Appl. https://doi.org/10.1007/s11042-019-7367-0
28. Tripathi MK, Maktedar DD (2020) A role of computer vision in fruits and vegetables among various horticulture products of agriculture fields: a survey. Inf Proc Agric 7:183–203
29. Rehman TU, Mahmud MS, Chang YK, Jin J, Shin J (2019) Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. Comput Electr Agric 156:585–605
30. Venkat PS, Avinash BL, Jabber B (2020) Crop yield prediction based on Indian agriculture using machine learning. In: International conference for emerging technology, Belgaum, India
31. Nagaveni V, Raghavendra BK (2019) A review on machine learning classification techniques for plant disease detection. In: 5th International conference on advanced computer and communication system, 978-1-5386-9533-3/19/$31.00 ©2019. IEEE
32. Sanches D, Júnior AL, da Costa CC, de Castro Victoria D, Inamasu RY, Grego CR, Ferreira VR, Ramirez AR (2020) Precision and digital agriculture: adoption of technologies and perception of Brazilian farmers. Agriculture 10:653. https://doi.org/10.3390/agriculture10120653
33. Más FR (2020) From smart farming towards agriculture 5.0.: a review on crop data management. Agronomy 10:207. https://doi.org/10.3390/agronomy10020207. www.mdpi.com/journal/agronomy
34. Bjerge KN, Sepstrup JB, Helsing-Nielsen MV, Høye F (2021) An automated light trap to monitor moths (Lepidoptera) using computer vision-based tracking and deeplearning. Sensors. https://doi.org/10.3390/s21020343
35. Figueiredo V, Campos A, Mafra S, Rodrigues J (2020) A proposed IoT smart trap using computer vision for sustainable pest control in coffee culture. arXiv 2020
36. Saoud Z (2020) Can we estimate insect identification ease degrees from their identification key paths. Ecol Inf 55:101010. ISSN: 1574-9541
37. Geetha K (2021) An integrated approach for crop production analysis from geographic information system data using SqueezeNet. J Soft Comput Paradigm 4:308–321
38. Krishnan GH, Rajasenbagam T (2021) A comprehensive survey for weed classification and detection in agriculture lands. J Inf Technol 3(4):281–289
39. Yadav R, Seth A (2022) A review for investigation on soil features using IoT and ML. In: 2022 4th international conference on smart systems and inventive technology (ICSSIT)
40. Abeywardhana DL, Dangalle CD, Nugaliyadde A et al (2022) An ultra-specific image dataset for automated insect identification. Multimed Tools Appl 81:3223–3251.https://doi.org/10.1007/s11042-021-11693-3
41. Espinoza K, Valera DL, Torres JA, López A, Molina-Aiz FD (2016) Combination networks as a novel approach for the identification of *Bemisia tabaci* and Frankliniella of image processing and artificial neural occidentalis on sticky traps in greenhouse agriculture. Comput Electron Agric 127:495–505
42. https://towardsdatascience.com/understanding-cnn-convolutional-neural-network-69fd626ee7d4
43. Sharma A (2022) Introduction to the YOLO family. In: Chakraborty D, Chugh P, Gosthipaty AR, Haase J, Huot S, Kidriavsteva K, Raha R, Thanki A (eds) PyImageSearch. https://pyimg.co/3cpmz

44. Wu X, Zhan C, Lai Y-K, Cheng M-M, Yang J (2019) IP102.: a large-scale benchmark dataset for insect pest recognition. In: 2019 IEEE/CVF conference on computer vision and pattern recognition, pp 8779–8788. https://doi.org/10.1109/CVPR.2019.00899
45. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K, SqueezeNet.: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size
46. Khan RU, Zhang X, Kumar R (2019) Analysis of ResNet and GoogleNet models for malware detection. J Comput Virol Hack Tech 15:29–37. https://doi.org/10.1007/s11416-018-0324-z

Attendance Portal Using Face and Speaker Recognition

Sahil Sharma, Shivam Prajapati, Merin Meleet, and B. S. Rekha

Abstract There is a strong correlation between attendance of school and offices and its attendees' performance and success. The traditional ways to maintain attendance for organizations are time consuming and cumbersome. A novel way of doing this task is proposed in this paper where attendance of a person is marked based on his/her face and voice based on voice and speaker recognition. Both the biometrics are preprocessed to feed the combination as a datapoint to the Convolutional Neural Network. This ensures that proxy attendances are avoided and the shallow network is able to perform well. The model achieved an accuracy of above 90%. A python based interface facilitates the entire process of person registration, attendance marking and database maintenance.

Keywords Face recognition · Speaker recognition · Convolutional neural network · Spectrogram · Cascade classifier · Haar features

1 Introduction

Attendance forms an integral part of human gatherings. Children that miss school frequently often fall behind both academically as well as professionally. In offices attendance is used to keep track of the number of leaves the employee has taken. In seminars and conferences attendance is a measure of its popularity and people's

S. Sharma · S. Prajapati (✉) · M. Meleet · B. S. Rekha
Department of Information Science and Engineering, R V College of Engineering, Mysore Road, Bengaluru 560059, India
e-mail: shivamp.is19@rvce.edu.in

S. Sharma
e-mail: sahilsharma.is19@rvce.edu.in

M. Meleet
e-mail: merinmeleet@rvce.edu.in

B. S. Rekha
e-mail: rekhabs@rvce.edu.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_61

interest and seriousness towards it. Hence keeping track of attendance is necessary wherever applicable. Conventional pen paper system for taking attendance is tiring, time consuming and prone to errors and proxies. Newer methods like Radio Frequency Identification described in [1], iris recognition given in [2] and fingerprint recognition are costly. As every person has several unique biometric features like face and voice, any of them can be used to identify them. To avoid a person marking attendance for someone else, a combination of face and voice is a good biometric.

Face recognition is a computer vision technology which detects and visualizes human faces in digital images. Object recognition is a superset of this domain which attempts to monitor an instance of a semantic object. These objects can be classes of people, cars, animals or anything else. Face recognition technology is important in many areas like marketing and security. For object detection Cascade Classifiers and Haar Features are the widely popular methods used. Michael Jones and Paul Viola proposed face detection as in [3], which is also used. The algorithm in which a cascade method is trained using a lot of images is a machine learning algorithm called cascade classifier. These images are categorized into two, one containing the target object called positive images and the other not containing the target object called the negative images. There are different types of cascade classifiers based on different target objects, human face being the target is a popular one. With the human face as the target it has to extract features of the human face. Haar features are convolution kernels which are basically permutations of black and white rectangles. Most of the features calculated are irrelevant so only the relevant ones are identified by Adaboost.

Speaker recognition has two broad categories, that are verification and identification of the speaker. The process of determining given audio corresponds to which registered speaker is called speaker identification. And the process by which the system either accepts or rejects the identity needed by the speaker is called speaker verification. Most applications where voice is used to verify speaker identity are classified as speaker verification. Every person has different characteristics in their speech which is caused by differences in anatomy and behavioural patterns. The conventional techniques use characteristics of the human voice to uniquely identify them. ML algorithms might not be able to detect sound without normalization when the range of values of signals varies remarkably. Feature scaling is the method that is employed to generalize the range of unconstrained variations or data attributes. The data is then scaled to bring all of it to the same scale. Another factor that affects the performance of the speaker recognition system is the number of channels in the audio file. These files are recorded in either of two formats, mono and stereo. The former contains only one channel, while the latter has more than one channel. System's performance can be significantly improved by converting files from stereo to mono format. Removing silent phases in the recording is also used for better model performance. The next step is feature extraction where distinctive features of a speech signal are identified, based on ones voice's frequency, pitch and energy with respect to time, and a compact representation of the raw form of acoustic signals are made. Some of the techniques that are used for the extraction of notable features from audio file of a person are Linear Predictive Coding, Gammatone Frequency Cepstral Coefficients, Mel Frequency Cepstral Coefficient (MFCC) and Power Normalized Cepstral Coefficients. MFCC is the most commonly used feature extractor in speech recognition tasks. Its working is similar to a human ear, wherein sound is represented in both linear and non-linear cepstrals. We can also use an unsupervised ML model that is mainly used for finding the solution to tasks such as data mining and clustering, i.e. we can use Gaussian Mixture Model abbreviated as GMM for the task of speaker recognition as well. This model relies on a certain number of Gaussian distributions, each of which represents a separate cluster. The grouping of data instances from a single distribution are the main focus of this model. The task of speaker recognition provides much better accuracy when we combine different techniques of feature extraction like MFCC and GMM. For the training of GMM the expectation maximization algorithm is used. In this algorithm gaussian mixtures are created by using maximum likelihood estimates for the updation of gaussian means. GMMs are typically trained with audio samples from a particular speaker to distinguish between individual audio features. If the GMM is trained with a large set of voice samples, it can learn common voice features and convert them to a universal background model (UBM). A better handling of alternative speech that may be encountered during speaker recognition like whispering, fast or slow speech can be achieved when GMM and UBM models are combined, applicable to both the quality and type of speech, as well as the composition of the speaker.

The advancements in deep learning algorithms using Convolutional Neural Networks present many possibilities. Not only images but CNNs are capable of working with image representations of sound as well. Using this capability, a novel method using a combination of face and voice for taking attendance in organizations is presented in this paper.

2 Literature Survey

Harikrishnan et al. [4] worked on Development of surveillance systems that take real time attendance using artificial neural networks with user-friendly graphical user interface. 74% was the max accuracy for recognition that was achieved. The entire graphical user interface that was developed was very intuitive and was built in such a way that it could be used with small-sized (pocket sized) computers such as Raspberry Pi. A feature to store the attendance of the user to the server automatically was also provided by the system. Smitha and team [5] also used face recognition to build an attendance recording system for their class. The four phases of the system were creation of the database, followed by facial detection, and then recognizing the face and ending with updating the attendance along with attendance mailing feature to the faculty after each session. The users of the system had the option to utilize three different features in the graphical interface, that are, student registration, faculty registration, and mark attendance. Authors in [6] used smart glasses to implement face recognition. For face detection Haar-features method was used which has an accuracy of 98%. For face recognition, transfer learning was used upon AlexNet, which gave the accuracy to be almost 98.5% by making use of 2500 different images in a class.

In [7], Yuan et al. analysed various identity authentication methods and particularly explored speaker recognition methods in detail. A new CNN architecture was developed by Salehghaffari [8] for verifying the speaker and simultaneously capturing and discarding the speaker information and background noise. His given method outperformed the primitive speaker verification techniques in which background models are directly used to make models for speakers as described in the previous section. Shah and team [9] aimed to use voice as biometrics, but they gave a technique for taking attendance that uses a very small data set. Report generation facility was also provided by the automated attendance recording system [10, 11]. Depending on the quality of the input taken during the initial stage the accuracy changed. The current accuracy was obtained to be 80%. Wang et al. [12] Compared relatively small Convolutional Neural Networks (CNN) and evaluated effectiveness of speaker recognition using existing models on edge devices with application of transfer learning technique to deal with a problem of limited training data. The preliminary results proved that the chosen model adapted the benefit of computer vision tasks by using CNN and spectrograms to perform speaker classification with precision and recall of around 84% in time less than 60 ms on mobile devices with Atom Cherry Trail processor. In Gomes and team's [13] work CNN was used to identify the person from the speech input given by him by making use of speech dataset. The dataset was made up of voice recordings of 60 subjects. For the task of recognizing the speaker, they used CNN networks that were MobileNet v1 and Inception v3. It was done on spectrograms of the audio dataset. Inception v3 provided an accuracy of 82.9% on the test data when the step size was 4000 and MobileNet v1 provided an accuracy of 81.5% respectively. The accuracy for test data was increased to some extent for both the networks when there was a reduction in the step size to 2000. Becker et al. [14] used an already existing technique of layer wise relevance propagation abbreviated as LRP in which they made use of the human's neural networks knowledge in the domain of audio. A gender classifier based upon a spectrogram was used to form a hypothesis and assess it to know about features that are used by the network. The networks' choices depend only upon a very small part of the data when classifications are done for raw waveform LRP.

Jacksi and team [15] built a Web-based application for attendance management which can be used to take attendance of students electronically to track the activity of students in class and further store their attendance in a database. Features were provided to give warnings to students for a specified period about their attendance percentage on the basis of the data analysis and attendance statistics about the student's absences [16]. Laravel Framework was used to develop an intuitive, attractive and user friendly graphical user interface.

3 Methodology

The methodology contains three major steps which are described as follows.

3.1 Data Gathering and Preprocessing

Data was collected from around 30 scholars of a class. Each one uploaded five of their front view pictures and five of their voice recordings speaking pre-defined sentences, four of which were chosen so as to cover all English alphabets and one sentence including their full name. Figure 1 shows graphical representation of an audio recording and Fig. 2 shows the Fourier transform.

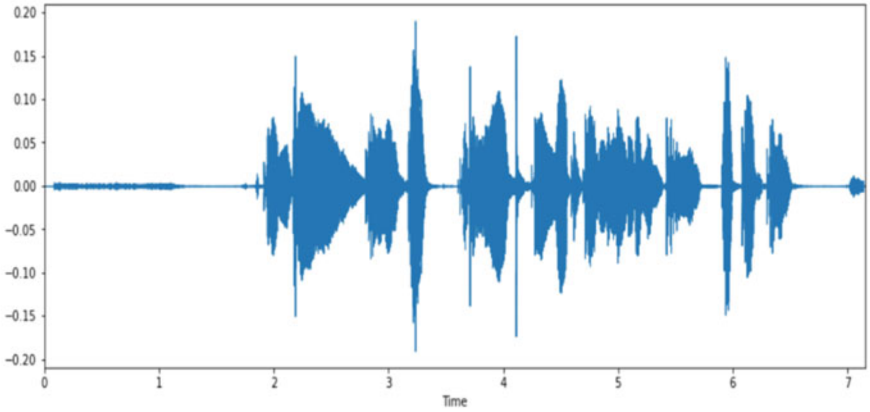


Fig. 1 Audio time series at a sampling rate of 4410

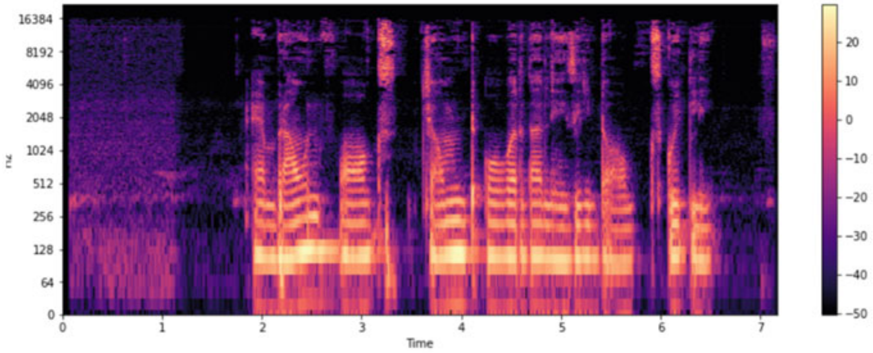


Fig. 2 Short time Fourier transform with log scale on *y*-axis

Figure 3 shows a voice recording converted to a spectrogram, which is a chart that shows the amount of each frequency at each time in an audio file. Spectrogram is a Fourier transform for small intervals of time obtained using the frequency time distribution of an audio. This chart has capability for identifying features of a person's voice to recognize them through the frequency energy distribution over time which is unique for each person.

The images of students were cropped to include just the face using the cascade classifiers and Haar features. Using the Haar features, a face is detected and a rectangular area is defined which encloses the face. This facial area is cropped and saved as a separate file which is later used to form a datapoint for model training. This was converted to grayscale and resized to a standard 500 × 500 pixel size. The captured audio in wav format was converted to spectrogram. The so-formed processed arrays were combined as shown in Fig. 4 to form a single datapoint and the process was repeated for each image and corresponding audio file.

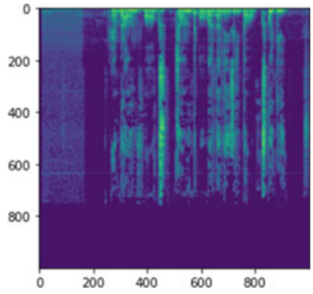


Fig. 3 Spectrogram obtained from an audio recording

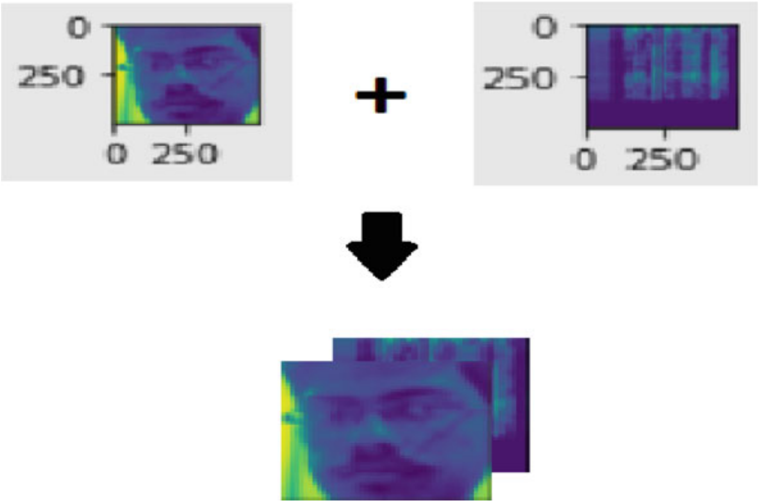


Fig. 4 Pre-processing overview: combining image and spectrogram into a data point

3.2 Model Building and Training

Convolutional Neural Networks work best with images. Having datapoints which are face in the form of image and even the sound in the form of image, CNN was decided to be used for building the model as CNN is capable of working with image and with image representations of sound as well.

A combination of three convolutional layers with maxpool layer was taken followed by a flatten and fully connected layers for building the CNN model as shown in Fig. 5. As O'Shea and Nash describes in [17], Due to similarity of neurons that self-optimise themselves through learning Convolutional Neural Networks (CNNs) are similar to primitive Artificial Neural Networks (ANNs). A countless number of ANN's are produced because of scalar product that is followed by a nonlinear function in which each neuron receives an input and will execute an operation. The weight is the single perceptive score function that is expressed by the entire network as the finalized result of the class score from the raw input image vectors. The classes in the last layer are dependent upon loss functions that are contained in the last layer of CNN and all of the common techniques employed by traditional ANNs work in the same way in CNN as well. The only significant variation of CNNs from primitive ANNs depends upon the feature of CNN that are mainly used for the recognition of patterns in an image. The parameters that are needed to make a working model are further reduced in CNN while at the same time making the network more accurate for image-focused tasks by making the architecture encoded with image specific features.

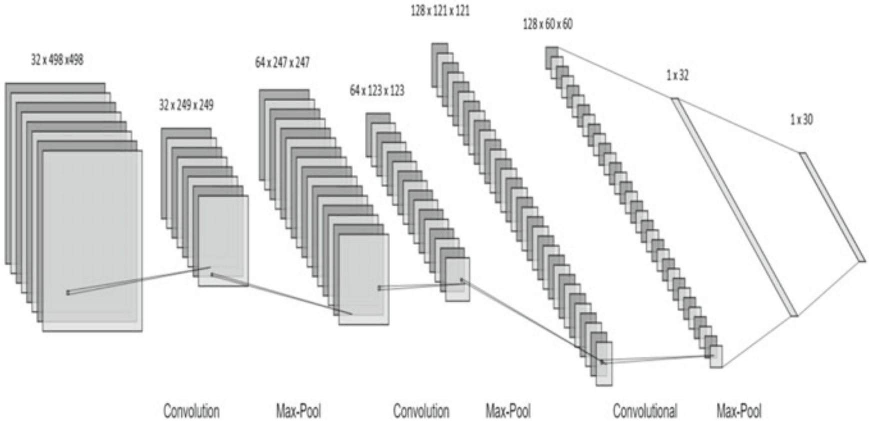


Fig. 5 CNN model architecture

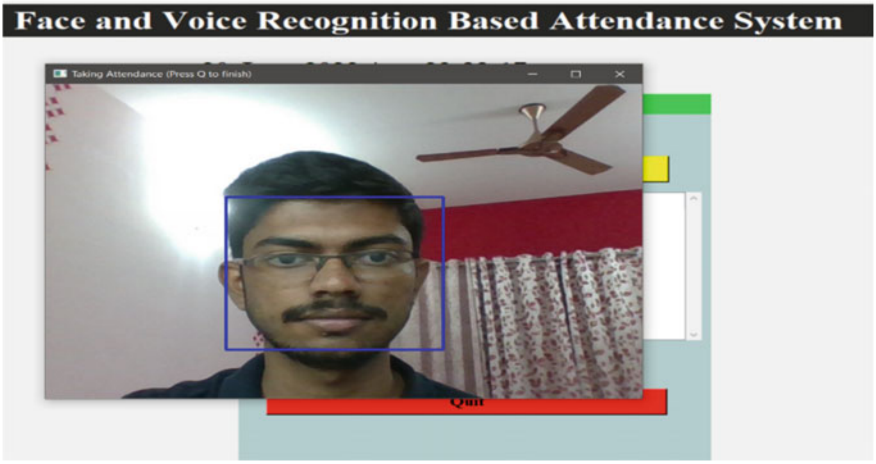
The pre-processed dataset was split into 80% training and 20% validation data ensuring that the latter includes each student's data point. Adam optimizer is used to calculate accuracy metric and the categorical cross entropy loss function is compiled with the data points that are inputted in this network. Category cross entropy is employed by multiclass classification problems as a loss function. It is used to determine the class of an example task out of all the possible classes. It shows the variation between the probability distributions of any two combinations of classes among the given classes. A stochastic gradient descent technique depending upon adaptive estimation of 1st order and 2nd order moments is Adam optimization. This technique works well because the network process single training example which is easier to fit in the memory also being computationally fast. Since the changes to the parameters are very frequent, it converges sooner for large dataset. These frequent updates helps to get out of the local minimum of the loss function, owing to the oscillations that happen because of the steps taken towards the minima of the loss function. According to Kingma et al. [16] this method is unaffected by diagonal rescaling of gradients, it has very small memory requirements, "computationally efficient" and is best for problems that are large in terms of data and parameters.

3.3 Interface Development

An interactive interface using python's Tkinter library was developed. The admin has access to log into the system and enable students entering the classroom or attendees entering the room to mark their attendance. The candidates can then capture their face as shown in Fig. 6, record audio and get a confirmation of their attendance being marked as shown in Fig. 7. Attendance of each session will be recorded in a folder named by the date of that day, containing comma separated files in the name of the course entered while marking the attendance. Ultimately, these files of each course will contain the ID and name of candidates along with the timestamp at which attendance was marked. Additionally, these files are then used to generate reports based on the queries entered into the system by the admin.

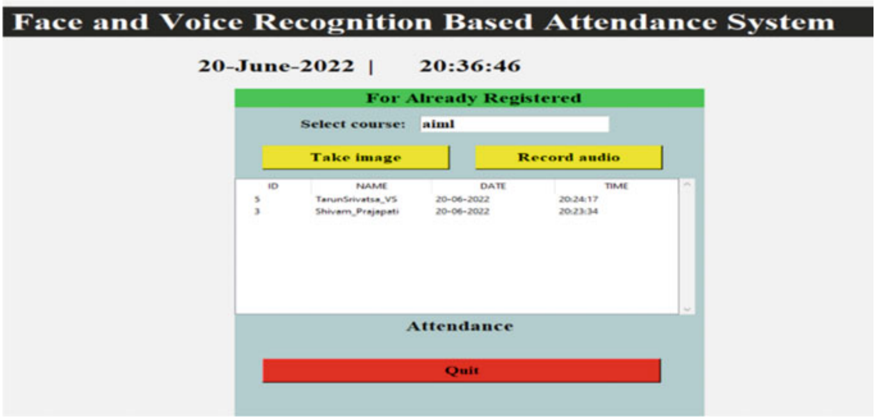
4 Results

The trained Convolutional Neural Network model achieved a combined accuracy of 95.45% and a validation accuracy of 82.35% when trained with a batch size of 32. The training loss and validation loss showed a steady decrease over the 25 epochs, over which training took place, indicating no overfitting. The learning rate taken was 0.01, with exponential decay rate of 1st moment as 0.9 and exponential decay rate of 2nd moment as 0.999 and a constant for numerical stability as 1e−7.



Face and Voice Recognition Based Attendance System

Fig. 6 Capturing face for marking attendance



Face and Voice Recognition Based Attendance System

20-June-2022 | 20:36:46

For Already Registered

Select course: aiml

ID	NAME	DATE	TIME
5	TarunSrivatsa_VS	20-06-2022	20:24:17
3	Shivam_Prajapati	20-06-2022	20:23:34

Attendance

Fig. 7 Interface for the portal

The training of the model presents a curve with increasing accuracy over the number of passes initially which eventually stabilizes near accuracy value of 1 and the curve flattens as shown in Fig. 8. More number of passes from that point will result in a possible overfit. Contrary to this graph trend, the loss vs epoch graph shows a steep decrease in both training and validation loss initially as the model is learning as shown in Fig. 9. The loss eventually stabilizes near the loss value of 0 and the graph flattens. Further passes will increase the validation loss while training loss decreases, indicating a possible overfit. Overfitting occurs when the model fits the training data too well. It learns the noise and random fluctuations in the data as features which in turn badly impacts the accuracy of the model on testing data.

After experimenting with multiple epoch ranges, different learning rates, batch sizes, optimizers and model architectures, the final training parameters were arrived at.

The user interface built to use the model had a smooth flow with user friendly and proxy avoiding aspects in the attendance management system. The overall system including the interface, storage and the model worked successfully in coherence when tested in real time.

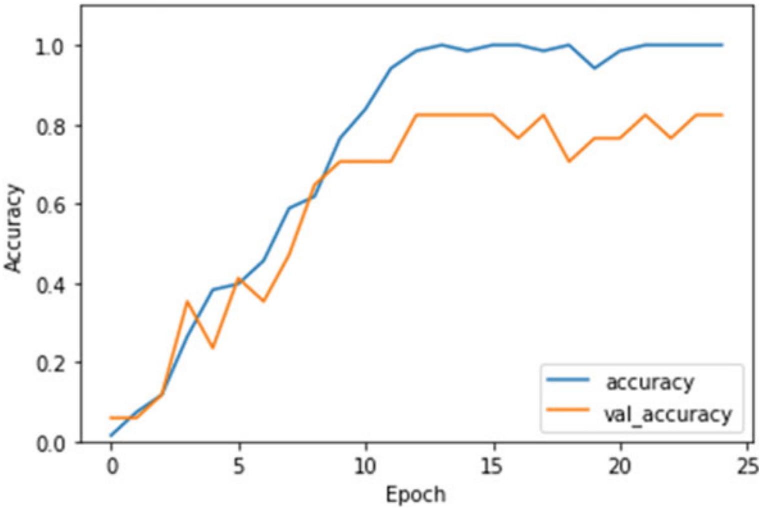


Fig. 8 Training and validation accuracies of the model over the epochs

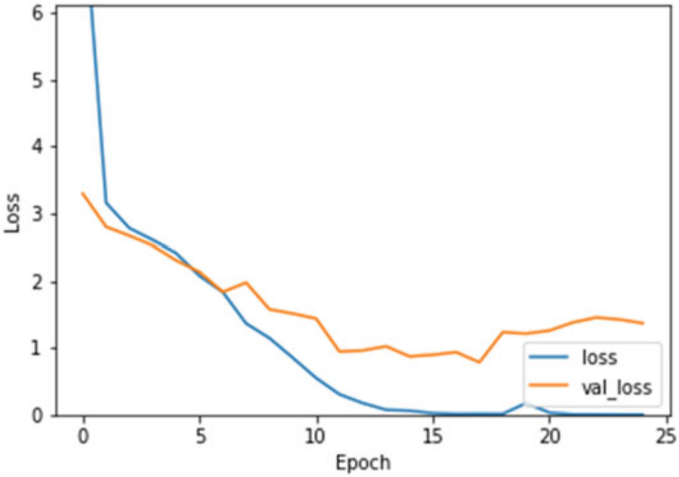


Fig. 9 Training and validation losses of the model over the epochs

5 Conclusion

Recognizing people uniquely with least effort and expense remains a vast field to explore having some critical applications in criminal cases. Attendance also employs a person's unique features to identify him/her. The modern day attendance systems can be innovated with the proposed methodology in this paper. With larger dataset and more powerful CNN models there remains a scope to further improve upon this approach.

References

1. Akbar MS et al (2018) Face recognition and RFID verified attendance system. In: 2018 international conference on computing, electronics and communications engineering (iCCECE). IEEE (2018)
2. Okokpujie KO et al (1999) Design and implementation of a student attendance system using iris biometric recognition. In: Foster I, Kesselman C (eds) 2017 International conference on computational science and computational ıntelligence (CSCI). IEEE, 2017 The grid: blueprint for a new computing ınfrastructure. Morgan Kaufmann, San Francisco
3. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, pp I-I. https://doi.org/10.1109/CVPR.2001.990517
4. Harikrishnan J, Sudarsan A, Sadashiv A, Ajai RAS (2019) Vision-face recognition attendance monitoring system for surveillance using deep learning technology and computer vision. In: 2019 ınternational conference on vision towards emerging trends in communication and networking (ViTECoN), pp 1–5. https://doi.org/10.1109/ViTECoN.2019.8899418
5. Harikrishnan J, Sudarsan A, Sadashiv A, Ajai RAS (2020) Face recognition based attendance management system. Int J Eng Res Technol (IJERT) 9(05). ISSN: 2278-0181
6. Khan S, Hammad Javed M, Ahmed E, Shah SAA, Umaid Ali S (2019) Facial recognition using convolutional neural networks and ımplementation on smart glasses. In: 2019 ınternational conference on ınformation science and communication technology (ICISCT), pp 1–6. https://doi.org/10.1109/CISCT.2019.8777442
7. Yuan X, Li G, Han J, Wang D, Zhi T (2021) Overview of the development of speaker recognition. J Phys Conf Series, ICETIS 2021, 1827:012125.https://doi.org/10.1088/1742-6596/1827/1/012125
8. Salehghaffari H (2018) Speaker verification using convolutional neural networks. In: arXiv:1803.05427v2 [eess.AS] 10 Aug 2018
9. Shah J, Salunkhe V, Saturwar J, Parab O (2020) Voice ınput based attendance system. In: International journal of recent technology and engineering (IJRTE). ISSN: 2277-3878 (Online), vol 9, Issue 1
10. Dhaya R (2021) Efficient two stage ıdentification for face mask detection using multiclass deep learning approach. J Ubiquit Comput Commun Technol 3(2):107–121
11. Sathesh A (2019) Typıng eyes: a human computer ınterface technology. J Electron Inf 1(2):80–88
12. Wang M, Sirlapu T, Kwasniewska A, Szankin M, Bartscherer M, Nicolas R (2018) Speaker recognition using convolutional neural network with minimal training data for smart home solutions. In: 2018 11th international conference on human system ınteraction (HSI), pp 139–145. https://doi.org/10.1109/HSI.2018.8431363
13. Gomes J, Fernandes H, Abraham S, Chavan S (2021) Person identification based on voice recognition. In: 2021 4th biennial ınternational conference on nascent technologies in engineering (ICNTE), pp 1–5. https://doi.org/10.1109/ICNTE51185.2021.9487756

14. Becker S, Ackermann M, Lapuschkin S, Müller K-R, Samek W (2019) Interpreting and explaining deep neural networks for audio signal classification. In: arXiv:1807.03418v2 [cs.SD] 22 Oct 2019
15. Jacksi K, Ibrahim F, Ali S (2018) Student attendance management system. In: Scholars J Eng Technol (SJET). ISSN: 2347-9523 (Print). ISSN 2321-435X (Online). Published: 15.02.2018. https://doi.org/10.21276/sjet.2018.6.2.1
16. Kingma DP, Ba J, Adam: a method for stochastic optimization. In: arXiv:1412.6980 [cs.LG]
17. O'Shea K, Nash R (2015) An ıntroduction to convolutional neural networks. In: arXiv:1511.08458v2 [cs.NE] 2 Dec 2015

Blockchain-Enabled Network for 6G Wireless Communication Systems

Nazanin Moosavi and Hamed Taherdoost

Abstract 6G wireless network is going to revolutionize wireless systems by introducing several innovative services such as virtual reality (VR), 16K Video, Vehicle to Vehicle communication, and Internet of Everything (IoE) on a commercial scale to increase end-user experiences. Hence, network infrastructure needs upgrading to provide higher data rates, massive connectivity, and more secure wireless systems to meet the use case requirements. Distributed ledger technology and blockchain, which is known to be a disruptive technology enabler, can address the challenges and functional needs of 6G technology. In this work, we investigate the opportunities of those blockchain-enabled services in the 6G network, along with the shortcomings and limitations that need to be discussed in further researches.

Keywords Blockchain · Wireless communication · 5th generation (5G) mobile network · 6th generation (6G) mobile network · Wireless network · Distributed ledger technology

N. Moosavi
Hamta Group|Hamta Business Corporation, Vancouver, Canada
e-mail: nazanin@hamta.ca

H. Taherdoost (✉)
University Canada West, Vancouver, Canada
e-mail: hamed.taherdoost@gmail.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_62

1 Introduction

The fifth-generation (5G) network is now available in some countries around the world. 5G services are divided into three main groups eMBB for Enhanced Mobile Broadband, mMTC for massive Machine Type Communication, and URLLC for Ultra Reliability and Low Latency Communication. However, the initial goals for 5G networks are not realized as an example massive internet of things (IoT) networks which was supposed to yield the way to IoE are not available with sufficient scale [1].

As a result, academic research projects are working to shape the 6G technology. The 6th generation of the mobile network is supposed to have higher data rates with lower latency and in advance, it can propose diverse services and innovative applications. In comparison with 5G, 6G data rate is around 1 Tb/s which is around 1000 times larger than 5G, its end to end delay is less than 1 ms versus 5G being around 5 ms, its mobility is 1000 km/h versus 5G being 500 km/h and its reliability is around 10−9 which is 104 better than 5G [2].

6G security and privacy have gained a lot of attention in recent studies. Since 6G is going to propose various services in both industry and high-end users, satisfying security, privacy, and reliability KPIs is of high importance, which requires embedding a trust model into the network. This model should have coordination between different parts of the network, from network entities and providers to end-users [3, 4].

Recently blockchain technology has attracted a lot of attention in both industry and academia. Some of the main specifications of blockchain technology are offering decentralized networks, transparency, immutability and irreversibility [5]. Bitcoin, which is the most famous blockchain network, become popular these years since it makes an evolution to financial systems. However, blockchain can offer solution to many industries such as energy, agriculture, neurosceience and telecommunication [6]. Consequently, blockchain can be regarded as a promising solution to offer trust and security for diverse applications in 6G wireless network [7].

In this paper, we present some applications of blockchain in 6G networks for both technology improvement such as resource management and spectrum sharing, network virtualization, edge computing and artificial intelligence and in different use cases such as industrial application, smart healthcare, vehicle-to-vehicle communication, Unmanned Aerial Vehicles (UAVs) and smart grid. In addition, we discuss the challenges that may encounter when using blockchain technology in 6G networks.

The rest of the paper is organized as follows: Section 2 outlines some general issues in the 6G network, Sections 3 introduces the blockchain technology, Sections 4 proposes the blockchain solutions for 6G services, Sect. 5 reveals the possible challenges of blockchain in 6G and finally Sect. 6 concludes the paper.

2 6G General Specification

In this section, we discuss some early features that 6G network is going to propose. It is noted that Understanding these features gives us a good insight into how 6G services stand to gain from blockchain-based proliferation [8].

2.1 Higher Data Rate

Data rate is defined by the number of bits transmitted per second in the network. It is usualy showed by bits or bytes per second and described the speed of data transmission in a wireless network. Introducing higher data rates is one of the main features of each generation of wireless communication. With the advent of new use cases in 6G such as autonomous vehicles, 16 K video, and Virtual Reality (VR) services need more data rate with more data consumption, which leads the way to more enhanced network infrastructure and optimization.

2.2 Massive Number of Users

Massive machine type communication especially in the industrial IoT solutions require a strict network design to handle a massive number of users and unprecedented data traffic. In addition, machine-type communication needs a more secure design to avoid data leakage.

2.3 Security Requirements

The future wireless networks for example IoT services may expose many security threats. Due to a massive number of devices, encryption techniques should change to become lightweight. However, these lightweight symmetric keys are subject to privacy risks in the network. Besides, the large volume of data will transmit between the nodes in the network so the eavesdropping may be more which affects the integrity of the system. In addition, system availability, which is one main feature of 6G network, may increase the risk of DDoS attacks. All of these challenges show that 6G systems need an upgraded version of network security tools.

3 Blockchain Terminology

A blockchain network is a series of blocks that holds transactional records through distributed ledgers like public ledgers. In recent years, digitalized ledgers have been used to store data with centralized ownership, but blockchain technology introduces a distributed way to store data records [5]. Blockchain is a sequence of some blocks that produces a chain. Each block in a blockchain sequence contains two parts, block header, and block data, which are transactional data, stored in each block [7]. Figure 1 shows the block sequence and the relationship between blocks and the block parts are as follow:

- Block version—4 bytes: Indicates some rules to validate the block
- Merkle tree hash—32 bytes: Hashing method to represent hash of block data
- Parent block hash—32 bytes: The hash value of the previous block
- Timestamp—4 bytes: The block creation time since 197001-01T00:00 UTC
- nBits—4 bytes: Current hashing target in a compact form
- Nonce-4 bytes: A number which is varied to create a unique hash for each block data and usually starts with zero
- Data-32 bytes: transaction counters and transactions.

3.1 Consensus Protocols

A new block can be added to the network when all the users in the network verify it by utilizing a consensus protocol, which is a set of rules that should be followed during the transactions. Proof of work (PoW) is one famous consensus protocol that is used in the Bitcoin network. Since it uses a lot of computing power, it is not environmentally friendly. The number of transactions per second for Bitcoin's PoW is around 7 and it takes 10 min for a new block to be confirmed. Since it has 6 block confirmation latency, each node shoud wait around 1 h to confirm transactions [9]. Hence there are other kinds of consensus algorithms with less limitations such as

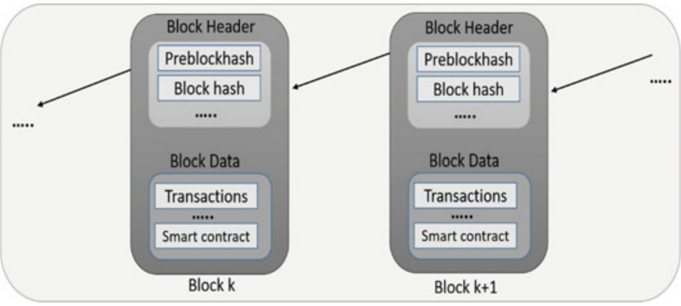


Fig. 1 Block sequences

Table 1 Comparison between different consensus protocols

Consensus algorithm/use cases	Latency	Transaction	Advantage	Disadvantage
Proof of-Work (PoW)/Bitcoin, Ethereum	6–60 min	10 s	Full decentralization	Low energy efficiency
Proof of-Stake (PoS)/Peercoin	10–60 min	10 s	High energy efficiency	Threat to security
Byzantine fault tolerance(BFT)/ Hyperledger	1–6 s	1000 s	High efficiency	Higher communication complexity[1]
Delegated Proof of-Stake (DPoS)/EOS	≤1 s	≥1000 s	Higher energy efficiency, vulnerability	Less decentralization

proof of stake (PoS) which are more used in blockchain application-based services [10–12]. Table 1 gives a comparison for a different consensus algorithm. It is noted that choosing the right protocol depends on each use case requirements and should be considered precisely.

3.2 Blockchain Classification

Blockchain network is classified into two models based on the permission to add a new block to the chain. If any user in in the network is able to add a new block, the blockchain is permissionless and if only a few users with authority are able to add a new block to the network the blockchain is permissioned. In simple words, a permissionless blockchain is like a public internet where anyone can access it, but a permissioned blockchain is like a corporate Intranet that is under control and is often used by groups or organizations [13]. A public blockchain is one kind of permissionless blockchain where anyone at any time can enter or exit the network and any user can involve in the consensus mechanism. Private Blockchain and consortium blockchain are two main categories of permissioned blockchain. Former is for a group that just some authorized users have exclusive access to the blockchain and the latter is a network that only some pre-selected nodes can join the consensus mechanism [5].

[1] Since PBFT algorithm provides heavy system overhead and decreases the consensus efficiency, its communication complexity is high.

3.3 Smart Contracts

One of the main technological innovations of blockchain is smart contracts, which is an effective and significant feature to use with other technologies like 6G. Smart contracts are predefined roles and programmable codes. It is software-defined contracts between users in the network that represents terms of agreements triggered automatically when certain conditions are met. It can be applied in many use cases such as industry, insurance, telecom and energy trading [14].

3.4 Blockchain Features

Blockchain features are defined as follow [3, 5, 15–17]:

- **Decentralization**: It means that there is no central authority to make a decision. Everyone in the network can access the information and users can make a direct transaction which each other without the need for third-party.
- **Privacy**: Privacy or transparency means that the real identity of users is not shown to others and is secured [17].
- **Immutability**: When the data enters the blockchain, it cannot be deleted or edited so it is tamper-proof.
- **Distributed ledger**: All the database is broadcasted to all the users in the network so each of users has the same copy of the database.
- **Irreversibility**: It means that once a task is done it cannot be retrieved.
- **Auditability**: Since all the records of transations are in the distributed ledger, it is possible to audit past records by accessing one node in the network [5].

4 Blockchain Solutions for 6G

Blockchain is known as an innovative and disruptive technology that can solve security and trust issues in wireless communication networks, paving the way for more creative services and applications. In the first section, 6G technology services that can benefit from blockchain are proposed, and then some blockchain-enabled 6G applications are given.

4.1 Blockchain-Enabled 6G Technology

Resource Management: Resource management is a challenging issue in 6G massive network with many connections and various kinds of services. Network resources such as spectrum and infrastructure are limited and should be used more efficiently.

Resource sharing allows multiple categories of users to share their spectrum or network infrastructure. Sharing resources needs an open market where every part of the network from the end-users to the spectrum providers, infrastructure providers, and the service provider can safely exchange their resources.

Consequently, all these resources need to be shared between the networks to make the best use of them without any security concerns. Blockchain can offer a promising secure solution for network sharing between entities by offering smart contracts which is representing terms of agreements triggered automatically when certain conditions are met [18, 19].

Network Virtualization: There are new kinds of cloud processing in the wireless networks such as software-defined network (SDN) or network function virtualization (NFV) which increases the network resource for example computing and storage resources. NFV can be used by mobile operators to virtualize some network functions such as routing, load balancing and policy management to be transferred to virtual servers and SDN is a solution that uses software-based controllers to communicate with hardware infrastructure. By means of network virtualization, the number of physical servers is reduced which results in less electrical consumption and so energy efficiency increses and hence capacity. Network softwerization and virtualization can provide a level of abstraction which improvise not only cost effectiveness, but also the end-to-end reliability in 6G network. Since the security functions are implemented in the software, it provides some new security challenegs that can be addressed by blockchain [3, 4]. By enabling SDN and NFV, the network slicing technique can happen easily. Network slicing technology enables end-to-end control of the network for special use cases, which can be easily provided by service providers on top of a shared network to support different applications in various verticals and industries. As an example, one slice can be used for low latency vehicular communication, one slice can be used for high bandwidth 4 K video and one slice is optimized for low power IoT network. Blockchain can help for immutability and security of managing and sharing virtual network slices [20, 21].

Edge Computing: Cloud processing and edge computing make computing capacities valuable resources in the network. Blockchain can help to provide a secure network to manage and share resources efficiently. Also by means of edge, computing heavy computational processes are offloading to remote servers. Since this computation may include sensitive information, guaranteeing privacy is so important. Blockchain technology can make trust between users and remote edge servers [22, 23].

Artificial Intelligence (AI): AI-enabled 6G promotes intelligent services, which produces a large amount of data and storage capacity. This huge data is more vulnerable to attackers. Blockchain can add more security levels to the system. Besides, by integrating AI and blockchain in 6G services, the quality of intelligent services is more optimized [24]. Consequently in summery blockchain benefits 6G services by:

- Build trust between users and servers
- Guruntee integrity of remote servers

- Enhance security for spectrum management and sharing
- Gurunteе trust between providers and market users
- Eliminate the third parties and intermediaries
- Provide more security level to the system.

4.2 Blockchain-Enabled 6G Use Cases

Here are some 6G technology use cases that can benefit from blockchain technology in Fig. 2.

Industrial Application: Industrial applications especially beyond industry 4 applications are known to play an important role in 6G use cases. Blockchain can meet the needs for decentralized architecture in massive connectivity or remote maintenance of industrial applications [25].

Smart Healthcare: The application of blockchain in smart healthcare has attracted a lot of attention these years. With a lot of advancement in wearable devices, remote monitoring is accessible through healthcare data. Following the maintenance of medical records, utilizing blockchain to address the privacy of users' data is became the most important case study [26]. Authors in [27] introduces an Ethereum smart contracts project which aims to record the health data of users with blockchain technology.

Vehicle-to-Vehicle Communications: 6G network will provide a great opportunity for Vehicle-to-Vehicle network technology since it ensures service availability for example by combining satellite communication with current wireless communication. Blockchain can provide a trustful solution for vehicle management [28].

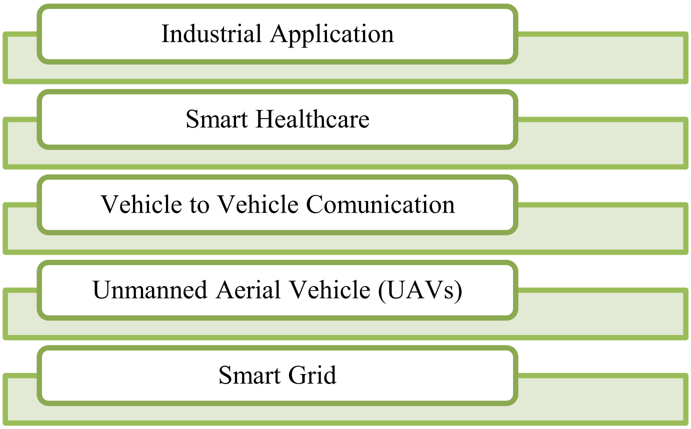

Industrial Application
Smart Healthcare
Vehicle to Vehicle Comunication
Unmanned Aerial Vehicle (UAVs)
Smart Grid

Fig. 2 Blockchain-enabled 6G use cases

Unmanned Aerial Vehicles (UAVs): UAVs play an important role in future wireless application. They provide coverage in non-accessible areas with low latency. Although they are more vulnerable to attacks since the information gathered by them is so attractive for cyber-attacks. Blockchain technology can provide solutions for security, air traffic, and insurance of UAVs [29].

Smart Grid: Smart grid is introduced as a new technology to provide energy anywhere and at any time by utilizing distributed decentralized energy providers. The transformation to a decentralized mechanism from the traditional centralized mechanism in the smart grid needs a proper and secure wireless connection, which can be addressed by 6G network. However, it has a some to be considered. Consequently, blockchain is a proper technology to address these issues and can provide secure connections for smart grid users [30, 31].

5 Challenges

It is obvious from the previous section that blockchain technology has an important effect on improving 6G network services. However, it should be noticed that integrating blockchain into 6G might face some challenges and issues that should be considered precisely. Here is the description of some challenges.

5.1 Storage

Due to the decentralized nature of blockchain technology, each node should have the full copy of all the networks. It needs a lot of storage capacity, which affects the massive connectivity of 6G, especially in the IoT solutions where each node cannot save many transactions. Hence applying blockchain technology to IoT services needs further investigation especially selecting the right consensus algorithm is of high importance. Also, it should be noticed that a large storage capacity costs a lot which is not proper for many use cases [14].

5.2 Delay

Blockchain technology requires a high number of messages being passed and broadcasted between users. Hence, it may add some delays to the network. This is not suitable for some delay stringent applications unless there would be a trade-off between delay and the privacy that blockchain would guarantee.

5.3 Security Risks

Although blockchain can improve security and privacy in the network by containing transparency and immutability features, there are some security problems in this technology that should be considered when adopting it as a core technology in 6G. Here are some security risks to be considered. Majority attack: This attack happens when attackers try to control the main processing power of the network. For the case of PoW is it 51% which means when one group has 51% of the system power processing, the network is under control of them, and they can reverse the transactions and reverse the blockchain.

Double Spending: By this attack, one user aims to break the integrity of the blockchain. Particularly, in cryptocurrencies, it happens when one user completes two different transactions with just the same amount of currency.

Privacy Leakage: When a transaction takes place, there is no information about the user's real identity. However, some studies reveal that since all the transactions are apparent to the users, the transactional privacy of the users is not guaranteed. Besides, there are some methods to find the user IP address which is not favorable for many users [13].

5.4 Scalability

Two main problems of the most famous blockchain network, Bitcoin are its low throughput and high transaction latency. These two problems affect the scalability of blockchain-enabled services, especially in 6G network that delay, and throughput are so important. However, there are some researches on finding solutions to improve the scaling problem in blockchain network but further investigation needs to be done [32].

6 Conclusion

In this paper, at first, we introduced some general issues of 6G technology. Then blockchain is proposed as a disruptive solution to enhance 6G services. Hence, we studied the usage of blockchain in 6G network. We outlined a list of 6G services that can be improved by blockchain technology and then we discussed the challenges they may counter by using blockchain in different 6G services and applications. In conclusion, it should be noted that blockchain can help the growth of 6G applications by giving proper solutions for security attacks and privacy issues, but further investigation should be considered to mitigate its challenges.

References

1. Saad W, Bennis M, Chen M (2019) A vision of 6G wireless systems: applications, trends, technologies, and open research problems. IEEE Netw. 34(3):134–142
2. Nguyen T, Tran N, Loven L, Partala J, Kechadi M-T, Pirttikangas S (2020) Privacy-aware blockchain innovation for 6G: challenges and opportunities. In: 2020 2nd 6G wireless Summit (6G SUMMIT), 2020. IEEE, pp 1–5
3. Wang M, Zhu T, Zhang T, Zhang J, Yu S, Zhou W (2020) Security and privacy in 6G networks: new areas and new challenges. Digit Commun Netw 6(3):281–291
4. Lu Y (2020) Security in 6G: the prospects and the relevant technologies. J Ind Integr Manag 5(03):271–289
5. Taherdoost H (2022) A critical review of Blockchain acceptance models—Blockchain technology adoption frameworks and applications. Computers 11(2):24
6. Taherdoost H (2022) Neuroscience and Blockchain. Arch Neurol Neurosci 2022. https://doi.org/10.33552/ANN.2022.12.000794
7. Rajasekaran AS, Azees M, Al-Turjman F (2022) A comprehensive survey on blockchain technology. Sustain Energy Technol Assess 52:102039
8. Hewa T, Gür G, Kalla A, Ylianttila M, Bracken A, Liyanage M (2020) The role of blockchain in 6G: challenges, opportunities and research directions. In: 2020 2nd 6G wireless Summit (6G SUMMIT). IEEE, pp 1–5
9. Lepore C, Ceria M, Visconti A, Rao UP, Shah KA, Zanolini L (2020) A survey on blockchain consensus with a performance comparison of PoW, PoS and pure PoS. Mathematics 8(10):1782
10. Xiao Y, Zhang N, Lou W, Hou YT (2020) A survey of distributed consensus protocols for blockchain networks. IEEE Commun Surv Tutorials 22(2):1432–1465
11. Maksymyuk T et al (2020) Blockchain-empowered framework for decentralized network management in 6G. IEEE Commun Mag 58(9):86–92
12. Qiao L, Dang S, Shihada B, Alouini M-S, Nowak R, Lv Z (2021) Can blockchain link the future? Digit Commun Netw
13. Zheng Z, Xie S, Dai H-N, Chen X, Wang H (2018) Blockchain challenges and opportunities: a survey. Int J Web Grid Serv 14(4):352–375
14. Ling X, Wang J, Bouchoucha T, Levy BC, Ding Z (2019) Blockchain radio access network (B-RAN): towards decentralized secure radio access paradigm. IEEE Access 7:9714–9723
15. Hoffman MR, Ibáñez L-D, Simperl E (2020) Toward a formal scholarly understanding of blockchain-mediated decentralization: a systematic review and a framework. Front Blockchain 3:35
16. Click K, Singh A, Parizi RM, Srivastava G, Dehghantanha A (2020) Immutable and secure IP address protection using blockchain. In: Blockchain cybersecurity, trust and privacy. Springer, pp 233–246
17. Suma V (2019) Security and privacy mechanism using blockchain. J Ubiquit Comput Commun Technol (UCCT) 1(01):45–54
18. Maksymyuk T, Gazda J, Han L, Jo M (2019) Blockchain-based intelligent network management for 5G and beyond. In: 2019 3rd international conference on advanced information and communications technologies (AICT), 2019. IEEE, pp 36–39
19. Haavisto J, Arif M, Lovén L, Leppänen T, Riekki J (2019) Open-source RANs in practice: an over-the-air deployment for 5G MEC. In: 2019 European conference on networks and communications (EuCNC), 2019. IEEE, pp 495–500
20. Javed F, Antevski K, Mangues J, Giupponi L, Bernardos CJ (2022) Distributed ledger technologies for network slicing: a survey. IEEE Access
21. Xu H, Klaine PV, Onireti O, Cao B, Imran M, Zhang L (2020) Blockchain-enabled resource management and sharing for 6G communications. Digit Commun Netw 6(3):261–269
22. Dai Y, Xu D, Maharjan S, Chen Z, He Q, Zhang Y (2019) Blockchain and deep reinforcement learning empowered intelligent 5G beyond. IEEE Netw 33(3):10–17

23. Guo S, Dai Y, Xu S, Qiu X, Qi F (2019) Trusted cloud-edge network resource management: DRL-driven service function chain orchestration for IoT. IEEE Internet Things J 7(7):6010–6022
24. Li W, Su Z, Li R, Zhang K, Wang Y (2020) Blockchain-based data security for artificial intelligence applications in 6G networks. IEEE Netw 34(6):31–37
25. Zhang Z et al (2019) 6G wireless networks: vision, requirements, architecture, and key technologies. IEEE Veh Technol Mag 14(3):28–41
26. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A (2020) Blockchain in healthcare and health sciences—a scoping review. Int J Med Inf 134:104040
27. Ekblaw A, Azaria A, Halamka JD, Lippman A (2016) A case study for Blockchain in healthcare:“MedRec” prototype for electronic health records and medical research data. In: Proceedings of IEEE open & big data conference, vol 13, p 13
28. Khan AS, Balan K, Javed Y, Tarmizi S, Abdullah J (2019) Secure trust-based blockchain architecture to prevent attacks in VANET. Sensors 19(22):4954
29. Gupta R, Nair A, Tanwar S, Kumar N (2021) Blockchain-assisted secure UAV communication in 6G environment: architecture, opportunities, and challenges. IET Commun 15(10):1352–1367
30. Mollah MB et al (2020) Blockchain for future smart grid: a comprehensive survey. IEEE Internet Things J 8(1):18–43
31. Madhura S (2020) A secure protocol for smart meters using IoT enabled distribution networks and blockchain security mechanism. J Ubiquitous Comput Commun Technol (UCCT) 2(01):48–58
32. Zhou Q, Huang H, Zheng Z, Bian J (2020) Solutions to scalability of blockchain: a survey. IEEE Access 8:16440–16455

Machine Learning Based Automated Disaster Message Classification System Using Linear SVC Algorithm

N. Merrin Prasanna, S. Raja Mohan, K. Vishnu Vardhan Reddy, B. Sai Kumar, C. Guru Babu, and P. Priya

Abstract This paper presents the machine learning based automated disaster message classification system. Machine learning is be used to identify such information and provide valuable information for aiding disaster response during emergency events. Disaster management prediction systems (DMPS) are computer systems for determining when and where to deploy mitigation measures in the event of an emerging natural or man-made hazard, while accounting for and mitigating human factors that may compromise operational effectiveness for providing fast services to handle this high volume and velocity of urgent information. Till date diverse techniques used for disaster and pandemic management are available using the technologies like satellite-based systems, cellular networks, Internet of things (IoT), smartphone-based systems, 5G and cellular networks. Linear support vector machines are an efficient way to learn discriminative models, which is especially useful in data where the number of attributes is large or is not known. Linear SVC (Support Vector Classifier) is one of the most successful linear models, not only because it is quite fast to train and compute, but also because it can achieve excellent performance in high dimensional problems. Linear SVMs can make use of a wide variety of learning algorithms.The proposed work uses Linear SVC Algorithm with strategy of self-training that learns from available datasets with the labeled data. Finally, the paper gives the message classification based on the emergency to the relevant disaster.

Keywords SVC · Disaster management · Machine learning

N. Merrin Prasanna (✉)
Associate Professor, Department of ECE, Annamacharya Institute of Technology and Science, Rajampet, Annamayya, Andhra Pradesh, India
e-mail: nmp@aitsrajampet.ac.in; nagadasari.merrin@gmail.com

S. Raja Mohan · K. Vishnu Vardhan Reddy · B. Sai Kumar · C. Guru Babu · P. Priya
Software Engineer, HCL, Bangalore, India

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*, Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_63

1 Introduction

Due to sudden disasters many people lost their lives and left with injuries to avoid such condition disaster are to be predicted, detected and human evacuation should be done very fast. Emergency Rescue Evacuation Support System (EREES) is an American company that develops, builds, and installs advanced communications systems that allow fire departments to communicate with each other and with emergency response agencies during fire, emergency medical and other rescue missions. the term may be applied to events as various as tsunamis, natural floods, hurricanes, earthquakes, avalanches, volcanoes, or the like, though the term usually describes a natural disaster that results in a significant amount of human loss of life [1]. The term may also be applied to man-made disasters such as nuclear disasters. According to the Centre for Research on the Epidemiology of Disasters, the most frequent cause of natural disasters is weather. Over half of the natural disasters in the last 50 years, and more than half of the losses from natural disasters are estimated to have been caused by just 3% of causes: “storms, earthquakes, and volcanoes”. The World Health Organization has reported that between 1981 and 2002 there were more than 50 million deaths directly or indirectly caused by earthquakes. The World Meteorological Organization has stated that earthquakes are the leading natural cause of death worldwide. In 2010, earthquakes accounted for 17% of deaths and disasters. Most of the world’s deadliest disasters have been attributed to earthquakes. As far as economic losses are concerned, over half of the costs are incurred because of economic damages from natural disasters, the majority of which are due to the loss of property. These include the direct loss of property (including losses in human life) and indirect losses (e.g. destruction of property, losses to business assets, and damage to public infrastructure, to name a few). The term disaster can also be used for human activities that lead to loss of life, and this is especially true of war and terrorism. The definition of a disaster differs according to the context. There are various definitions based on the type of disaster: earthquake, flood, tsunami, volcanic eruption, etc. As described in [Disaster Management: An introduction to response and recovery. Routledge, London.], a disaster is defined as: an event or chain of events that causes great harm, loss, damage, or destruction; a calamity. something that is extremely bad, frightening, or harmful, generally involving death or injury, a calamity. a state of being harmed, injured, or killed by a severe accident, disaster, or catastrophe, or by sudden illness, injury, or death; harm; mishap; as a consequence of a catastrophe; to suffer harm or loss. an event that causes unexpected loss or damage, a misfortune. According to [“A study of current and historical definitions of disaster.” Disaster management and prevention. Springer-Verlag, London.], a disaster is defined as: A calamitous event that causes considerable human or material loss or damage that is not the normal outcome of a complex social system. An event that involves considerable suffering of human lives or loss of property and infrastructure, but which can be anticipated and usually does not cause total loss of the affected community. “Disaster”, “calamity”, and “accident” are terms with similar but different meanings. Whereas “calamity” is “a misfortune that befalls a group of people or a place”, “disaster” is a “calamity that befalls an entire society, especially a very large and populous one". A calamity may be called a disaster, but not every disaster is a calamity. A natural disaster (or natural catastrophe) is an event with severe effects resulting from natural causes, particularly those resulting from weather, and especially those caused by meteorological events or their immediate result: earthquakes, tornadoes, hurricanes, and tsunamis. Disasters or catastrophes may be categorized as natural, social, or technological [2]. The first type is caused by natural events such as earthquakes, volcanoes, tornadoes, hurricanes, flooding, and other natural disasters that cause loss of life and property. ML algorithms are often able to perform predictive analysis using statistical models based on large datasets, but only require a few minutes to hours to complete processing. The main components of ML algorithms are the model, the data base, and the software that operates on the data and produces the model. The model is used to predict future values for output variables given the input variables. The models are constructed based on the data and may be built using either supervised or unsupervised learning [3].

To create models that are useful for predicting and acting on future events, ML algorithms may need to learn from samples (a sample is just a subset of the data used for training). Each sample is associated with a target variable and several input variables. If more input variables are available, ML algorithms may be able to identify relationships among the variables more quickly. ML algorithms can be "unsupervised" or "supervised" in how they learn from samples. Unsupervised algorithms use the relationship between input and target variables to create a model. Supervised algorithms may need more input information about the target variable. Because ML algorithms can work on many types of data, you may need to adapt your data to fit the type of ML algorithm that you use. For example, a model may use only continuous data. You can convert your data into continuous variables, but also into variables based on the categories of the data. You may need toadjust the data before modeling [4]. You may need to build your own data collection, transformation, and preprocessingsystem. At the time of disaster emergency, the responders and the helping organizations or the teams uses their own organizational polices and mechanism for responding the crises. All the organizations have standard operating procedures, command structure centralized and internal vetting standards for attaining the emergency responses. While are those mechanisms are not optimized to the current needs and expectations of efficiency, speed, knowledge but still recovered millions of people but to make it 100%, systems need to be improved to current speed expectation and efficiency using the machine learning algorithms [5]. The responder can help the people based on the big crisis data and prioritize the information.

2 Literature Survey

Sirinrat Khwanpheng et. al. proposed a system for cross checkingthe information collecting from the multiple sources and the victims are sorted on priorities based. The proposed system got the moderate satisfaction and with adequate precision [6].

Improved disaster management distribution is proposed by authors Zarei et al. [7] to disaster base maintenance method that optimizes the breakage rate, minimum outage period. The survey on disaster management and the role of artificial intelligence as proposed by the authors Nunavath and Goodwin in the paper [8]. In all the techniques of computing technologies applied in day-to-day life as well as diverse domains of the industries is machine learning (ML), it as applications of the artificial intelligence that works with the algorithms on the data available that are used to predict the future [9]. There are various the algorithms are employed in ML to make the fastest reliable decisions that have wide variety of the applications that reduces the human interventionsmachine learning algorithms and research is focused on disaster management, major area is flood disaster and earthquakes. ML models are used to help understand underlying relationships in datasets. One way to do this is to look at statistical relationships. Another is by using expert models. Expert models can be based on other expert models or on the relationships that you have created, which in turn can be derived from the relationships that you have already created. If you don't know what the relationships between variables are or how to make these relationships, you should build these models [10]. The advantages of using the Machine learning Machine-agnostic: Each application is developed for one or a few machines, not a whole range of machines. High performance algorithms [11] can be executed in an "embarrassingly parallel" manner on multiple machines. Modularity: Each application can use the ML algorithms, and the ML framework for all classes of problems. Transparency: Application developers can see the results of training their ML algorithms and debug them as needed. Scalability: Algorithms that learn faster tend to be more scalable. The algorithms that are mainly and widely used in the prediction and detection are Support Vector Machines (SVMs) [12] are supervised machine learning algorithms that provide a technique for classification by finding an optimal hyperplane that separates data into two classes. The algorithm involves the use of linear decision boundaries, which can provide high quality results with minimal over fitting. SVMs tend to have limited classification problems and are not capable of handling nonlinear problems. To improve disaster management, we can consider supervised neural networks. We have shown that we can improve the prediction performance of neural networks by using the ensemble of tree models. When we use supervised neural networks, we need to prepare a data set that describes a possible disaster. The data set must be generated by a computer simulation. For instance, we use the following data set, and our goal is to predict the death count by the flood, wind, and fire. IoT and ML-based Models for Disaster Prediction are proposed by Chen et al. [10]. The world is becoming more connected. As the world grows more connected, it also becomes more complex [13]. One result of this increased complexity is that it's become harder to make predictions of any kind. Predictions become all but impossible to make, as they take on a degree of subjectivity that was previously considered as irrelevant. We have long ago lost the ability to truly tell what's going to happen in the future, which is a pity. The world is changing, and it's time to face that fact. As society becomes more complex, it also becomes more vulnerable. This is the primary reason why the world is becoming more and more complex: because it's become increasingly necessary to ensure safety. This fact is true for anyone living in any part of the world, and this is a fact that most people never really think about. hybrid of SVM and K-means to detect disaster risk by the authors in [14] these methods are not suitable for application in practice because of the following reasons. First, traditional algorithms do not address the issue of the quality of the data or noise in the process of training. In a real disaster, the time and space resources for collecting the data are extremely limited, and data are susceptible to error. Traditional methods tend to ignore the problem of noise, and the data-to-label conversion also increases the time complexity of the algorithms. Second, the parameters of some algorithms cannot be learned using the actual situation of a disaster. This may cause the algorithm to fail to detect the risk in a real disaster. Third, some methods cannot effectively detect the complex disaster risk level. For example, the method using the SVM neural network is designed for a single-class problem; it is not suitable for detecting the risk of multi-source disasters. Finally, some of the algorithm cannot analyze multiple dimensions simultaneously and cannot achieve a global–local perspective of risk level.

3 Proposed System of Disaster Prediction

Figure 1 shows the proposed architecture of disaster prediction based on the messages. For the systems that are based on the AI and ML need the training that is called as training data. Training data are the data sets given by the user. In the next stage the data will be loaded and preprocessed for feature extraction.After extraction the SVC will divide and classifies the data to the best fit. The classification and best fit is detected then it shows the type of predicted disaster.

The following are steps to be followed that are shown in the flow diagram Fig. 2 as discussed, the system architecture in Fig. 1 the machine will be trained first with the different tweets and messages in the next the systems take the message files and extracts the message feature in the next step in collects the test features now the algorithms with SVC classifier in the trained machine. Now the system compares the with the existing data available in the database and finally give the predicted output of the users and response teams to act accordingly.

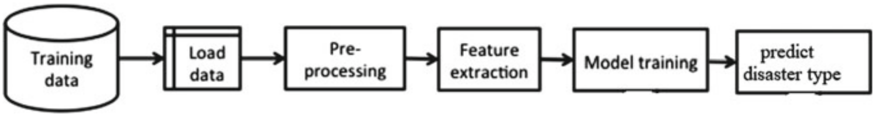
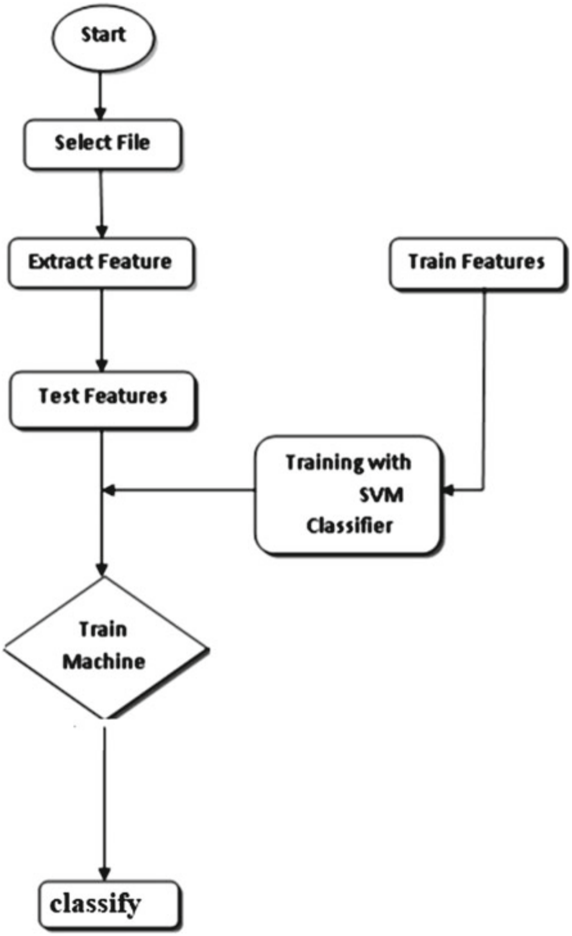


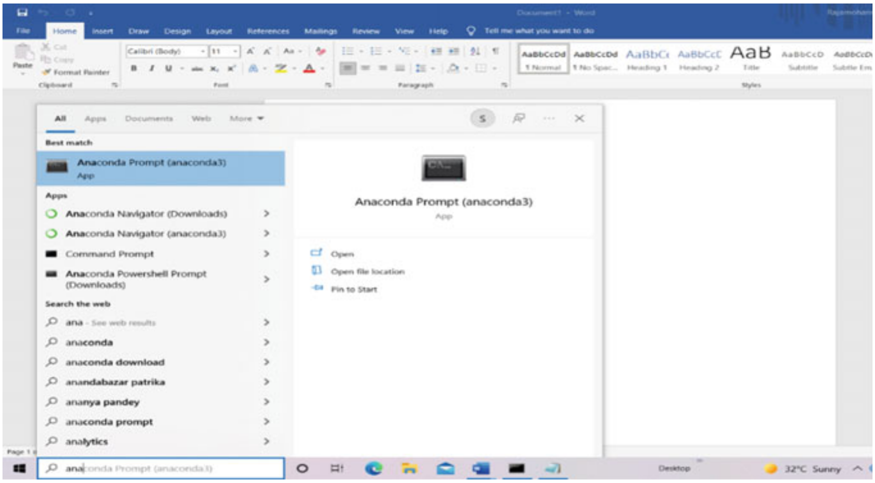
Fig. 1 Systems for prediction using machine learning

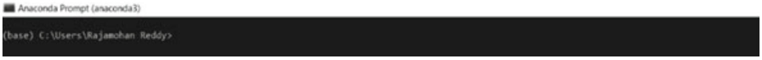
Fig. 2 Proposed system architecture in classification and prediction of disaster management



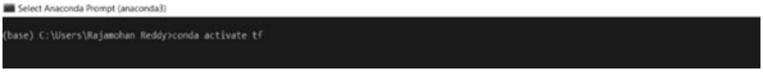
4 Method of the Proposed System

The proposed method is the implemented in the anaconda prompt with python programming by using the advantages of ease of doing stuff, easy reading and easy understanding. The configuration, programming and the loading of the external data are as shown below.

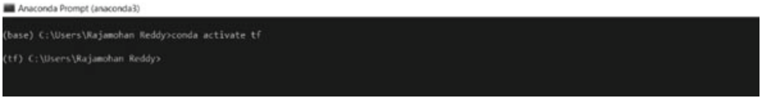



Anaconda Prompt (anaconda3)

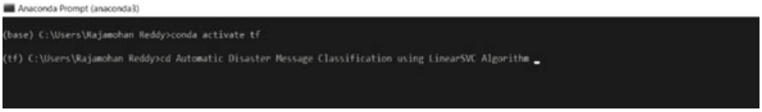
```
(base) C:\Users\Rajamohan Reddy>
```


Select Anaconda Prompt (anaconda3)

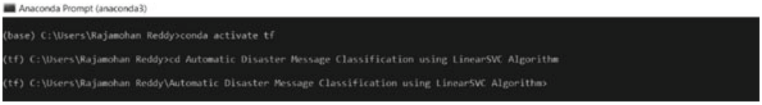
```
(base) C:\Users\Rajamohan Reddy>conda activate tf
```


Anaconda Prompt (anaconda3)

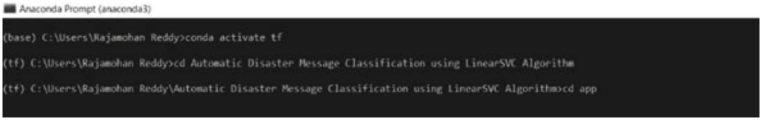
```
(base) C:\Users\Rajamohan Reddy>conda activate tf

(tf) C:\Users\Rajamohan Reddy>
```


Anaconda Prompt (anaconda3)

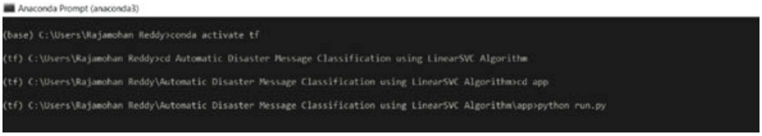
```
(base) C:\Users\Rajamohan Reddy>conda activate tf

(tf) C:\Users\Rajamohan Reddy>cd Automatic Disaster Message Classification using LinearSVC Algorithm
```


Anaconda Prompt (anaconda3)

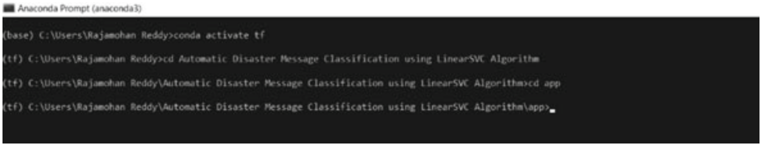
```
(base) C:\Users\Rajamohan Reddy>conda activate tf

(tf) C:\Users\Rajamohan Reddy>cd Automatic Disaster Message Classification using LinearSVC Algorithm

(tf) C:\Users\Rajamohan Reddy\Automatic Disaster Message Classification using LinearSVC Algorithm>
```



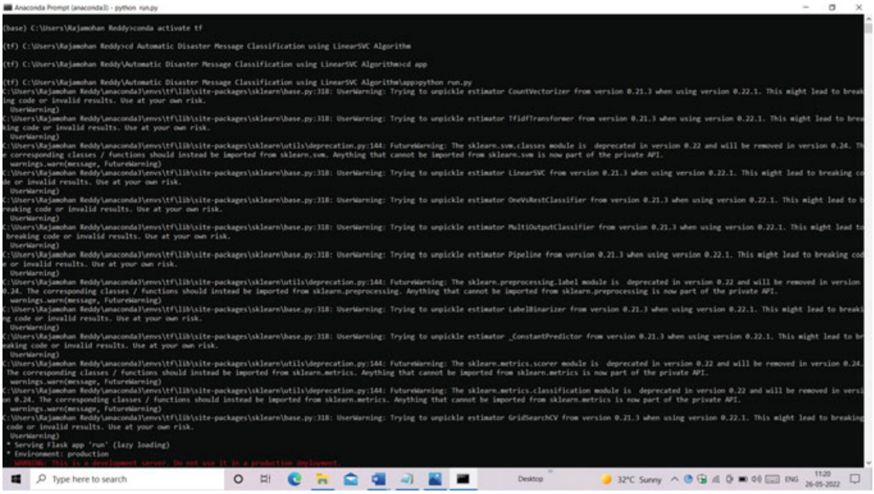
```
(base) C:\Users\Rajamohan Reddy>conda activate tf

(tf) C:\Users\Rajamohan Reddy>cd Automatic Disaster Message Classification using LinearSVC Algorithm

(tf) C:\Users\Rajamohan Reddy\Automatic Disaster Message Classification using LinearSVC Algorithm>cd app
```



```
(base) C:\Users\Rajamohan Reddy>conda activate tf

(tf) C:\Users\Rajamohan Reddy>cd Automatic Disaster Message Classification using LinearSVC Algorithm

(tf) C:\Users\Rajamohan Reddy\Automatic Disaster Message Classification using LinearSVC Algorithm>cd app

(tf) C:\Users\Rajamohan Reddy\Automatic Disaster Message Classification using LinearSVC Algorithm\app>python run.py
```



```
(base) C:\Users\Rajamohan Reddy>conda activate tf

(tf) C:\Users\Rajamohan Reddy>cd Automatic Disaster Message Classification using LinearSVC Algorithm

(tf) C:\Users\Rajamohan Reddy\Automatic Disaster Message Classification using LinearSVC Algorithm>cd app

(tf) C:\Users\Rajamohan Reddy\Automatic Disaster Message Classification using LinearSVC Algorithm\app>_
```



```
(base) C:\Users\Rajamohan Reddy>conda activate tf

(tf) C:\Users\Rajamohan Reddy>cd Automatic Disaster Message Classification using LinearSVC Algorithm

(tf) C:\Users\Rajamohan Reddy\Automatic Disaster Message Classification using LinearSVC Algorithm>cd app

(tf) C:\Users\Rajamohan Reddy\Automatic Disaster Message Classification using LinearSVC Algorithm\app>python run.py
C:\Users\Rajamohan Reddy\anaconda3\envs\tf\lib\site-packages\sklearn\base.py:318: UserWarning: Trying to unpickle estimator CountVectorizer from version 0.21.3 when using version 0.22.1. This might lead to break
ing code or invalid results. Use at your own risk.
  UserWarning)
C:\Users\Rajamohan Reddy\anaconda3\envs\tf\lib\site-packages\sklearn\base.py:318: UserWarning: Trying to unpickle estimator TfidfTransformer from version 0.21.3 when using version 0.22.1. This might lead to brea
king code or invalid results. Use at your own risk.
  UserWarning)
C:\Users\Rajamohan Reddy\anaconda3\envs\tf\lib\site-packages\sklearn\utils\deprecation.py:144: FutureWarning: The sklearn.svm.classes module is  deprecated in version 0.22 and will be removed in version 0.24. Th
e corresponding classes / functions should instead be imported from sklearn.svm. Anything that cannot be imported from sklearn.svm is now part of the private API.
  warnings.warn(message, FutureWarning)
C:\Users\Rajamohan Reddy\anaconda3\envs\tf\lib\site-packages\sklearn\base.py:318: UserWarning: Trying to unpickle estimator LinearSVC from version 0.21.3 when using version 0.22.1. This might lead to breaking co
de or invalid results. Use at your own risk.
  UserWarning)
C:\Users\Rajamohan Reddy\anaconda3\envs\tf\lib\site-packages\sklearn\base.py:318: UserWarning: Trying to unpickle estimator OneVsRestClassifier from version 0.21.3 when using version 0.22.1. This might lead to b
reaking code or invalid results. Use at your own risk.
  UserWarning)
C:\Users\Rajamohan Reddy\anaconda3\envs\tf\lib\site-packages\sklearn\base.py:318: UserWarning: Trying to unpickle estimator MultiOutputClassifier from version 0.21.3 when using version 0.22.1. This might lead to
 breaking code or invalid results. Use at your own risk.
  UserWarning)
C:\Users\Rajamohan Reddy\anaconda3\envs\tf\lib\site-packages\sklearn\base.py:318: UserWarning: Trying to unpickle estimator Pipeline from version 0.21.3 when using version 0.22.1. This might lead to breaking cod
e or invalid results. Use at your own risk.
  UserWarning)
C:\Users\Rajamohan Reddy\anaconda3\envs\tf\lib\site-packages\sklearn\utils\deprecation.py:144: FutureWarning: The sklearn.preprocessing.label module is  deprecated in version 0.22 and will be removed in version
0.24. The corresponding classes / functions should instead be imported from sklearn.preprocessing. Anything that cannot be imported from sklearn.preprocessing is now part of the private API.
  warnings.warn(message, FutureWarning)
C:\Users\Rajamohan Reddy\anaconda3\envs\tf\lib\site-packages\sklearn\base.py:318: UserWarning: Trying to unpickle estimator LabelBinarizer from version 0.21.3 when using version 0.22.1. This might lead to breaki
ng code or invalid results. Use at your own risk.
  UserWarning)
C:\Users\Rajamohan Reddy\anaconda3\envs\tf\lib\site-packages\sklearn\base.py:318: UserWarning: Trying to unpickle estimator _ConstantPredictor from version 0.21.3 when using version 0.22.1. This might lead to br
eaking code or invalid results. Use at your own risk.
  UserWarning)
C:\Users\Rajamohan Reddy\anaconda3\envs\tf\lib\site-packages\sklearn\utils\deprecation.py:144: FutureWarning: The sklearn.metrics.scorer module is  deprecated in version 0.22 and will be removed in version 0.24.
 The corresponding classes / functions should instead be imported from sklearn.metrics. Anything that cannot be imported from sklearn.metrics is now part of the private API.
  warnings.warn(message, FutureWarning)
C:\Users\Rajamohan Reddy\anaconda3\envs\tf\lib\site-packages\sklearn\utils\deprecation.py:144: FutureWarning: The sklearn.metrics.classification module is  deprecated in version 0.22 and will be removed in versi
on 0.24. The corresponding classes / functions should instead be imported from sklearn.metrics. Anything that cannot be imported from sklearn.metrics is now part of the private API.
  warnings.warn(message, FutureWarning)
C:\Users\Rajamohan Reddy\anaconda3\envs\tf\lib\site-packages\sklearn\base.py:318: UserWarning: Trying to unpickle estimator GridSearchCV from version 0.21.3 when using version 0.22.1. This might lead to breaking
 code or invalid results. Use at your own risk.
  UserWarning)
 * Serving Flask app 'run' (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
```

Using the python the User interface (UI) is created for the giving the input data as shown below. By entering the IP address in the google crome the project window gets opened as shown in Fig. 3.

In the next step user can be able to enter the message in the comment box provided and click on the classify the message. Once the classify is clicked Support Vector Machines will analyze the data that are powerful supervised machine learning models which have been well researched and implemented in the various tools available. Support Vector Classification which uses Support Vector Machines has gained a lot

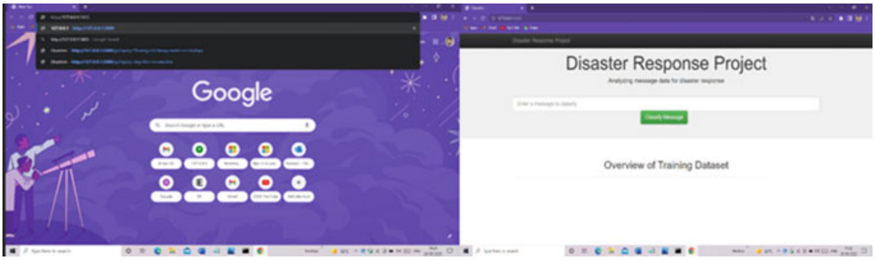
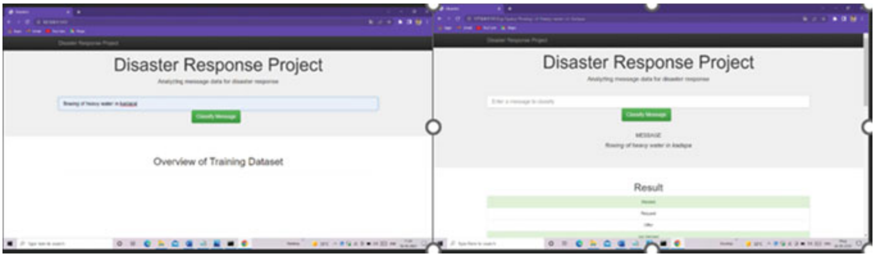

Google
Disaster Response Project
Analyzing message data for disaster response
Classify Message
Overview of Training Dataset

Fig. 3 Disaster response project


Disaster Response Project
Analyzing message data for disaster response
Classify Message
Overview of Training Dataset
Disaster Response Project
Analyzing message data for disaster response
Enter a message to classify
Classify Message
MESSAGE
Result

flowing of heavy water in kadapa

Fig. 4 User input and classification of result

of popularity due to the better performance on sparse data, high scalability, and low training complexity. A Support Vector Classification model is used to classify a new unlabeled instance given by the user. A key component of a Support Vector Classification model is the support vectors, which are used to form the decision boundaries that separates the training data into two classes. A Support Vector Classification model uses the Support Vectors to make classification decisions. A Support Vector Classification model is said to learn a non-parametric boundary as compared to a parametric boundary. User input and classification of result as shown in Fig. 4.

5 Results

After they are taken by giving different input to the system and the results are acquired as shown in Figs. 5, 6 and 7. The proposed system is evaluated using the diverse test case in that some of them are shown below. The first case is heavy rain fall in Kadapa the classifier as shown the results are aid related and water floods related, they are given the response team. In the same way it was evaluated in other conditions like big fire in America and heavy winds in vizag and the results can be observed in Figs. 5, 6 and 7.

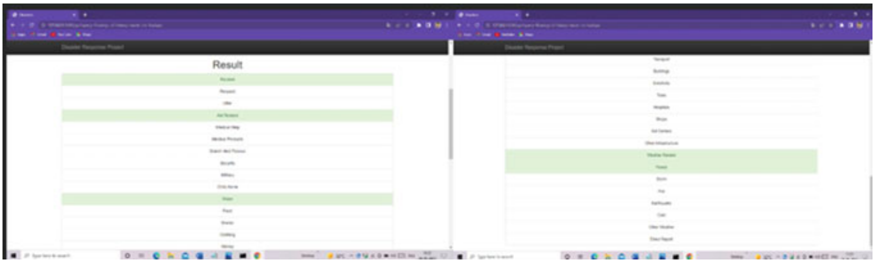


Fig. 5 Results of the 1st input

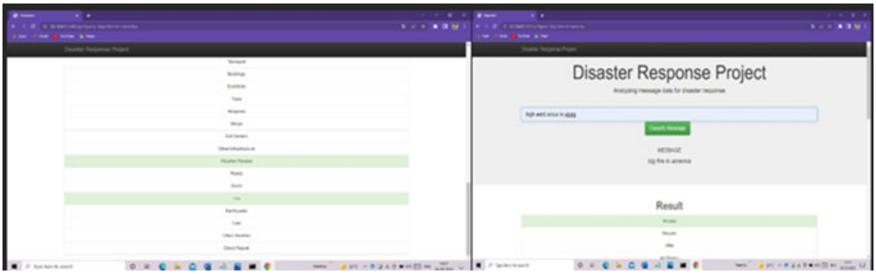

Disaster Response Project
Disaster Response Project
Analyzing message data for disaster response
Classify Message
MESSAGE
big fire in america
Result

Fig. 6 Results of the second input big fire in America

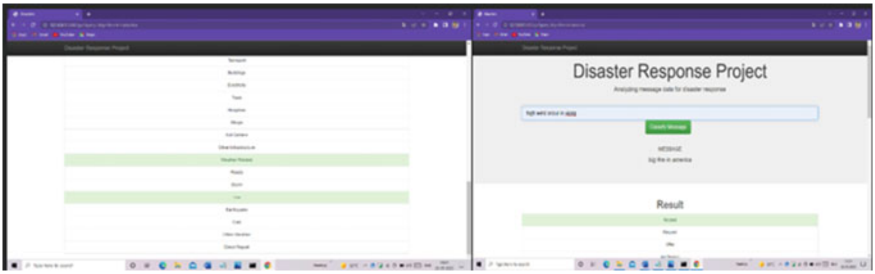

Disaster Response Project
Disaster Response Project
Analyzing message data for disaster response
Classify Message
MESSAGE
big fire in america
Result

Fig. 7 Results for high wind in Vizag

6 Conclusion

Disaster may happen at any time, and any were for this the response teams have to predict the exactly and necessary steps are to be taken. The proposed systems in implemented in SVC and the results are obtained in the accurate, precision.

References

1. Krishna PG et al (2019) An approach of tomato leaf disease detection based on SVM classifier. Int J Recent Technol Eng 7(6):697–704
2. Caragea C, Silvescu A, Tapia AH (2016) Identifying informative messages in disaster events using convolutional neural networks. In: Proceedings of 13th international ISCRAM conference. ACM
3. Blum A, Mitchell T (1998) Combining labeled and unlabeled data with co-training. In: Proceedings of the eleventh annual conference on computational learning theory, pp 92–100
4. Beigi G, Hu X, Maciejewski R, Liu H (2016) An overview of sentiment analysis in social media and its applications in disaster relief. In: Sentiment analysis and ontology engineering. Springer, pp 313-340. BBC news http://www.bbc.com/news/blogs-trending-34836214
5. Nagadasari MP, Bojja P (2021) Industrial IoT enabled fuzzy logic based flame image processing for rotary kiln control. Int J Pervasive Comput Commun. https://doi.org/10.1108/IJPCC-10-2020-0161
6. Khwanpheng S, Lwin KT, Chaisricharoen R, Temdee P (2015) Collaborative crosschecking system of observed loss estimation for disaster relief management. In: 2015 9th International conference on software, knowledge, ınformation management and applications (SKIMA), pp 1–5. https://doi.org/10.1109/SKIMA.2015.7400049
7. Zarei V, Behjat V, Baziar R, Azimizadeh MR (2015) Pre-disaster base maintains of distribution transformers for improving crisis management. In: 2015 20th Conference on electrical power distribution networks conference (EPDC), pp 179–183. https://doi.org/10.1109/EPDC.2015.7330492
8. Nunavath V, Goodwin M (2018) The role of artificial ıntelligence in social media big data analytics for disaster management—initial results of a systematic literature review. In: 2018 5th International conference on ınformation and communication technologies for disaster management (ICT-DM), pp 1–4. https://doi.org/10.1109/ICT-DM.2018.8636388
9. Wikipedia Contributors (2020) Machine learning—wikipedia the free encyclopedia. Retrieved from https://en.wikipedia.org/w/index.php?title=Machine_learning&oldid=954000510. Accessed 30 Apr 2020
10. Chen N, Qiu T, Zhou X, Li K, Atiquzzaman M (2019) An intelligent robust networking mechanism for the Internet of Things. IEEE Commun Mag 57(11):91–95
11. Nagadasari MP, Bojja P (2022) Industrial IoT enabled fuzzy logic based flame image processing for rotary kiln control. Wirel Pers Commun.https://doi.org/10.1007/s11277-022-09677-z
12. Krishna PG et al (2018) Smart farming based on embedded technology for detection of leaf disease and control. J Adv Res Dyn Control Syst 10(02):534–540
13. Narayana BV, Ravi KS, Krishna PG (2019) An advanced crop field monitoring system ın agriculture through java beans and gsm modem 1771
14. Ashktorab Z, Brown C, Nandi M, Culotta A (2014) Tweedr: mining twitter to inform disaster response. In: Proceedings of 11th international conference on ınformation systems for crisis response and management (ISCRAM 2014). University Park, PA, pp 354–358. BBC Trending (2015)

Intelligent Healthcare System

M. Senthamil Selvi, K. Abinaya, N. Jemy Sharon, and R. Lakshmi Pooja

Abstract Heart and kidney are the two major organs in a human cardiovascular disease (CVDs) and Chronic Kidney Disorders (CKDs) are the leading by causing death and health related issues globally. Mostly, cardiovascular diseases or any heart abnormalities can be prevented by addressing some of the risk factors such as using tobacco, unhealthy diet which leads to obesity, physical inactivity and massive consumption of alcohol. CKD means the kidneys are damaged and losing their ability to keep our body healthy by filtering the blood. CKDs can only be treated, by early clinical-diagnosis and treatment. So that it's possible to slow down or stop the progression of kidney disease. The heart helps to pump the blood filled with oxygen through all parts of the body, including the kidneys. As we know, the kidneys help cleaning the blood, by removing the waste products and extra water in the body. Without the help of kidneys, the blood in our body would contain too much waste and extra water that is unnecessary, which can lead to be fatal infections sometimes. So, only by the proper functioning of both heart and kidney would help maintain our body functionalities properly. It is very important to understand that everybody with renal disease is at risk for heart problems, which can increase your chances of developing heart disease. The Intelligent health care system works as a web application in which users can enter some blood test parameters and blood pressure (BP) in order to find if there is any abnormality or not in their heart and kidney. This application creates awareness about the heart and kidney health. It is better to follow prevention than getting cured is the strategy followed here.

M. Senthamil Selvi (✉) · K. Abinaya · N. Jemy Sharon · R. Lakshmi Pooja
Sri Ramakrishna Engineering College, Coimbatore, India
e-mail: hod-it@srec.ac.in

K. Abinaya
e-mail: Abinaya.1805002@srec.ac.in

N. Jemy Sharon
e-mail: jemysharon.1805038@srec.ac.in

R. Lakshmi Pooja
e-mail: lakshmipooja.180505050@srec.ac.in

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_64

Keywords CVD · CKD · Blood pressure · Blood test parameters

1 Introduction

The project's goal is to combine ML with web technology so that ordinary people can use it in their daily lives. The aim of the project is to help the patients in emergency situations. It is used to detect heart and kidney abnormalities using ML algorithms integrated with the web app. Supervised ML model is being used in order to train the model The dataset is trained to predict whether the individual has heart or lung abnormalities and display it by using certain parameters from his/her blood test as input. The project's goal is to combine ML with web technology so that common people can use it on a daily basis without having any knowledge of AI or ML just with the help of a User Interface (UI).

Heart disease has surpassed cancer as the top cause of mortality in the world over the last few decades, and it is now the main cause of death not only in India but worldwide. As a result, there is a current need for a system that can reliably identify and treat such illnesses. The heart, as we all know, is a vital organ in our bodies. If a human's heart does not function properly, it will impact other body organs such as the brain, kidneys, and so on. It works more like a pump, allowing blood to move in and out of the body [1]. When blood circulation in the body is poor, organs like the brain suffer, and if the heart stops working totally, death occurs within minutes. The effective functioning of the heart is absolutely necessary for life.

The kidney is also linked to the heart, thus any problems with the heart will have an impact on the kidney. They are interconnected organs in the human body. Over 10% of the world's population has CKD, and millions of people die each year owing to a lack of affordable treatment choices. It is possible that becoming aware of their heart and renal issues would prevent them from becoming impacted and resulting in any unfortunate circumstances [2].

Websites for health care are one of the most crucial resources a person can have in an emergency nowadays. The web app will feature sections for users to enter data based on blood test results and BP, which will be fed into a ML system that will forecast whether there is a cardiac or kidney issue. As a result, the user may get the idea to check on their body and prevent any unfortunate events, potentially saving their life. People will be more mindful of their health as a result of this.

2 Literature Review

A. Louridi et al. proposed a system "Machine learning-based identification of patients with a cardiovascular defect" [3]. This research paper proposes an effective intelligent medical system based on machine learning techniques to assist doctors in accurately diagnosing whether or not a patient has CVD.

B. Rindhe et al. proposed a system "Heart Disease Prediction Using Machine Learning" [4]. This work includes proper data processing and analysis of the heart disease patient dataset. Then, with the highest possible scores, three models were trained and tested, Support Vector Classifier: 84.0%, NN: 83.5%. A SVM can make some errors to avoid over-fitting. It tries to minimize the no of errors that will be made to SVM classifiers that are applied in many apps. The purpose of a neural network is to learn network parameters so that the anticipated outcome matches the ground truth. Random Forest is a machine learning algorithm that is supervised. This method can be used for both regression and classification tasks, but it excels at classification.

C. Thary Al-Ghrairi et al. proposed a system "An Application of Web-based E-Healthcare Management System Using ASP.Net" [5]. This paper provides a website for the medical healthcare system. It has two key components: client and server. The client-side refers to everything that the user sees; it was built as a website using HTML, CSS and JavaScript. Server-side refers to how the site, such as servers and databases, makes changes and updates in the browser.

D. Xavier et al. proposed a system "Heart Disease Prediction using Machine learning and Data Mining Technique" [6]. To develop a good classifier, this research proposed a classification algorithm that predicts a restricted set of relationships between qualities in databases. Based on assumed parameters and the best associative classification algorithm, a skilled system is designed for end-users to check the risk of heart illnesses.

E. Imteaj and Hossain proposed a system "A smart phone-based application to improve the health care system of Bangladesh" [7]. This paper mainly focuses to provide an effective health care system in the city of Bangladesh with the help of a mobile app. By availing this app people will be able to get various benefits like finding the information about the city's hospital, cabin availability, paying for a cabin in the hospital, intelligent tips on choosing a suitable hospital, locating a doctor, emergency service calling, first aid information, medication alert system, BMI calculator, and so on.

F. Marin et al. proposed a system "Web Application for self-diagnosis and drug recommendation based on user's symptoms" [8]. This research paper is detailed about an app that is designed to give an online self-diagnosis and drug recommendation tool based on natural language processing of the user's symptoms. This platform aids in the automation of the search process and the provision of the most relevant information to the user by removing the need for manual data interpretation.

G. Kuo et al. proposed a system "Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning" [9]. To predict kidney function, this work uses a Resnet model pre-trained on an ImageNet dataset in their NN architecture. Based on 4,505 renal ultrasound pictures, predict kidney function. It's a crucial step toward fulfilling the potential of renal ultrasound imaging as a real-time, remote screening tool.

H. Kolandaisamy and Noor proposed a system "Web-based Online Medical Diagnosis System (WOMEDS)" [10]. This paper proposed an app has a feature for users to do diagnostics for the health problem and will provide some health monitoring and tips for the user to follow. This system will concentrate on Registration and Administration, Diagnosing and Treatment, Health Monitor and Tips. Each patient will have their own username and a password for themselves, where they can login and look at their information and diagnosis. The database is built using SQL Server languages, making it simple for patients to access their health records.

I. Sobrinho et al. proposed a system "Design and evaluation of a mobile application to assist the self-monitoring of the chronic kidney disease in developing countries" [11]. This research proposes an app to aid in the early diagnosis and self-monitoring of CKD while taking into account quality features such as safety, effectiveness, and usability. When CKD risk evaluations are available, the app also allows individuals to communicate their results with nephrologists.

3 Methodology

The proposed flowchart is shown in Fig. 1.The goal of the system is to combine web technology with machine learning to build seamless software to identify any abnormalities in the heart and kidney, and a web page with a UI for the user to enter parameters from the blood test report and the BP.

A variety of illnesses that affect our heart are referred to as heart disease. Various harmful habits, such as excessive cholesterol, obesity, increased triglyceride levels, hypertension, and so on, raise the risk of heart disease. All of these symptoms resemble other diseases, such as those that affect the elderly, making it difficult to make a precise diagnosis, which could lead to death in the near future.

A person with kidney disease is more likely to develop heart disease. The most common cause of death which prevails among patients with kidney disease is because of the heart disease. Kidney disease majorly occurs among the patients when the kidneys get damaged and when they are unable to filter the blood as effectively as they should. Wastes might build up in the body as a result of kidney damage.

When given an input of age, gender, BP, blood sugar level, and blood urea level, this web app executes a supervised ML Algorithm on the backend. The output indicates whether or not the person has heart or kidney abnormality. Thus, future fatal situations can be avoided. As a result, the user may develop the idea of checking in on their body and avoiding any potentially catastrophic situations, as well as saving their life. People will be more concerned about their health now that they are aware that they may have a problem. The system includes the following steps to identify the heart and kidney has abnormality or not.

a. Dataset collection
b. Dataset pre-processing
c. Model Development

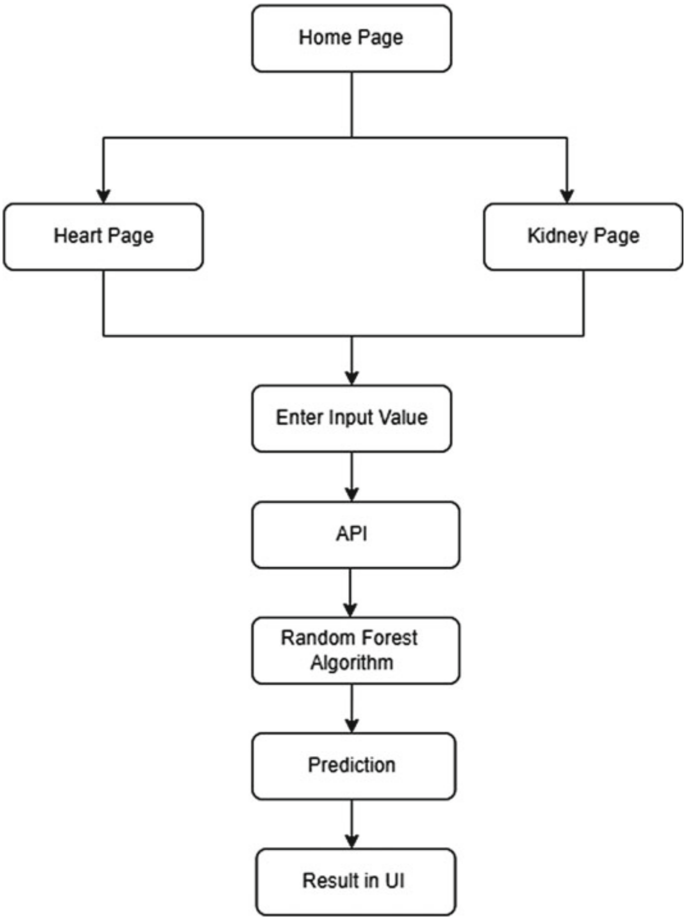


Fig. 1 Flowchart

- Machine Learning
- Front End Support
- Backend Support

(1) Dataset collection

There are 450 items in the dataset. In which 225 data sets are for heart and 225 datasets are for kidney. Heart and kidney each have several parameters taken into consideration for identifying abnormalities.

The parameters in the kidney dataset are age, blood pressure, specific gravity, serum creatinine, albumin, sugar, sodium, potassium, blood glucose, blood urea, diabetes mellitus, red blood cell, platelet count, pus cell clumps, packed cell volume. The heart and kidney datasets are shown in the Tables 1 and 2.

Table 1 Heart dataset

Age	Sex	Cp	Trestbps	Chol	Fbs	Restecg	Thalach	Exang	Oldpeak	Slope	Ca	Thal	Target
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Table 2 Kidney dataset

Age	Bp	58	Al	Su	Rbc	Pc	Pcc	Bgr	Bu	Sc	Sod	Pot	Pcv	Dm	Classification
62	70	1.025	3	0	Normal	Abnormal	Notpresent	122	42	1.7	136	4.7	39	yes	CKD
54	70						Notpresent	233	50.1	1.9				yes	CKD
47	80						Notpresent	114	87	5.2	139	3.7		no	CKD
43	60	1.025	0	0	Normal	Normal	Notpresent	108	25	1	144	5	43	no	NotCKD
42	100	1.015	4	0	Normal	Abnormal	notpresent		50	1.4	129	4	39	no	CKD
60	50	1.01	0	0		Normal	Notpresent	261	58	2.2	113	3		no	CKD
29	80	1.02	0	0	Normal	Normal	Notpresent	83	49	0.9	139	3.3	40	no	NotCKD

The parameters in the heart dataset are age, gender, chest pain, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiography, maximum heart rate, exercise induced angina, ST depression induced by exercise, slope of the peak exercise, no of major vessels colored by fluoroscopy, thalassemia.

1. No. of datasets for Heart = 225
2. No. of datasets for Kidney = 500
3. No. of parameters for Heart = 14
4. No. of parameters for Kidney = 16
5. Percentage of Training Data = 70%
6. Percentage of Testing Data = 30%.

(2) Dataset Pre-processing

The columns that are not necessary are removed. For kidney only, Age, Blood Glucose, BP, Sugar Level, Serum Creatine, Sodium, Potassium, Blood Urea, and Diabetes are only needed so other parameters are removed. For Heart only, Age, Gender, Chest Pain, Resting BP, Cholesterol, and Fasting Blood Sugar are used. The noisy data has been removed from both heart and kidney datasets. Some values were missing in both datasets; these values were handled during data preprocessing. Some values have been filled with the average value in order to avoid overfitting issues. Labelling of data with an abnormality as 1 and no abnormality as 0 was also done in the preprocessing. The first and most important stage in ML is to acquire datasets. Next, the dataset for this project is separated into training and testing data using a split function. This project's training data is fed into the model's regression to predict cardiac and renal abnormalities. An API connects the ML algorithm to the front end. In the front end, the user can enter the blood test parameters in the input area. The input values are provided to the API, which does the classification in the backend which contains the ML model. The results of the model are displayed in the front end using the API, it shows whether the person has abnormalities in heart or kidney. Heart and kidney dataset after preprocessing are shown in the Tables 3 and 4.

(3) Model Development

The front end consists of a home page which has 2 buttons called Heart and Kidney. If the heart button is clicked it redirects to the heart abnormality page. It has several parameters listed which need to be entered by the user. It also has a home button

Table 3 Heart dataset after pre-processing

	Age	Sex	Cp	Trestbps	Chol	Fbs	Target
0	63	1	3	145	233	1	1
1	37	1	2	130	250	0	1
2	41	0	1	130	204	0	1
3	56	1	1	120	236	0	1
4	57	0	0	120	354	0	1

Table 4 Kidney dataset after pre-processing

	Age	Bp	Su	Sc	Sod	Pot	Bgr	Bu	Dm	Classification
0	62.0	70.0	0.0	1.7	136.0	4.7	122.0	42.0	1.0	1
1	54.0	70.0	0.0	1.9	138.0	4.4	233.0	50.1	1.0	1
2	47.0	80.0	0.0	5.2	139.0	3.7	114.0	87.0	0.0	1
3	43.0	60.0	0.0	1.0	144.0	5.0	108.0	25.0	0.0	0
4	29.0	80.0	0.0	0.9	139.0	3.3	83.0	49.0	0.0	0

that redirects the user back to home page. When the user clicks the kidney button it leads to the kidney abnormality page and it also has a home button. The parameters listed are blood test parameters along with BP. When the user enters the parameters in the input field and clicks the abnormality or not button, the data entered by the user are sent to the ML algorithm via an API called the flask. This web framework helps in connecting the front end to the back end. The back end is a python file that contains the trained model for the random forest algorithm that takes the input and produces a result whether there is heart or kidney abnormality and this data is sent to the front-end using flask and is displayed on the web page.

(3.1) Machine Learning

Random forest falls under the category of Supervised ML Algorithm and is can be seen widely used in classification and regression problems. It helps in building various decision trees on different samples of data which is given and it takes an average in case of regression and a majority vote for classification.

One of the Random Forest Algorithm's most pivotal feature is that it is capable of handling the data set which consists of continuous variables if you're working on the case of regression problems and categorical variables if you're working on the case of classification problems. We can see that Random Forest Algorithm performs better for classification problems than regression problems. Random forest algorithm is shown in Fig. 2.

Steps involved in random forest algorithm

Step 1: Random Forest consists of n number of random records which are chosen from the data set having k number of records.

Step 2: For each sample, a unique decision tree is created. Step 3: Each and every decision tree will produce its own set of results.

Step 4: For classification and regression problems, the final output is mainly based on majority of voting or averaging accordingly.

(3.2) Front end

The front end has a home page with 2 buttons one for heart and the other for kidney as in Fig. 3.

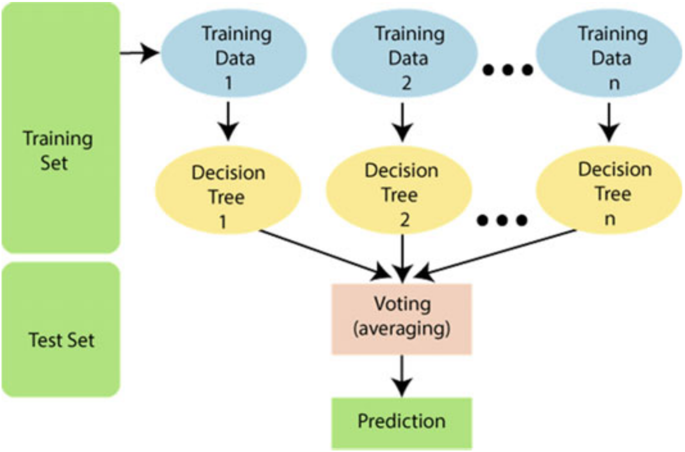
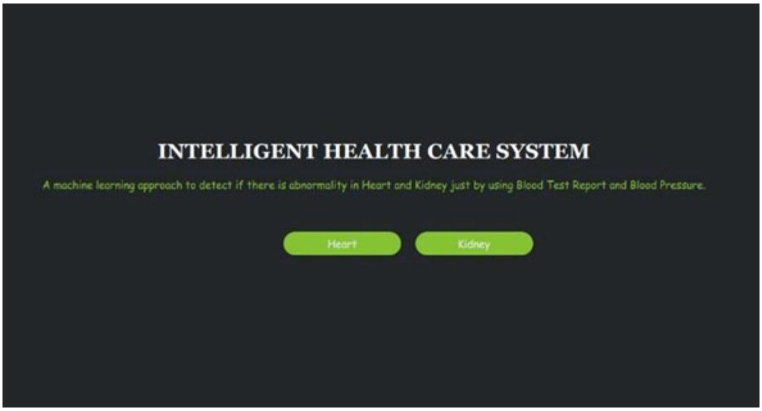


Fig. 2 Random Forest algorithm



INTELLIGENT HEALTH CARE SYSTEM

A machine learning approach to detect if there is abnormality in Heart and Kidney just by using Blood Test Report and Blood Pressure.

Fig. 3 Home page

When the Heart button is clicked, it gets redirected to the Heart Abnormality detection page as in Fig. 4. It shows input fields related to the blood test report that the user needs to fill. A button to find whether there is abnormality or not is there. When that button is clicked, the input parameters are sent to computation.

When the Kidney button is clicked, it gets redirected to the Kidney Abnormality detection page as in Fig. 5. It shows input fields related to the blood test report that the user needs to fill. A button to find whether there is abnormality or not is there. When that button is clicked the input parameters are sent to computation.

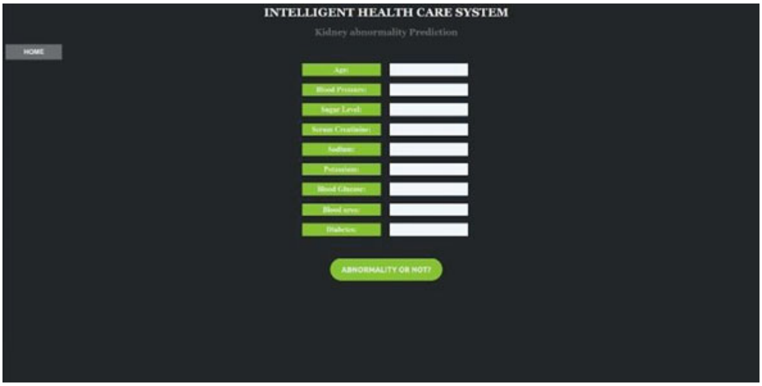


Fig. 4 Heart abnormality page

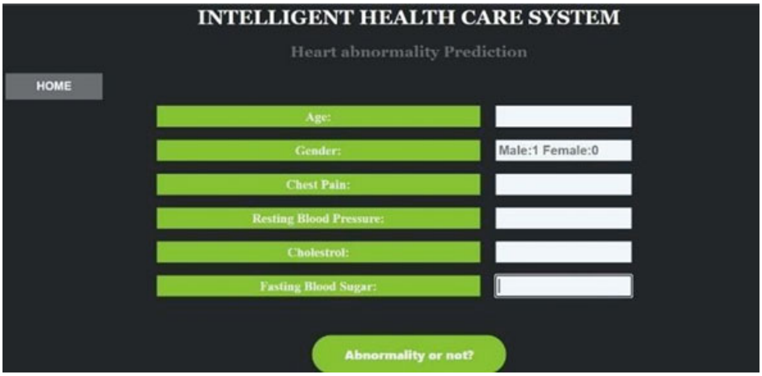


Fig. 5 Kidney abnormality page

(3.3) Back end

The flask is being used and here the reference to it https://palletsprojects.com/p/flask/ backend process consists of the ML model which is trained to predict heart and kidney abnormalities through blood tests. Few major parameters which are the first indicators of the abnormalities are first jotted down and checked with several data and have been used to train the dataset. Both models have been trained with the Random Forest Classifier. This model has proved to be very efficient for our dataset. An API has been developed for both the models, to take the input from the user and to run the model and to deliver the result in the frontend. The API has been created with the help of flask framework, using the GET request.

(4) Merits

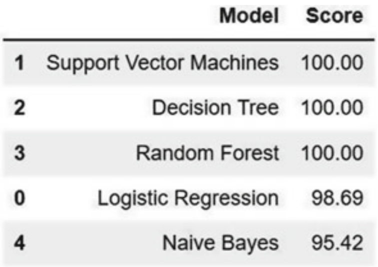
The app can be used for common people even if they do not have any idea in ML or AI. It has a simple UI that will be easy for the users to use. It can help people to acknowledge the abnormality before it gets severe. It just uses blood test parameters and BP in order to find the abnormality. The app can be used in order to know about someone's heart and kidney health and if the user has abnormality, it is better to visit the doctor to get an opinion on one's health, so that one can avoid the severity of disease that may be dangerous.

(5) Demerits

The data set used in training the ML model is health related, it is sensitive in nature. Health related data needs to be protected and patients' details shouldn't be misused. The availability of data is low and the model overfits for the heart abnormality detection alone. This app works under the blood test results. If the user does not have blood test results with them, they cannot use this app. The app might produce some result that might be wrong and as it is a health app it is a major risk that needs to be taken care of.

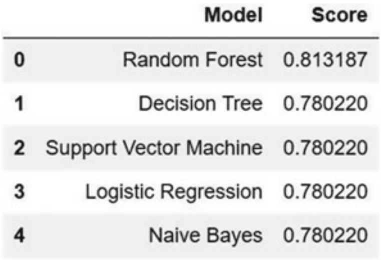
4 Experimental Result and Discussion

Different ML algorithms were used for testing the accuracy and a random forest algorithm was chosen since it has a good accuracy as in Figs. 6 and 7. It avoids overfitting since it uses multiple trees. It can handle several features easily. It is much faster than many other algorithms like SVM. Random forests can handle categorical data very well. It has less false positive rates compared to other algorithms.



	Model	Score
1	Support Vector Machines	100.00
2	Decision Tree	100.00
3	Random Forest	100.00
0	Logistic Regression	98.69
4	Naive Bayes	95.42

Fig. 6 Heart accuracy



	Model	Score
0	Random Forest	0.813187
1	Decision Tree	0.780220
2	Support Vector Machine	0.780220
3	Logistic Regression	0.780220
4	Naive Bayes	0.780220

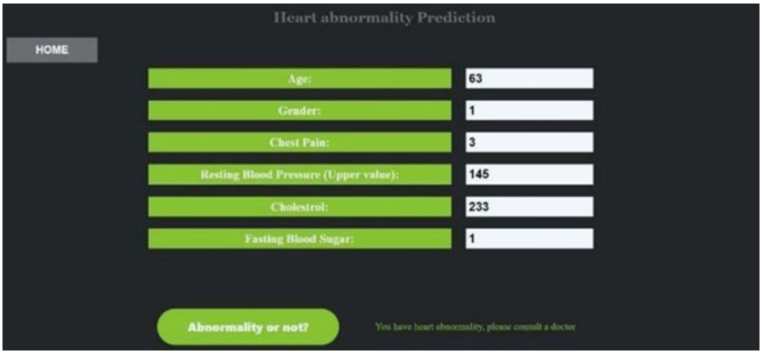
Fig. 7 Kidney accuracy

As they are only limited number of datasets for heart, as it a sensitive data, the data augmentation cannot be done. Hence, there is an overfit in the model. But if there are given an extra 200 real-time data for the heart. The accuracy of the model will be perfect. It will show more accurate results. Currently, the heart dataset consists of 225 data. Whereas on the other hand, the kidney dataset consists of 500 real-time data and thus making it more efficient in producing results than the heart model, when it is applied on the Random Forest Algorithm.

The front end was implemented with a home page having 2 buttons called heart and kidney abnormalities. When the heart abnormality button is clicked it goes to the abnormality page and the user needs to enter the following parameter for heart, Age, Gender, Chest Pain, Resting BP, Cholesterol, Fasting Blood Sugar.

When the kidney Button is clicked it goes to the kidney abnormality page and the user needs to enter the following parameter for kidney abnormality detection, Age, BP, Sugar Level, Serum Creatine, Sodium, Potassium, Blood Urea, Diabetes, Blood Glucose.

When the parameters are entered and the abnormality or not button is clicked, the input values are sent to the ML algorithm using the app Programming Interface (API). The connection is established using the web framework flask. The ML algorithm will run in the back end with the data sent and give the output as 1 for abnormality and 0 for No abnormality and that is sent to the front end using the API and that is printed in the UI as output. Many algorithms were run and the algorithm with the best accuracy was chosen i.e., the Random Forest for implementation. The validation results are shown in the Figs. 8, 9, 10 and 11.

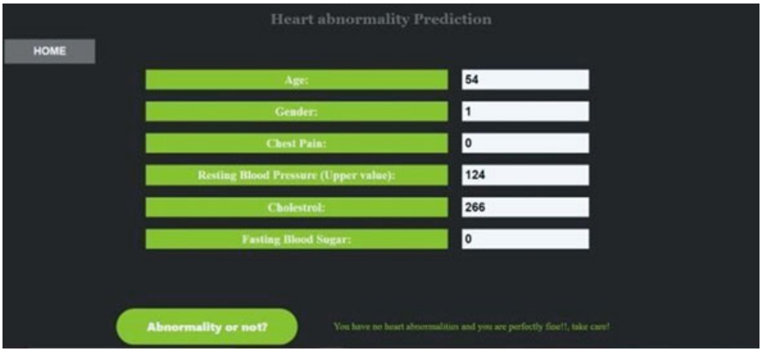


Heart abnormality Prediction

Age:	63
Gender:	1
Chest Pain:	3
Resting Blood Pressure (Upper value):	145
Cholestrol:	233
Fasting Blood Sugar:	1

You have heart abnormality, please consult a doctor

Fig. 8 Having heart abnormality



Heart abnormality Prediction

Age: 54

Gender: 1

Chest Pain: 0

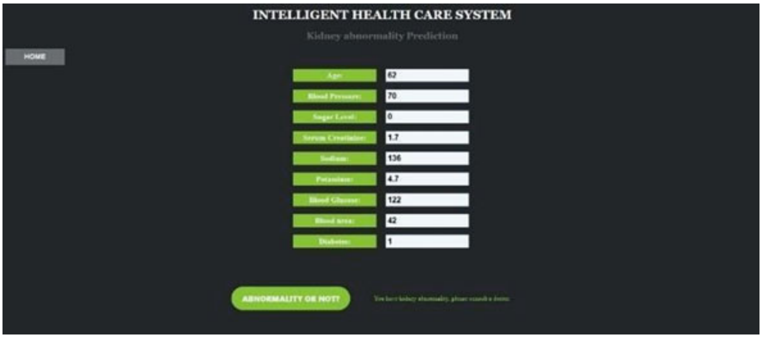
Resting Blood Pressure (Upper value): 124

Cholestrol: 266

Fasting Blood Sugar: 0

You have no heart abnormalities and you are perfectly fine!!, take care!

Fig. 9 Not having heart abnormality



INTELLIGENT HEALTH CARE SYSTEM

Kidney abnormality Prediction

Age: 62

Blood Pressure: 70

Sugar Level: 0

Serum Creatinine: 1.7

Sodium: 136

Potassium: 4.7

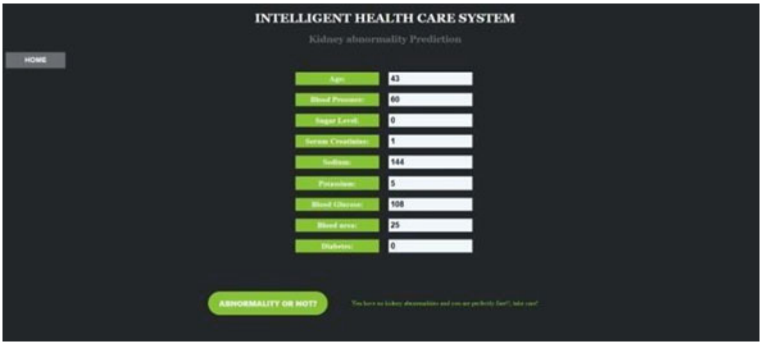
Blood Glucose: 122

Blood urea: 42

Diabetes: 1

You have kidney abnormality, please consult a doctor.

Fig. 10 Having kidney abnormality



INTELLIGENT HEALTH CARE SYSTEM

Kidney abnormality Prediction

Age: 43

Blood Pressure: 60

Sugar Level: 0

Serum Creatinine: 1

Sodium: 144

Potassium: 5

Blood Glucose: 108

Blood urea: 25

Diabetes: 0

You have no kidney abnormalities and you are perfectly fine!!, take care!

Fig. 11 Not having kidney abnormality

5 Conclusion

This project provides deep insight into ML techniques for identification of heart and kidney abnormalities. It raises public awareness of kidney and heart abnormalities so that they can be treated at time and the larger problem avoided The ML is a complex topic that can be understood only by professionals and in order for the common people to use we have linked the ML model to a web page. Thus, the user does not need to be concerned about ML and can simply use this website to determine if they have any abnormalities in their heart or kidney by entering the parameters from a blood test, and if the app suggests abnormalities, the user should consult a doctor as soon as possible.

6 Future Scope

List of Hospitals, Doctors, Appointment bookings, pharmacies, Scan centers can be listed based on the user's location and the user can search the doctor they need to visit and get appointments as well. Other diseases like diabetes can be included as part of the project using some parameters to find if there are any complications. Try to obtain more data from blood test reports and use it in our project to substantially improve the performance metrics. The link between heart and kidney disorder can be found out by developing a meta-algorithm using any ML and implementing it into our project. There are several possibilities for further additional research that would considerably improve and enhance the functionality of the current study and research.

References

1. Rathore PS, Sharma BK (2022) Improving healthare delivery system using business intelligence. J IoT Soc Mob Anal Cloud 4(1):11–23
2. Kumar D, Smys S (2020) Enhancing security mechanisms for healthcare informatics using ubiquitous cloud. J Ubiquit Comput Commun Technol 2(1):19–28
3. Louridi N, Douzi S, El Ouahidi B (2021) Machine learning-based identification of patients with a cardiovascular defect. J Big Data 8(1). Article no: 133
4. Rindhe BU, Ahire N, Patil R, Gagare S, Darade M (2021) Heart disease prediction using machine learning, pp-62–71
5. Al-Ghrairi AHT, Mohammed AA, Saeed HM (2021) An application of web-based E-healthcare management system using ASP.Net. Webology 18(1):285–298
6. Xavier A, Sadat S, Chakalakal S (2021) Heart disease prediction using machine learning and data mining technique, vol 7. Article no: 67
7. Imteaj A, Hossain MK (2020) A smart phone-based application to improve the health care system of Bangladesh, vol 10
8. Marin I, Goga N, Stanciu RC (2019) Web application for self-diagnosis and drug recommendation based on user's symptoms. J Adv Technol Eng Res 5(2):62–71

9. Kuo CC, Chang CM, Liu KT, Lin WK, Chiang HY, Chung CW, Ho MR, Sun PR, Yang RL, Chen KT (2019) Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning. NPJ Digit Med 2(1):1–9
10. Kolandaisamy R, Noor RM (2019) Web-based online medical diagnosis system (WOMEDS), Vol 1
11. Sobrinho Á, da Silva LD, Perkusich A, Pinheiro ME, Cunha P (2018) Design and evaluation of a mobile application to assist the self- monitoring of chronic kidney disease in developing countries. BMC Med Inf Decis Making 18(1):1–4

Intelligent Predictive Maintenance for Industrial Internet of Things (IIoT) Using Machine Learning Approach

Umesh W. Hore and D. G. Wakde

Abstract The Industrial Internet of Things is a intricate area which comprises feature like information and operation technology, statistics, and engineering. The industrial data management system uses five basic layers like things layer, edge layer, fog Layer, communication layer, and cloud services to build a system for industrial operation. The cloud assisting in fetching and acquiring vast industrial data generated by several devices in the industry on the shop floor and retrieve necessary information based on context aware approach to create a smart enterprise based on industrial scenario. The paper presented a new solution for industry using IoT for predictive and remote maintenance provision for various industrial environmental parameters and assisting in increasing the work carried out by hand as well as productivity in industry using a machine learning approach. More specifically, this IIoT solution captures air quality, outdoor temperature humidity, boiling temperature from one sensor node and object detection, indoor temperature, humidity, smoke and light intensity data sensors from other sensor node in the system base on controllers and analyses them in the fog layer to provide a timely evaluation of intelligence require to operate the system which is helpful in increasing the productivity in the production line. The proposed experimentation illustrated the design of the IIoT solution, described the prototype industrial plant in normal and abnormal operation, analyzed with supervised machine learning approach and presented the sensor data analysis to create a smart enterprise.

Keywords Industrial data management system · IIoT solution · Cloud services industrial scenario · Smart enterprise

U. W. Hore (✉) · D. G. Wakde
Department of Electronics and Telecommunication Engineering, P.R. Pote College of Engineering and Management, SGBAU, Amravati, Maharashtra, India
e-mail: umeshhore@gmail.com

J. Hemanth et al. (eds.), *Intelligent Cyber Physical Systems and Internet of Things*,
Engineering Cyber-Physical Systems and Critical Infrastructures 3,
https://doi.org/10.1007/978-3-031-18497-0_65

1 Introduction

The IoT applications for industry is a significant importance as per as the introduction and integration of IoT is concerned. The intention of designing a industrial production systems is for progress-and not for failure-. Even though similar system can sustained little or no interruption or disruption. In IoT application standardized API or SDK is used to exchange data and instructions. These techniques are not enough as additional technologies in components such as communication protocols including wired and wireless, operating systems, databases, file systems to provide security and connecting devices like, smart phones, touch monitors and tablets. These are the top issue in industrial IoT environments where IoT technologies are required to solve context-specific requirements using the same setting. The main focus is on electronics components and data exchange, data storage and availability to analysis of data and prediction as well as user experience. Considering these all aspects introducing IoT in the industry is a risky job in terms of success since even most of the time incompatibility with pre-existing equipment, and not useful in a condition. Also various attribute like environment, accuracy, speed, latency also contribute a significant rise in workload as part of configuration, maintenance and operation. Even though someone manage to reach design requirements, it is not an easy job and that is why their may arise unrelated or insignificant issue. These issue resulted into non retrieval of required knowledge by data analysis and also not delivering the maximum user experience. Any of this above mentioned issue lead to the inadequate usage of IoT-based improvements and their eventual action [1]. Technology such as information and communication has brought new revolution in the operations of industry and productions. The size of industry is small or large certainly have the need of artificial intelligence and machine learning methods to process the large data generated by sensors, actuators in industrial management systems [2].

2 Related Work

IoT solution in industry used in application in determining the worker productivity in meat processing plant. The design and implementation based on highly accurate SVM and KPI models in determining the workers efficiency [3, 4]. Artificial Intelligence plays a significant role as it has a potential of large computational power resulted into building up a deeper neural network and could handle any hurdles [5]. Machine learning plays a major role using optimization methods responsible for rise in perceptive ability to have the potential of representation of learning and reasoning [6]. Highly accurate Light information from multiple location can be obtained with a few sensors from definite area. Method can be implemented with machine learning approach for very good approximation ratio compared to real ambient measurements [7, 8]. Pattern base classification plays a significant role in anomaly detection in industrial environment for normal and abnormal behaviour which is useful in identifying the faults, malfunctions and impact of bad maintenance [9]. FPGA technology, supported in some embedded System on Chips (SoC), for transferring power hungry and computationally intensive smart operations [10]. Machine learning approach is useful in smart grid where requirement is proper big data handling and data extraction [11]. Animal species can be detected using ANN method using acoustic signals and machine learning model for classification purpose based on sound is used [12]. Multilayer Network approach is used for retrieving and collection of huge industrial data generated in industry. The layer structure such as physical, network middleware, database and application are used [13, 14]. Predictive maintenance with machine learning approach played a significant role in Industrial IoT as a performance indicator in big data analysis [15]. End to end quality of service is also one of the requirement in IoT application and such applications are also useful to determine performance indicator for other application also [16]. Unusual event can be detected in industrial cyber-physical system based on IoT using machine learning approach [17, 18]. Present scenario in industrial IoT demands predictive maintenance rather than preventive maintenance which can be help in detecting the failure of machine and also estimate the time of failure [19, 20]. Abnormality among the sensor based data using machine learning approach using IoT architecture and predicting supplemental values using other correlational sensor [1, 21]. An advanced machine learning model has achieved the state of art performance for feature learning on big data computation using deep computation [22]. Automatic monitoring of fetal movement for signal processing using wearable system of accelerometer is also implemented as IoT application [23]. For İnternet of Things (IoT) technology ML and Big data analytics are powerful tool for analyzing and securing purpose [24]. Use of machine learning classifier and wrapper subset evaluation technique with Naive-Bayes and Meta bagging methods for feature selection is used in multidimensional IIoT feature extraction [25]. Manufacturing industry needs intelligent application using big data analytics and artificial intelligence on daily based requirement and are capable of meeting out the user needs using cloud computing [26]. Modern industries enabled with embedded sensor with cloud based solutions generated large data during business management. So these data should provide knowledge for the purpose of data capturing and analysis using IoT and big data analytics [27, 28]. After making the brief literature review common research problem is its a risky job to introduce IoT in industry since there are multiple point of failure. The technical solution that are built for existing system is not compatible since there are number of issues that are raised during implementation of IoT like environment, accuracy, latency, speed which should be considered during operation and maintenance. The motive of proposed task is to organise the efforts and skills which is needed to deploy an IoT platform in an industrial environment rapidly and without problems. The effort is to seek to achieve the design of the IIoT solution, described the prototype industrial plant during normal and abnormal operation, analyzed with supervised machine learning approach and presented the sensor data analysis to create a smart enterprise.

3 Proposed Work Methodology

In order to carry out this proposed work the flow of work consist of layer wise application oriented IIoT architecture under which system are built to acquire, interpret and expertised data using the arrangement of sensor and different layer architecture using supervised machine learning approach to create an intelligence. Different machine learning approach are used for analysis and monitoring the system and also evaluate various performance parameters.

3.1 Layer Wise Application Oriented IIoT Architecture

In order to carry out the proposed work the proposed methodology is to build the prototype using the arrangement of sensors based on industrial requirement. Layer wise application oriented architecture as shown in Fig. 1, which consist of things layer, edge layer, fog layer, communication layer and cloud services. After acquiring data from edge layer decision making model is developed at fog layer and sent data to cloud and monitoring system is developed with user interface.

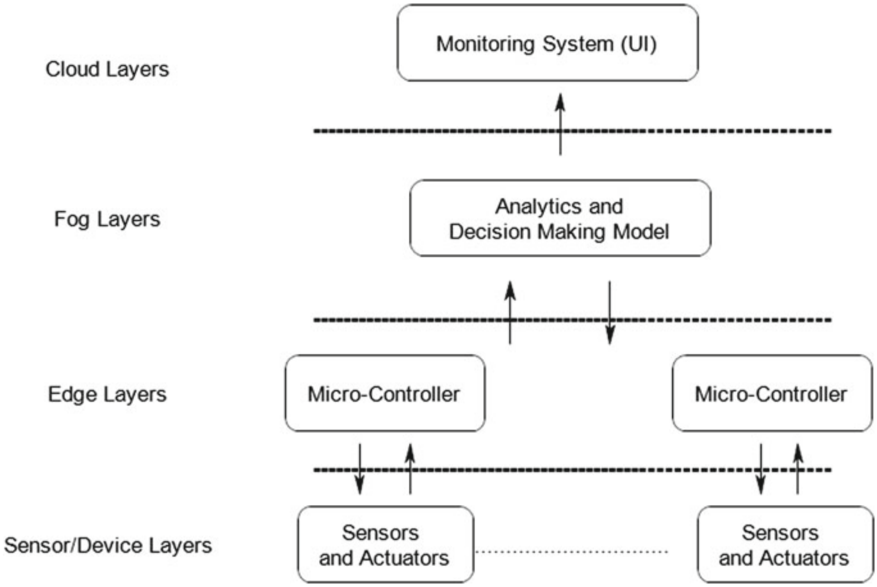


Fig. 1 Proposed layer wise application oriented architecture

3.2 Data Acquisition Using Experimental Set Up

The block diagram of experimental set up for data acquisition as shown in Fig. 2 where there are two Node of MCU ESP 32.The different sensors according to requirement are connected for indoor and outdoor environment in industry. The different sensors that are used in system are DHT11 Temperature and humidity, DS18B20, and MQ135 is a gas sensor for outdoor industrial scenario at node 1.Simillar sensor for indoor industrial scenario at node 2 are connected. DHT22 Temperature and Humidity sensor. MQ2 is useful for gas leakage detection in industry and helpful in dig out gases like H2, LPG, CH4, CO, Alcohol, Smoke or Propane. In Process industry capturing the light intensity of scenario during colour calibration measurement of item LDR is useful. Infrared proximity sensor such as E18 is used for existence of manufacturing object detection. The sensor which are connected at things layer to the microcontroller unit are useful in providing the information about the generated data. Embedded C programmes used for reading sensor data using Arduino IDE. IEEE 802.11 Wi-Fi are used for sending data through a Wi-Fi network. The supervisor is connected to the fog server for processing, and delivers data to the fog server. The work carried out by fog server is to first processes sensor data and temporarily stores it in a database before transferring data to the Cloud using MQTT protocol. A dashboard having user interface is used to display data on a, tablet, monitoring screen or mobile phone.

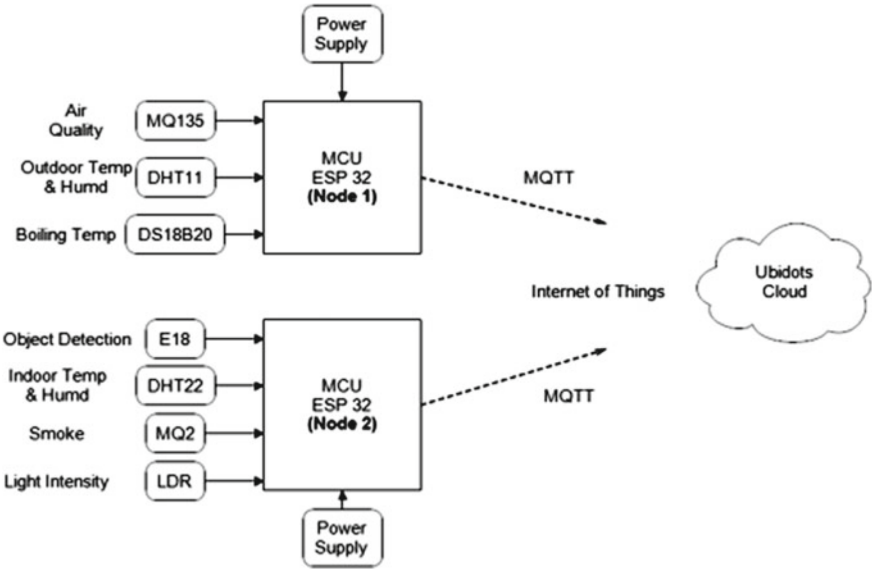


Fig. 2 Block diagram of IIoT experimental set up

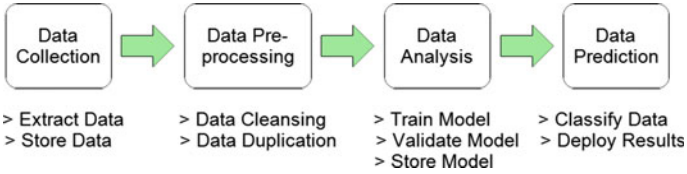


Fig. 3 Analytics and decision-making model

3.3 Data Analytics

The proposed methodology developed at fog layer for data analytics system using machine learning algorithms typically classification modeling is as shown in Fig. 3 and discussed in detail below.

Initially, generated data from sensors is acquired at the edge layer. Data gathered by IoT devices, particularly sensors, which can collect data in real-time or in small batches for two nodes. Data is collected and stored in a local target database. Preprocessing of stored data is used to clean and correct the data in which removal of irrelevant data, eliminate the duplicate copies of repeating data. Data classification depending on its intended purpose by initializing the classifier, later after training of machine learning algorithms, validation of classifier so as to store trained model. During prediction phase, calculations is performed on the classified data. Making decisions based on predictions and visualizing data in the form of reports or dashboards. Particularly, three machine learning algorithms is selected for implementation purpose, Support Vector Machine, Random Forest and Naïve Bayes algorithms.

3.4 Data Analytics Algorithm Steps

INPUT

"Features Set"

Input

node 1 parameters - air quality level (ppm), boiling temperature (Celsius), outdoor humidity (percent), outdoor temperature (Celsius).

node 2 parameters - object detection, smoke level (ppm), light intensity (lux), indoor humidity (percent), indoor temperature (Celsius)).

Output

Heating controller, ventilation fan, air conditioning, conveyer belt, light.

OUTPUT

Predicted Output with label values.

PROCEDURE

Step1: Collect, Prepare the feature data and label data from raw dataset values from Datasets.

Step2: Apply feature engineering to each feature data.

Find the missing and unknown values, replace the mean values.

Calculate the normalized value of all features set.

Scale all feature data into a specific range.

Step3: Select machine learning model for classification, SVM, RF and Naïve Bayes.

Step4: Choose the range of possible values for hyper-parameters of ML algorithms.

Step5: Optimize the values of the hyper parameters.

Step6: Evaluate and find the best score and best estimator for the selected classifier.

Step7: Validate the model using K-Fold Validation Learning Method.

Step8: Set best-selected hyper-parameters tuned for the ML training process.

Step9: Initialize the feature data and label data for the training dataset.

Step10: Train the model for respective ML algorithms.

Step11: Validate the model performance using the K-fold cross-validation method.

Step12: If validation is successful then save/deploy the trained model and if not repeat from steps 2 or 8.

Step13: Initialize the feature data for the testing dataset.

Step14: Load the trained model of ML algorithms.

Step15: Predict the results for its label values (classification).

Step16: Evaluate system performance using confusion matrix.

4 Results Analysis and Discussion

Sensor parameters as per requirement of industrial scenario in indoor and Outdoor situation are checked captured and examined in order to determine the success of the system for normal and abnormal case in the same environment and the representation of data have been statistically displayed on the cloud and also its presentation in numerical form is shown in Table 1.

Table 1 Mean value indoor and outdoor environmental parameter for a day of hours

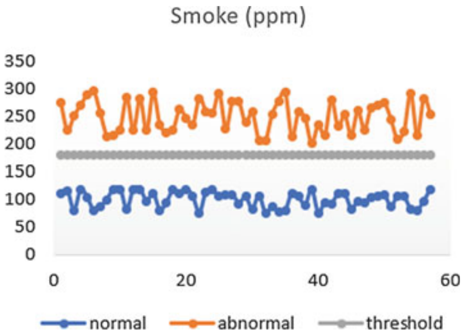
Scenario	Parameters	Units	Mean value	
			Abnormal	Normal
Indoor	Humidity	%	74.8	57.2
	Light intensity	Lux	502	988
	Object detection	Yes/no	–	–
	Smoke	Ppm	92	268
	Temperature	°C	22.5	31.5
Outdoor	Air quality	ppm	125.7	318.2
	Boiling temperature	°C	54.9	94.8
	Humidity	%	75.6	68.4
	Temperature	°C	28.1	38.7

Table 1 presents the mean values of indoor and outdoor parameter for normal and abnormal situations on regular basis for day of hours. these data represented used for anomaly detection in sensor data analytics for proposed system. The expertise data for different sensor arrangement is used for intelligence analytics using supervised machine learning approach and analysed for different parameters such accuracy, sensitivity, specificity, F-score, training and testing time. Figures 4, 5, 6 and 7 statistically display data on cloud for smoke, light intensity, temperature, humidity respectively for indoor environment.

Figures 8, 9, 10 and 11 statistically display data of boiling temperature, air quality, temperature and humidity respectively for outdoor environment.

After collecting data on a periodic basis for a day of hour data is trained and tested with different supervised machine learning approach with a combination of 70% for training and 30% for testing and analysed using various performance parameters. For that purpose Process1 and Process2 at node1 involved i/p from outdoor environment such as boiling temperature, air quality, temperature and humidity and for which system is making heating controller and ventilation fan on and off with the help of actuator as per intelligence detected. Similar Process1, Process2, Process3, at

Fig. 4 Smoke data indoor



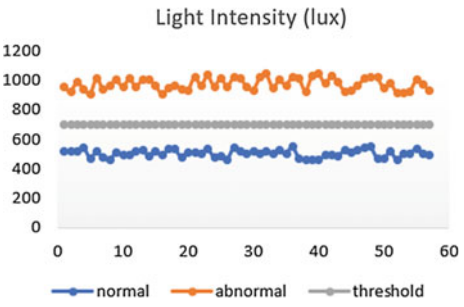


Fig. 5 Light intensity data indoor

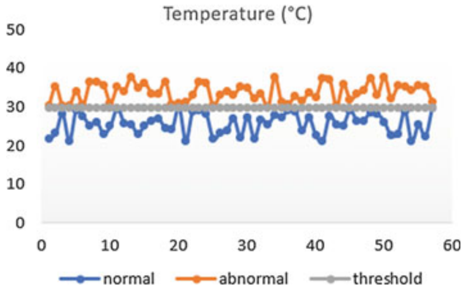


Fig. 6 Temperature data indoor

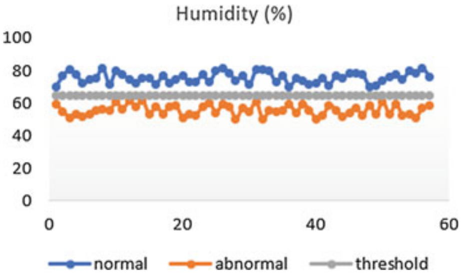


Fig. 7 Humidity data indoor

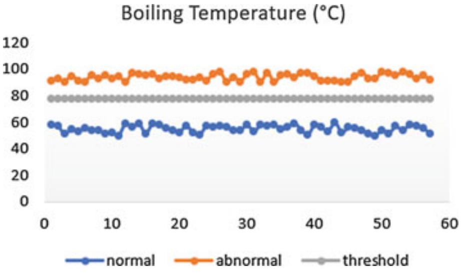


Fig. 8 Boiling temperature data outdoor

Fig. 9 Air quality data outdoor

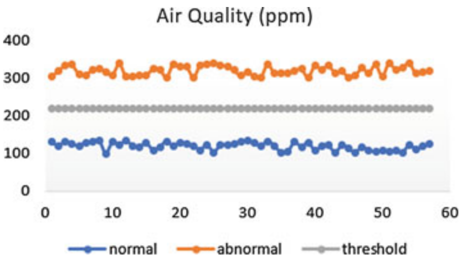


Fig. 10 Temperature data outdoor

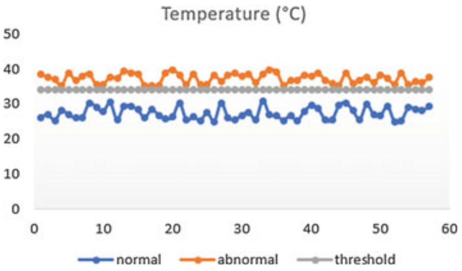
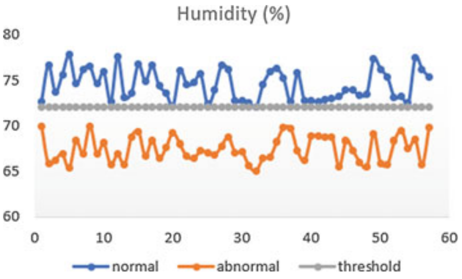


Fig. 11 Humidity data outdoor



node2 involved i/p from indoor environment such as object detection, smoke level, light intensity, indoor humidity, indoor temperature to make system on and off for conveyer belt, air conditioning, and light with the help of actuator as per intelligence detected.

After training and testing and validating the model the model is store and deploy the results for intelligence. The proposed experimental plan is implemented on a prototype that has been tested practically for various conditions. The programming of node controller is written in embedded C language with Arduino IDE. The analytics and decision-making model and classification is performed on a system having laptop with a 2.30 GHz Intel (R) Core (TM) CPU, 8 GB RAM, and Windows 10 (64 bit) operating system. The MATLAB IDE is used to program the intelligent model, in which statistics, and machine learning toolbox from MATLAB is utilized as tools (Tables 2 and 3).

Table 2 Training and testing time node 1

Performance/process		Process 1	Process 2
Training time (sec)	SVM	5.33	5.09
	RF	5.39	5.93
	NB	5.67	5.34
Testing time (sec)	SVM	5.03	4.85
	RF	5.13	4.74
	NB	5.35	4.53

Table 3 Accuracy, sensitivity, specificity, F-score for node 1

Performance/process	Method	Process 1	Process 2
Accuracy (%)	SVM	98.66	97.33
	RF	100	100
	NB	96.0	100
Sensitivity (%)	SVM	97.72	94.11
	RF	100	100
	NB	100	100
Specificity (%)	SVM	100	100
	RF	100	100
	NB	90.32	100
F-score (%)	SVM	99.53	98.76
	RF	100	100
	NB	94.82	100

The graphs of training and testing time, accuracy, sensitivity, specificity and F-score for node1 is shown in Figs. 12, 13, 14, 15 and 16 for node 1 (Tables 4 and 5).

The graph of training and testing time, accuracy, sensitivity, specificity and F-score for node 2 is shown in Figs. 17, 18, 19, 20 and 21 for node 2 (Fig. 22; Table 6).

5 Conclusion

The proposed experimentation illustrated the design of the IIoT solution, described the prototype industrial plant during normal and abnormal operation, analyzed with supervised machine learning approach and presented the sensor data analysis using different machine learning techniques with improving accuracy to create a context based smart enterprise which is useful in timely predictive maintenance and also

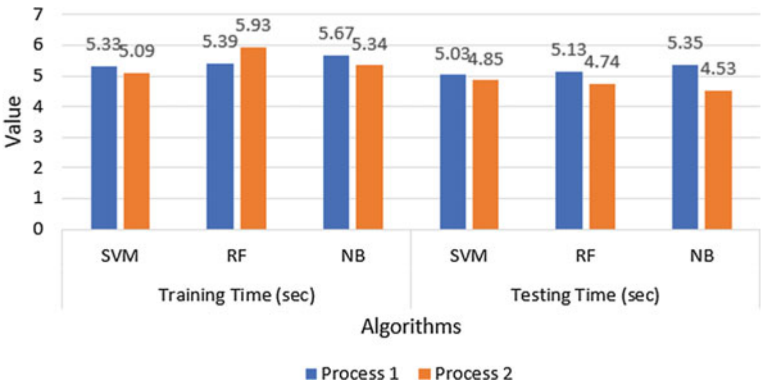


Fig. 12 Evaluation time for node 1

Fig. 13 Accuracy for node 1

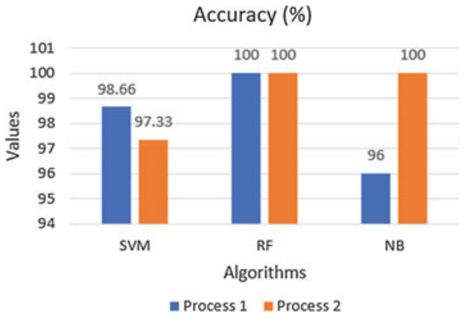


Fig. 14 Sensitivity for node 1

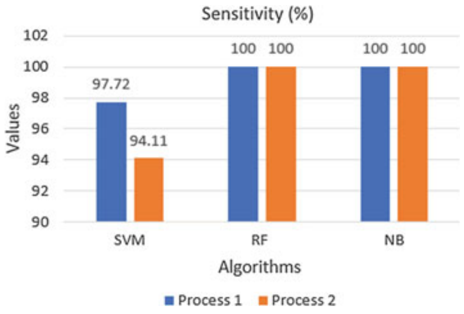


Fig. 15 Specificity for node 1

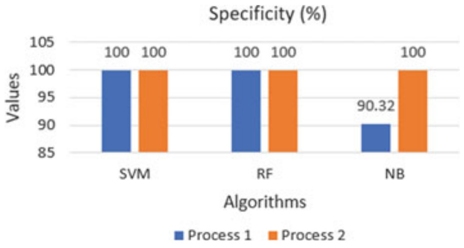


Fig. 16 F-score for node 1

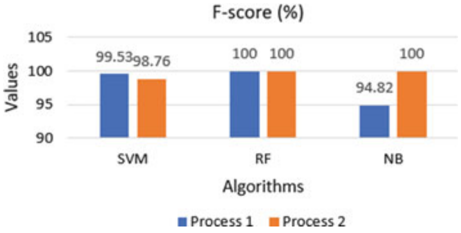


Table 4 Training and testing time for node2

Performance/process	Method	Process 1	Process 2	Process 3
Training time (sec)	SVM	5.03	5.39	5.29
	RF	5.99	5.68	5.38
	NB	4.92	5.31	4.46
Testing time (sec)	SVM	4.80	4.83	5.00
	RF	4.62	4.76	4.80
	NB	4.68	5.04	4.18

Table 5 Accuracy, sensitivity, specificity, F-score node2

Performance/process	Method	Process 1	Process 2	Process 3
Accuracy (%)	SVM	100	96	94.66
	RF	100	100	100
	NB	100	100	100
Sensitivity (%)	SVM	100	96.87	92.68
	RF	100	100	100
	NB	100	100	100
Specificity (%)	SVM	100	95.34	97.05
	RF	100	100	100
	NB	100	100	100
F-score (%)	SVM	100	94.51	96.44
	RF	100	100	100
	NB	100	100	100

useful in reducing the maintenance cost as well as supports for increase in productivity. Future scope of the work can be extended by increasing the no of parameters based on situations created in industrial environment using no of sensors and analyzed for intelligence.

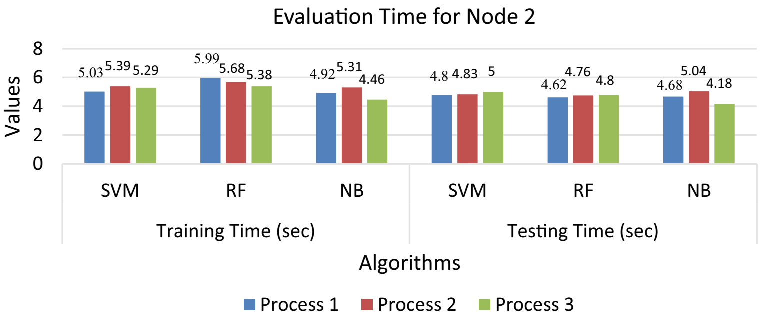


Fig. 17 Evaluation time for node 2

Fig. 18 Accuracy node 2

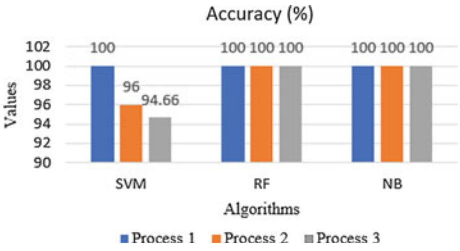


Fig. 19 Sensitivity node 2

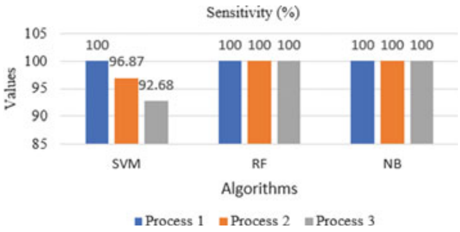


Fig. 20 Specificity node 2

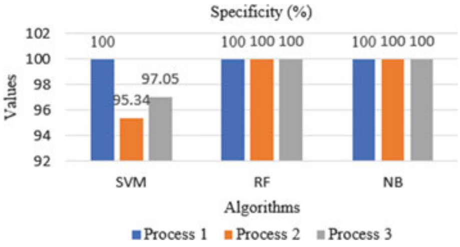
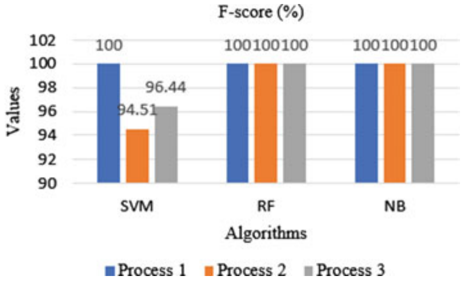


Fig. 21 F-score node 2



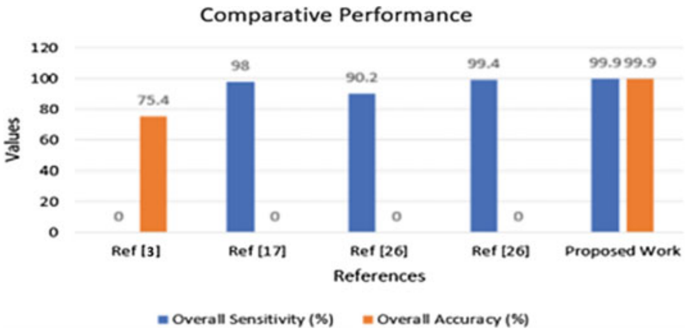


Fig. 22 Comparative performance of proposed method

Table 6 Comparative analysis

References	Models	Overall accuracy (%)	Overall sensitivity
Ref. [3]	Random forest	75.4	–
Ref. [17]	Multilayer perceptron	–	98
Ref. [25]	Classifier-subset with naive-bayes (CS-NB)	–	90.2
Ref. [25]	Subset with greedy-stepwise (Cfs-GS)	–	99.4
Proposed work	RF and NB	99.9	99.9

References

1. Vakaloudis A, O'Leary C (2019) A framework for rapid integration of IoT systems with industrial environments. In: 2019 IEEE 5th world forum on ınternet of things (WF-IoT), pp 601–605. https://doi.org/10.1109/WF-IoT.2019.8767224
2. Khan AI, Al-Badi A (2020) Open source machine learning frameworks for industrial Internet of Things. Procedia Comput Sci 170:571–577
3. Forkan ARM, Montori F, Georgakopoulos D, Jayaraman PP, Yavari A, Morshed A (2019) An Industrial IoT Solution for Evaluating Workers' Performance Via Activity Recognition. In: 2019 IEEE 39th international conference on distributed computing systems (ICDCS). Dallas, TX, pp 1393–1403

4. Ali MI, Patel P, Breslin JG (2019) Middleware for real-time event detection and predictive analytics in smart manufacturing. In: 2019 15th International conference on distributed computing in sensor systems (DCOSS), pp 370–376. https://doi.org/10.1109/DCOSS.2019.00079
5. Brennan RL (2019) AI, IoT hardware and algorithmic considerations for hearing aid and extreme edge applications. In: 2019 IEEE 62nd international midwest symposium on circuits and systems (MWSCAS), pp 841–844. https://doi.org/10.1109/MWSCAS.2019.8884886
6. Chen B, Wan J, Lan Y, Imran M, Li D, Guizani N (2019) Improving cognitive ability of edge ıntelligent IIoT through machine learning. IEEE Network 33(5):61–67. https://doi.org/10.1109/MNET.001.1800505
7. Drakoulelis M, Filios G, Ninos VG, Katsidimas I, Nikoletseas S (2019) Virtual light sensors in ındustrial environment based on machine learning algorithms. In: 2019 15th International conference on distributed computing in sensor systems (DCOSS), pp 709–716. https://doi.org/10.1109/DCOSS.2019.00126
8. Fahim M, Sillitti A (2019) Anomaly detection, analysis and prediction techniques in IoT environment: a systematic literature review. IEEE Access 7:81664–81681. https://doi.org/10.1109/ACCESS.2019.2921912
9. Ferrari P et al. (2019) Performance evaluation of full-cloud and edge-cloud architectures for Industrial IoT anomaly detection based on deep learning. In: 2019 II Workshop on metrology for ındustry 4.0 and IoT, pp 420–425. https://doi.org/10.1109/METROI4.2019.8792860
10. Fournaris AP, Alexakos C, Anagnostopoulos C, Koulamas C, Kalogeras A (2019) Introducing hardware-based intelligence and reconfigurability on industrial IoT edge nodes. IEEE Des Test 36(4):15–23. https://doi.org/10.1109/MDAT.2019.2908547
11. Hossain E, Khan I, Un-Noor F, Sikander SS, Sunny MSH (2019) Application of big data and machine learning in smart grid, and associated security concerns: a review. IEEE Access 7:13960–13988. https://doi.org/10.1109/ACCESS.2019.2894819
12. Hu NZ et al. (2019) Machine learning approach for robot diagnostic system. In: 2019 IEEE Eurasia conference on IOT, communication and engineering ECICE), pp 5–7. https://doi.org/10.1109/ECICE47484.2019.8942793
13. Saqlain M, Piao M, Shim Y, Lee JY (2019) Framework of an IoT-based industrial data management for smart manufacturing. J Sens Actuator Netw 8(2):25. https://doi.org/10.3390/jsan8020025
14. Liu Z et al. (2019) Intelligent station area recognition technology based on NB-IoT and SVM. In: 2019 IEEE 28th international symposium on industrial electronics (ISIE), pp 1827–1832. https://doi.org/10.1109/ISIE.2019.8781291
15. Liulys K (2019) Machine learning application in predictive maintenance. In: 2019 Open conference of electrical, electronic and information sciences (eStream), pp 1–4. https://doi.org/10.1109/eStream.2019.8732146
16. Sandström MK, Ericsson N, Rizvanovic L (2019) Analysing availability and QoS of service-oriented cloud for industrial IoT applications. In: 2019 24th IEEE international conference on emerging technologies and factory automation (ETFA), pp 1403–1406. https://doi.org/10.1109/ETFA.2019.8869274
17. Nardelli P et al. (2019) Framework for the identification of rare events via machine learning and IoT networks. In: 2019 16th International symposium on wireless communication systems (ISWCS), pp 656–660. https://doi.org/10.1109/ISWCS.2019.8877287
18. Qiu T, Wang H, Li K, Ning H, Sangaiah AK, Chen B (2019) SIGMM: a novel machine learning algorithm for spammer ıdentification in industrial mobile cloud computing. IEEE Trans Ind Inf 15(4):2349–2359. https://doi.org/10.1109/TII.2018.2799907
19. Shetty RB (2018) Predictive maintenance in the IoT era. IoT Predictive Maintenance and Services Group, SAP, San Francisco Bay Area, CA
20. Trakadas P et al (2020) An artificial intelligence-based collaboration approach in industrial IoT manufacturing: key concepts, architectural extensions and potential applications. Sensors 20:5480. https://doi.org/10.3390/s20195480

21. Tsai FK, Chen CC, Chen TF, Lin TJ (2019) Sensor abnormal detection and recovery using machine learning for IoT sensing systems. In: 2019 IEEE 6th international conference on industrial engineering and applications (ICIEA), pp 501–505. https://doi.org/10.1109/IEA.2019.8715215
22. Zhang Q, Yang LT, Chen Z, Li P, Bu F (April 2019) An Adaptive dropout deep computation model for industrial IoT big data learning with crowdsourcing to cloud computing. IEEE Trans Industr Inf 15(4):2330–2337. https://doi.org/10.1109/TII.2018.2791424
23. Zhao X et al. (2019) An IoT-based wearable system using accelerometers and machine learning for fetal movement monitoring. In: 2019 IEEE international conference on industrial cyber physical systems (ICPS), pp 299–304. https://doi.org/10.1109/ICPHYS.2019.8780301
24. Zolanvari M, Teixeira MA, Gupta L, Khan KM, Jain R (2019) Machine learning-based network vulnerability analysis of industrial Internet of Things. IEEE Internet Things J 6(4):6822–6834. https://doi.org/10.1109/JIOT.2019.2912022
25. Jayalaxmi PLS et al (2022) Machine and deep learning amalgamation for feature extraction in industrial Internet-of-Things. Comput Electr Eng 97:107610
26. Bashar A (2019) Intelligent development of big data analytics for manufacturing industry in cloud computing. J Ubiquit Comput Commun Technol (UCCT) 1(01):13–22
27. Suma V (2019) Towards sustainable industrialization using big data and İnternet of Things. J ISMAC 1(01):24–37
28. Hore UW, Wakde DG (2022) Context aware IoT enabled framework for monitoring parameters from ındustrial perspective. In: 2022 10th International conference on emerging trends in engineering and technology-signal and information processing (ICETET-SIP-22), pp 1–6. https://doi.org/10.1109/ICETET-SIP-2254415.2022.9791787