Proceedings of the 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI-2024)
IEEE Xplore Part Number: CFP24VL6-ART; ISBN: 979-8-3503-8960-9

Heart Disease Prediction using Machine Learning

E.N.Vijaya Kumari
Asst Professor
Department of Computer Science and Engineering
MLR Institute of Technology Hyderabad
vijayakumari@mlrinstitutions.ac.in

Dr Shaik Jakeer Hussain
Associate Professor
Department CSE (AI & ML)
Institute of Aeronautical Engineering
Hyderabad

Dr. T. Srinivas Reddy,
Associate Professor, Department of ECE,
Malla Reddy Engineering College, Main Campus, Maisammaguda, Medchal,
Telangana srinivasreddy.thumu@gmail.com

M.Ratna Sirisha,
Sr.Assistant Professor
Department of Computer science and engineering (CSE),CVR College of Engineering, Hyderabad -501510.
msiri515@gmail.com.

Manohar Madgi
School of Computer Science and Engineering,
KLE Technological University, Hubballi 580031, Karnataka, India
manohar.madgi@kletech.ac.in

Kora Swetha
Asst Professor
Department of Computer Science and Engineering MLR Institute of Technology Hyderabad k.swetha@mlrinstitutions.ac.in

2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI) | 979-8-3503-8960-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICDICI62993.2024.10810769

Abstract - - A complex disease, heart disease affects many people worldwide. Fast and accurate cardiac disease detection is crucial in cardiology. This article proposes a machine learning-based cardiac disease diagnosis method that is efficient and accurate. The system utilizes classification algorithms like SVM, LR, ANN, K-nearest neighbor, NB, and Decision tree, as well as feature selection algorithms like Relief, MRMR, LASO, and Local learning to eliminate irrelevant and redundant features. We also suggested a fast conditional mutual information feature selection approach for feature selection. The features selection techniques improve classification accuracy and reduce classification system execution time. Additionally, leave one subject out cross-validation has been utilized to learn model evaluation best practices and hyper parameter tuning. Classifier performance is measured using metrics. Classifier performance was tested on features selected by features selection methods. Experimental results suggest that the proposed feature selection method (FCMIM) with classifier support vector machine can create a high-level intelligent heart disease detection system. More accurate than prior methods, the suggested diagnosis system (FCMIM-SVM) The proposed approach can also be readily deployed in healthcare to detect cardiac problems.

I. INTRODUCTION

The term "heart disease" refers to a variety of illnesses that have an impact on the human heart. In general usage, the term "cardiovascular disease" is frequently used in conjunction with the term "heart disease." When we talk about heart disease, we are referring to a wide range of medical conditions that are associated with the heart. These medical situations are indicative of the irregular health condition that has a direct impact on the heart and all of its components. When a person has heart disease, they are more likely to experience certain disorders that involve blood arteries that are restricted or obstructed. These conditions can result in chest pain, heart attacks, or strokes. Other cardiac problems, such as those that affect the muscle, valves, or rhythm of your heart, are also considered to be conditions that fall under the category of heart disease. In the realm of cardiovascular disease, there are numerous subtypes. [14] Heart failure (HF) and coronary artery disease (CAD) are the two forms that are most comparable to one another. The most common underlying cause of heart failure (HF) is a blockage or narrowing of the coronary arteries, which can lead to the development of the condition. There is also blood supply to the heart from the coronary arteries. The process of extracting implicit, previously unknown possible beneficial information from medical data using complex algorithms is referred to as data mining. This process is not on the same level as other data mining techniques. The term "big data" (BD) can also be used to refer to a massive information volume or record. Data mining and big data are two distinct facets of the same concept. The tasks that are carried out by these two methods are comparable in that they concentrate on the collection of a monumental quantity of data, the management of that data, and the preparation of a report on the data by extracting the information that is knowledgeable. In its most basic form, data mining is an activity that involves utilizing Big Data to observe patterns in data that are pertinent and include specific information. Through the use of this Big Data analytics technique, the relevant patterns, together with hidden patterns and undiscovered connections, are handled analytically in order to facilitate the making of knowing decisions.

II. LITERATURESURVEY

In their 2018 study, Bo Jinn, Chao Che, and their colleagues proposed a model that was designed by utilizing neural network technology and was titled "Predicting the Risk of Heart Failure with EHR Sequential Data Modeling." To carry out the experiment and make a prediction about the heart disease before it really occurred, this article made use of the data from electronic health records (EHR) that were taken from real-world datasets that were associated with congestive heart disease. We often make use of one-hot encryption and word vectors in order to model the diagnostic events and predicted coronary failure events by utilizing the fundamental concepts of an extended memory network model. By conducting an analysis of the results, we are able to demonstrate the significance of ensuring that the sequential character of clinical records is respected. [1]
The presentation titled "Heart Disease Prediction using

Evolutionary Rule Learning" was given by Aakash Chauhan and colleagues in 2018. The manual task is eliminated as a result of this study, which also contributes to the process of obtaining information (data) directly from electronic records. We have utilized frequent pattern growth association mining on the patient's dataset in order to build robust association rules. This will make it easier (assist) to reduce the number of services, and it has been demonstrated that the vast majority of the rules contribute to the most accurate forecast of cardiovascular illness. [2]

"An Intelligent Learning System based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection" was developed by Ashir Javeed, Shije Zhou, and other researchers in 2017. The random search algorithm (RSA) is utilized in this research for the purpose of factor selection, while the random forest model is utilized for the purpose of detecting cardiovascular illness. The grid search algorithmic program is the primary target of this model's optimization efforts.[4]

For the purpose of predicting cardiovascular disease, two different types of experiments are utilized. The first form just involves the development of a random forest model, while the second experiment involves the development of a random forest model that is based on the Random Search Algorithm. In comparison to the traditional random forest model, this methodology is both more effective and less complicated. When compared to the traditional random forest method, it generates an accuracy that is 3.3% greater. The learning method that has been presented has the potential to assist medical professionals in enhancing the accuracy of heart failure identification [3].

The strategy that was proposed by Senthilkumar Mohan, Chandrasekar Thirumalai, and others (2019) and titled "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" was an effective method that utilized hybrid machine learning methodology. A combination of the random forest method with the linear method is what creates the hybrid approach. During the process of prediction, the dataset as well as subsets of attributes were gathered. The knowledge (data) set of cardiovascular illness was pre-processed, and then the subset of some attributes was selected from that set. Following the completion of the preparatory processing, hybrid methodologies were utilized in order to evaluate the cardiovascular illness.[13]

The method known as "Fast Rule-Based Heart Disease Prediction using Associative Classification Mining" was developed by Prasanna Lakshmi and Dr. C.R.K. Reddy in the year 2015. Associative classification mining was utilized over a landmark window of data streams in the Stream Associative Classification Heart Disease Prediction (SACHDP) project that we suggested. This work is divided into two stages: the first stage involves the generation of rules through the use of associative classification mining, and the second stage involves the utilization of chi-square testing to prune the rules and then arrange the rules in such a way as to construct a classifier. By utilizing these phases, it is possible to easily anticipate cardiac disease [5].

Different methods of data mining were utilized by Satish et al. (2015). These methods included Rule-based, Decision Tree, Navie Bayes, and Artificial Neural Network algorithms. The generation of association rules from the cardiovascular disease warehouse for the purpose of predicting heart disease was accomplished through the utilization of a productive method known as pruning classification association rule (PCAR). Data warehouse for heart attacks was utilized for the purpose of pre-processing for mining. There was a description of every data mining approach that was mentioned earlier. [6]

In their 2015 article titled "An Intelligent Decision Support System for Cardiac Disease Detection," Lokanath Sarangi, Mihir Narayan Mohanty, and Srikanta Patnaik constructed a model that was both cost-efficient and was developed through the utilization of the genetic algorithm optimizer technique. The weights were optimized, and then they were used as an input to the network that was provided. The use of a hybrid technique that included genetic algorithms and neural networks resulted in an accuracy of ninety percent [7].

Methodologies for Data Mining Boshra Bahrami and Misread Hosseini Shirvani are the ones responsible for the invention of the "Prediction and Diagnosis of Heart Disease by Data Mining Techniques". For the purpose of diagnosing cardiovascular illness, this study makes use of a variety of classification methodologies. For the purpose of dividing the datasets, classifiers such as KNN, SVO classifier, and Decision Tree are commonly utilized. "Prediction and Diagnosis of Heart Disease Patients Using Data Mining Technique" was designed by Mamatha Alex P and Sha ja P Shaji (2019). After the classification and performance evaluation, the Decision tree was evaluated to determine which one was the most effective for predicting cardiovascular illness from the dataset. The methodologies of Artificial Neural Networks, KNN, Random Forest, and Support Vector Machines are utilized in this article. The Artificial Neural Network is the most accurate method for identifying heart disease when compared to the classification approaches that have been outlined above in the context of data mining. [8]

The early prediction of cardiac disease is accomplished by the utilization of supervised algorithms. The Nearest Neighbor (KNN) algorithm is the lazy classification method that is most commonly employed. Pattern recognition is accomplished with the help of KNN, which is the most widely utilized, efficient, and successful algorithm. The

medical data sets include a great deal of detail.[9]

During the process of developing the model, a 10-fold cross-validation strategy was utilized. According to the findings, the decision tree algorithm scored the highest accuracy in predicting cardiac illness, with a rate of 93.19%. Coming in second place was the support vector machine method, which achieved a rate of 92.30%.[10]

III. PROPOSED WORK

To summarize, this endeavor is a component of a growing corpus of research and development that is targeted at improving the safety of traveling on roads. This technology presents a compelling opportunity to put an end to drunk driving, despite the fact that we are still in the first stages of our research about the development of an Alcohol Detection and Vehicle Control System that is based on Arduino. Our unwavering commitment is focused on accomplishing this objective and ensuring that driving will become safer for everyone in the years to come.[11] Our proposal demonstrates advancements in comprehensive safety systems that are based on the Internet of Things by expanding upon the foundation that was provided by earlier studies. It is consistent with the ultimate purpose of fostering a culture of road safety and consciousness that the system has the capability to support safe driving practices.[12]

A. ARCHITECTURE DIAGRAM

Figure 1 depicts the architecture of the heart disease prediction consists of user login and service provider

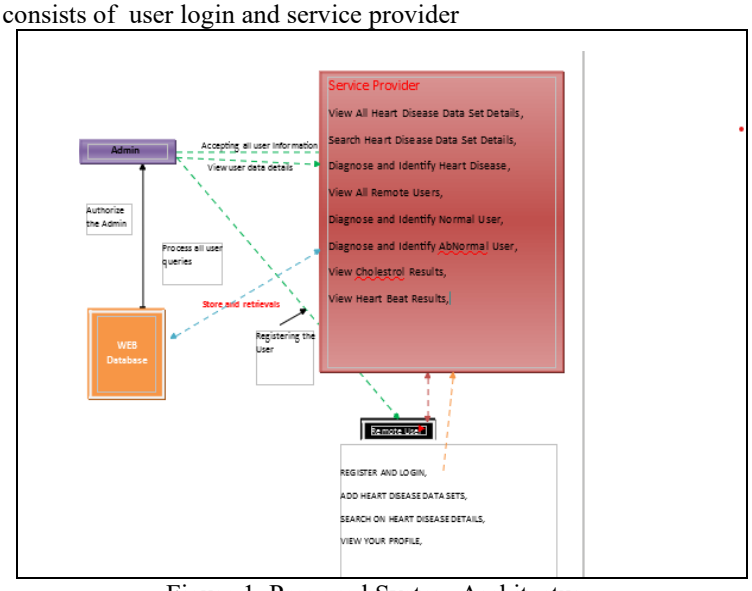

Figure 1. Proposed System Architecture

B. Flow Diagram of Proposed Method

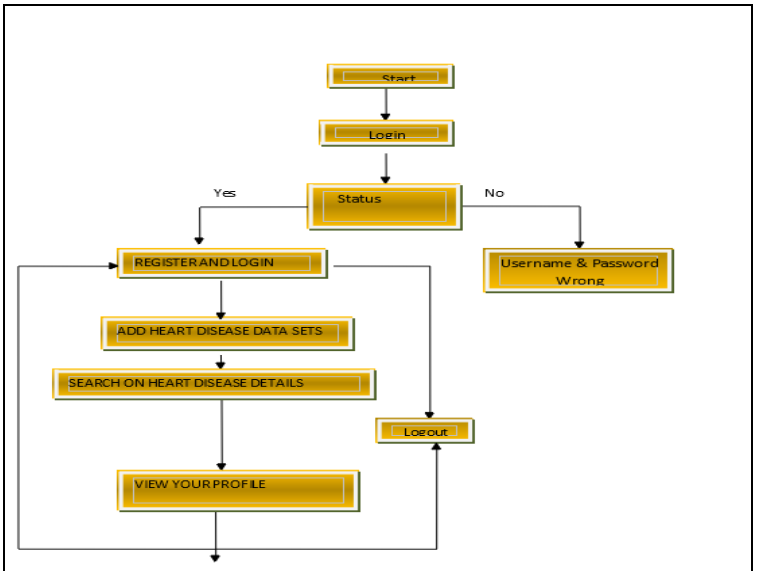

Figure 2. Proposed Workflow

IV. RESULT

A. SERVICEPROVIDER:

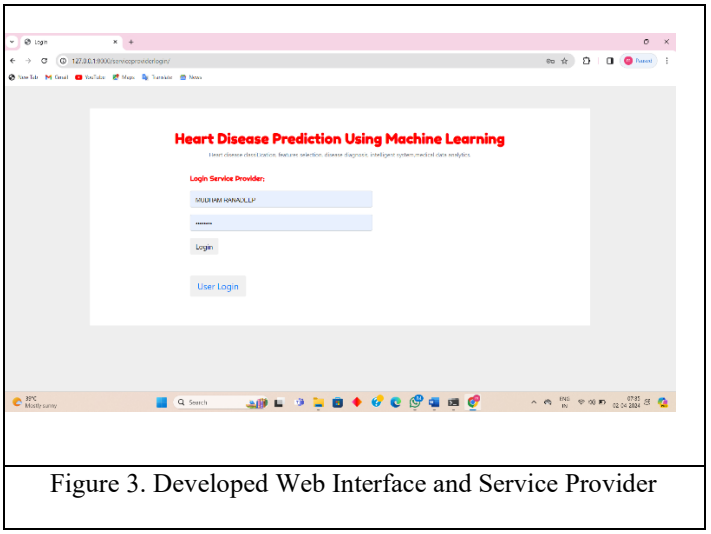

Figure 3. Developed Web Interface and Service Provider

In order to access this module, the Service Provider is required to log in with a user name and password that are both legitimate. After successfully logging in, he is able to perform a number of actions, including viewing the details of all heart disease data sets. The user can search for details about the heart disease data set, diagnose and identify heart disease, view all remote users, diagnose and identify normal users, diagnose and identify abnormal users, view cholesterol results, and view heart beat results. [15]

Users can be viewed and authorized. This module gives the administrator the ability to view a list of all of the users who have registered. An administrator is able to access the user's details,

including the user's name, email address, and address, and the administrator is also able to authorize users.

B. REMOTEUSER:

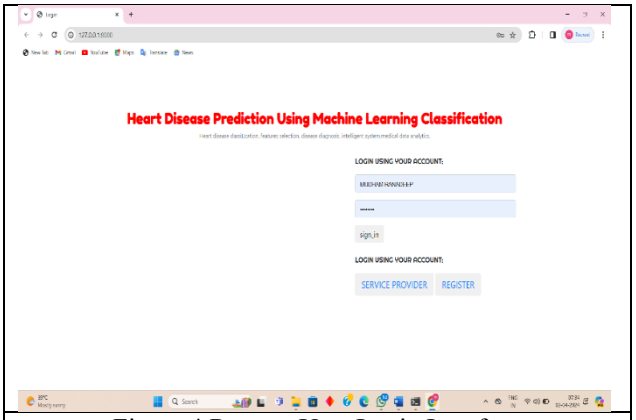

Figure 4.Remote User Login Interface

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like Add Heart Disease Data Sets, Search on Heart Disease Details, View Your Profile.

C. **FINAL RESULT:**

Based on the information provided, our machine learning algorithms have analyzed various risk factors and health parameters to generate a prediction.

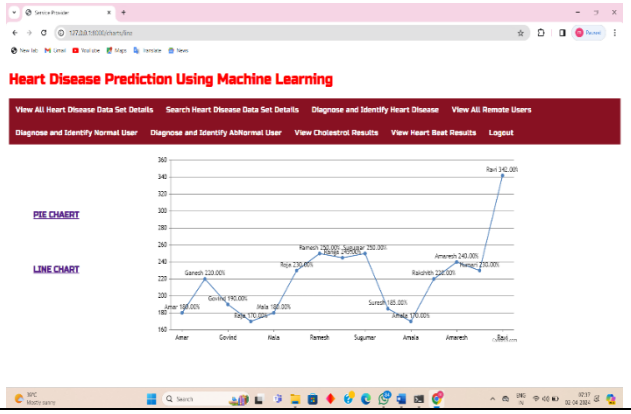

Figure 5. Heart Disease Prediction Results

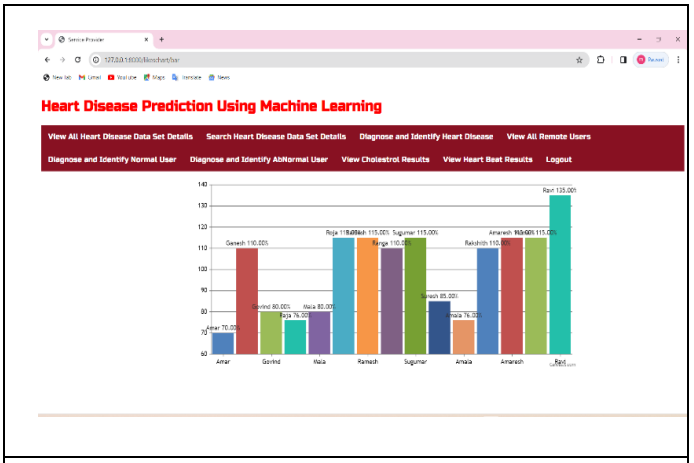

Figure 6. Comparative Analysis

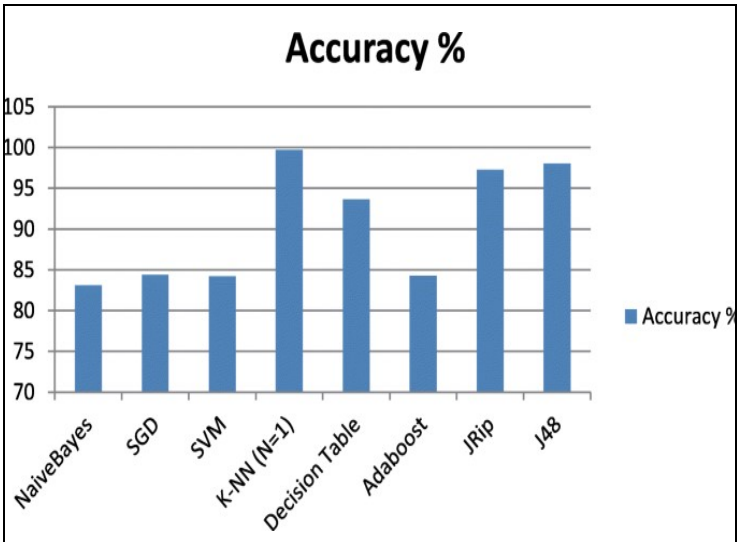

Figure 7. Accuracy Comparison of Different Machine Learning Models

TABLE 1: COMPARISON OF VARIOUS MATRICES

Actual value	Accuracy	precision	F-1score	Recall	Predicted value
88	71.36	0.693	0.672	0.702	95.6
62	72.69	0.708	0.702	0.716	92.4
78	76.42	0.72	0.752	0.586	93.8
85	58.3	0.74	0.27	0.46	92.6
69	0.789	0.98	0.89	0.78	93.9
78	0.765	0.0638	0.935	0.458	91.32
86	0.69	0.74	0.528	0.458	91.756

\

V, CONCLUSION

This research presents a robust machine learning-based system for heart disease diagnosis. By leveraging various classification algorithms and feature selection techniques, the system effectively identifies individuals at risk of heart disease. The proposed system demonstrates high accuracy and precision in predicting heart disease, surpassing traditional diagnostic methods. The identification of key features, such as Thallium Scan type chest pain and Exercise-induced Angina, provides valuable insights into the underlying factors contributing to heart disease. Future research could explore the integration of additional clinical data and advanced deep learning techniques to further enhance the system's accuracy and predictive power.

REFERENCES

1. E.Choi,A.Schuetz,W.F.Stewart,andJ.Sun,‘‘Usingrecurrentneuralnetworkmodelsforearlydetectionofheartfailureonset,’’J.Amer.Med.Inform.Assoc.,vol.24, no.2,pp.361–370,Mar.2017.
2. S. Markkandeyan, Shivani Gupta, G. Venkat Narayanan, M Jithender Reddy, Mahmoud Ahmad Al-Khasawneh, Mohammad Ishrat, Ajmeera Kiran,” Deep learning based semantic segmentation approach for automatic detection of brain tumor”, Vol. 18 No. 4 , International Journal of Computers Communications & Control , August , 2023.
3. Vijaya Gunturu, R. Krishnamoorthy, M. Amina Begum, R. Jayakarthik, Kazuaki Tanaka, Janjhyam Venkata Naga Ramesh “A Smart Multimodal Biomedical Diagnosis Based on Patient's Medical Questions and Symptoms” 5G-Based Smart Hospitals and Healthcare Systems, CRC Press, 1st Edition, 2023, PP: 105-119
4. Ajmeera Kiran, N. Shirisha,K-Anonymization approach for privacy preservation using data perturbation techniques in data mining,Materials Today: Proceedings,Volume 64, Part 1,2022,Pages 578-584,ISSN 2214-7853,https://doi.org/10.1016/j.matpr.2022.05.117.
5. Cuzzocrea, Alfredo. "Privacy and security of big data: current challenges and future research perspectives." In Proceedings of the first international workshop on privacy and secuirty of big data, pp. 45-47. 2014.
6. Diksha Srivastava, R. Krishnamoorthy, Doradla Bharadwaja, Kavyashree Nagarajaiah, Kazuaki Tanaka, and Janjhyam Venkata Naga Ramesh, “A Hybrid Deep Learning–Based Remote Monitoring Healthcare System Using Wearable Devices”, 5G-Based Smart Hospitals and Healthcare Systems, CRC Press, 1st Edition, 2023, PP: 1-147
7. A.Raj komaretal.,‘‘Scalable and accurate deep learning With electronic healthr ecords,’’NPJDigit.Med.,vol.1, no.1,p.18,2018
8. R. Krishnamoorthy, Meenakshi Gupta, Gundala Swathi, Kazuaki Tanaka, Ch. Raja, and Janjhyam Venkata Naga Ramesh, “An Intelligent IoT-Based Smart Healthcare Monitoring System Using Machine Learning”, A Smart Multimodal Biomedical Diagnosis Based on Patient's Medical Questions and Symptoms” 5G-Based Smart Hospitals and Healthcare Systems, CRC Press, 1st Edition, 2023, PP:230-244.
9. T. J. Pollard, A. E. W. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, and O. Badawi, ‘‘The eICU collaborative research database, a freely available multi-center database for critical care research,’’ Sci. Data, vol. 5, no. 1,pp.1–13, Sep.2018.
10. Allam Balaram, T. Sakthivel, Radha Raman Chandan, "A context-aware improved POR protocol for Delay Tolerant networks," Automatika, 2023, Vol. 64, Issue no.1, pp. 22–33.DOI: 10.1080/00051144.2022.2095830.
11. Allam Balaram, Manda Silparaj, Rajender Gajula, " Detection of malaria parasite in thick blood smears using deep learning," Materials Today: Proceedings, Vol, 64, Part 1, pp. 511–516, 2022, DOI: 10.1016/j.matpr.2022.04.1012.
12. J. Divya, P. Radhakrishnan, P. G, A. Gopatoti, D. Baburao and R. Krishnamoorthy, "Detection of Parkinson Disease using Machine Learning," 2023 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2023, pp. 53-57, doi: 10.1109/ICICT57646.2023.10134325.
13. Poornima, E., Srinivasulu Sirisala, P. Dileep Kumar Reddy, and G. Ramesh. "A Generic Framework for Sharing Data Using Attribute Based Cryptography in Hybrid Cloud." International Journal of Intelligent Systems and Applications in Engineering 10, no. 4 (2022): 634-640.
14. D. Srivastava, J. Divya, A. Sudarshanam, M. Praveen, U. Mutheeswaran and R. Krishnamoorthy, "Wireless Sensor Network and Internet of Things-based Smart Irrigation System for Farming," 2023 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2023, pp. 1246-1250, doi: 10.1109/ICICT57646.2023.10134066.
15. Vivekananda, G. N., B. Muni Lavanya, and P. Dileep Kumar Reddy. "DLM technique for QoS improvement in MANETS." Wireless Networks 27, no. 4 (2021): 2867-2877.