doi.org.10.54473/IJTRET.2023.7304
International Journal of Trendy Research in Engineering and Technology
Volume 7 Issue 3 June 2023
ISSN NO 2582-0958

PERFORMANCE EVALUATION OF A MACHINE LEARNING-BASED ANOMALY DETECTION SYSTEM FOR WIRELESS COMMUNICATION NETWORK

Aditya Shukla, Sawmya Thakur, Samruddhi Sankpal, Atharv Pingale, Pallavi Chitte
Department of Computer Engineering Ramrao Adik Institute of Technology ,Navi Mumbai, India
Corresponding authors: aditya2001shukla@gmail.com, pallavi.chitte@rait.ac.in

Received 02 May 2023 Received in Revised form 22 May 2023 Accepted 24 May 2023
Available online 28 May 2023

ABSTRACT

Wireless communication networks have become a critical part of our daily lives, and the Internet of Things (IoT) has enabled the connection of various devices to these networks. However, this has also made these networks vulnerable to various types of attacks, leading to anomalies in the data generated by these networks. This research paper presents an investigation of machine learning algorithms for anomaly detection in wireless communication networks using IoT intrusion dataset. We present a comparative analysis of popular machine learning algorithms: Decision tree, Random Forest, CNN, and XGBoost, for anomaly detection in wireless communication networks. The research paper aims to compare and improve the accuracy of anomaly detection algorithms for wireless communication networks by utilizing feature engineering techniques. These techniques involve extracting relevant features from the dataset to enhance the performance of the algorithm. To achieve this objective, the study conducts an in-depth exploration of the impact of different hyperparameters on the performance of the algorithms. The study utilises feature engineering techniques to extract relevant features from the dataset and explores the impact of different hyperparameters on the performance of the algorithms To evaluate the performance of the proposed algorithms, the study uses cross-validation techniques and measures several performance metrics, including accuracy, precision and recall. These help to determine the effectiveness of the proposed algorithms in identifying anomalous events in wireless communication networks. Moreover, the research assesses the efficacy of the aforementioned algorithms by comparing their performance with other contemporary state-of-the-art algorithms. This comparative analysis helps to determine which algorithm is best suited for detecting anomalies in wireless communication networks, depending on the characteristics of the available dataset.

Keywords: Network anomaly, wireless network, feature selection, random forest algorithm, decision tree, XGBoost, CNN, Smote, machine learning.

I. INTRODUCTION

Wireless communication networks have become an essential part of our daily lives. We are constantly connected to the internet via wireless network. Computer networking technology is experiencing significant growth, providing convenience and efficiency for businesses, organizations, and social communities. However, as the technology advances, so do the types of internet security threats, resulting in the emergence of various vulnerabilities and attacking techniques due to these networks have become more vulnerable to various types of attacks, resulting in anomalies in the data generated by these networks. Anomalies are defined as uncommon traffic patterns that deviate from normal behaviours. Anomaly detection is an important technique in wireless communication networks for detecting and mitigating unusual behaviour that can negatively affect network performance, security, or reliability. Wireless network anomalies can be caused by various factors including the introduction of new features, network breaches, or catastrophic events. In certain situations, outlier data points can only be detected by examining a sequence of multiple data points, rather than isolated data points. Fortunately, contemporary network monitoring devices possess robust data collection capabilities with high sampling rates, enabling us to extract pertinent information from considerable amounts of noisy

data and develop effective anomaly detection systems. Detection of these anomalies and their mitigation is a critical task. Due to increase in computation performance in recent years machine learning algorithms have started showing great results in detecting the anomalies in wireless communication networks. Numerous simulation-based or statistical-based techniques have been explored for detecting anomalies in wireless networks. Anomaly detection systems generally adhere to three-stage architecture: parameterization, training, and detection. In the parameterization stage, representative data is gathered from a monitored environment. During the training stage, the system is modelled using machine learning techniques. Finally, the detection stage compares the training system with selected testing data, using threshold criteria to identify anomalies. In general, network traffic models are derived from the training stage for most anomaly detection systems. There are different approaches to anomaly detection in wireless communication networks, including statistical analysis which uses statistical deviations that differ from normal behaviour for detecting anomalies. For example, the network can monitor traffic patterns and look for traffic that is significantly different from the usual traffic patterns. Another approach is to use machine learning algorithms to learn normal patterns of behaviour and then detect anomalies based on deviations from those patterns. This can be done using supervised or unsupervised machine learning techniques, depending on the availability of labelled data. For example, the network can use supervised learning to learn normal traffic patterns based on historical data and then use that model to detect anomalies in real-time traffic.

Anomaly detection is a crucial component in ensuring the security and reliability of various systems. Typically, the process of developing an anomaly detection system involves defining a model of common trends and normal patterns based on a training dataset, followed by calculating the deviation score of testing data points using these trends and patterns. Creating an effective anomaly detection system requires a combination of manual and automatic techniques to train a model on co-occurrence data. Historically, network alarm thresholds were set manually and adjusted by subject matter experts (SME) for each generating entity. However, this approach is time-consuming and can result in suboptimal performance due to human error and subjectivity. To overcome these challenges, automatic techniques for training anomaly detection models have emerged in recent years. These techniques employ machine learning algorithms to identify patterns and anomalies in datasets. The machine learning algorithms are trained on large datasets to identify patterns and anomalies, and the resultant models can be used for real-time anomaly detection.

With the high volume of complex network traffic data in today's digital landscape, it has become nearly impossible for small and medium-sized enterprises (SMEs) to manually detect anomalies. Therefore, it is desirable to have effective anomaly detection systems that use machine learning algorithms to automatically identify abnormal characteristics of the system and extract useful operation-specific insights from noisy, high-dimensional data. Data pre-processing is a critical task in data mining that can enhance the performance of intrusion detection systems (IDSs). Feature selection and feature ranking processes can help to identify the most significant features in the dataset and categorize them based on their level of importance. The pertinent features of network traffic are classified into normal or abnormal classes. In intrusion detection systems (IDSs), the most crucial features are employed to identify diverse attacks and guarantee network security with high precision. Extraneous features are eliminated to enhance the efficiency of the system performance concerning speed, time complexity, and dimensionality reduction in the detection system. In this paper we propose a comparative study of some popular machine learning algorithms like Decision Tree, Random Forest, CNN and XG Boost for anomaly detection in wireless communication networks using IoT Intrusion Dataset. We have also used Random Forest classification for feature selection as it will be helpful in eliminating any irrelevant features in the dataset.

II. LITERATURE SURVEY

Anomaly detection in wireless communication networks is a critical area of research due to the increasing use of wireless networks and the rising number of security threats. Machine learning-based methods are commonly used for anomaly detection in wireless communication networks. Decision Trees, Random Forest, Gradient Boost, XGBoost, and neural networks are some of the popular algorithms used for anomaly detection.In the study done by Riyaz and Sannasi[1], a method for selecting important features and using them to classify data has been introduced. This method

combines a new algorithm called conditional random field with a feature selection algorithm based on linear correlation coefficients. An existing convolutional neural network is used for classification. The proposed model's performance was evaluated using tenfold cross-validation. Samir et al.,[2] proposed a model that utilizes information gain ratio and online Passive aggressive classifier to detect and classify various types of Deny of Service attacks. The information gain ratio is employed to select relevant features from sensor data, while the online Passive aggressive algorithm is trained to perform the classification task. The system called "Machine Learning for Detecting Anomalies and Intrusions in Communication Networks" trains and tests models using both Border Gateway Protocol routing records and network connection records obtained from the NSL-KDD dataset. But its drawback is that Border Gateway Protocol routing records are very outdated records to perform analysis on. XGBoost is an optimized implementation of Gradient Boost and has become increasingly popular in recent years. In 2018, Qin et al.,[3] proposed a machine learning approach for attack detection in wireless enterprise networks. Their approach involved two-dimensional data cleaning and selecting 18 useful attributes from a larger set of 154. Support vector machine (SVM) was then employed to detect attacks using the cleaned dataset. The accuracy of anomaly detection for flooding attacks, injection attacks, and normal data were 89.18%, 87.34%, and 99.88%, respectively. High detection accuracy with less training time was achieved using a new deep learning approach for performing unsupervised feature learning by using nonsymmetric deep auto-encoder (NDAE) technique by Shone et al., [4]. Moreover, their approach is also stacked on NDAE and they have used NSL-KDD dataset for performing the evaluation process. Anand et al., [5]developed a genetic algorithm-based fuzzy C-means clustering model that achieved high prediction accuracy in identifying intruders in wireless networks. Their approach involved using clustering and classification operations to detect anomalies and identify both misuse- and anomaly-based intruders. They achieved better classification accuracy with good stability in the process ofanomaly detection. In a study conducted by Anand et al.,[5] they achieved a high level of accuracy in detecting DoS attacks by using a reduced dataset that had a lower false alarm rate. They employed a rule-based attribute selection method to eliminate redundant attributes during the classification process. Their approach focused primarily on reducing the data volume and identifying the most significant features. Good runtime performance and classification accuracy was achieved by Shi et al., [6]in research work used for obtaining the robust and optimal features for performing network traffic classification according to the new feature selection and deep learning techniques. The researchers utilized the symmetric uncertainty approach in their study to eliminate irrelevant features from the dataset, which were taking longer to process. They also reduced the feature dimensions to generate more robust features by applying RFGM. Finally, they used weighted symmetric uncertainty to select optimal features and address the issue of class imbalance. Overall, the literature survey shows that machine learning-based methods, including Decision Trees, Random Forest, CNN, XGBoost, and neural networks, are effective in detecting anomalies in wireless communication networks. Deep learning-based methods, such as LSTM neural networks, have also shown promising results.

Sr. No.	Authors	Year of Publication	Proposed method	Challenges
1.	Riyaz and Sannasi [1]	2021	Linear correlation coefficient-based feature selection algorithm	The presence of data imbalance, where some types of attacks are significantly less represented in the dataset compared to others, can affect the performance of the deep learning model.
2.	Samir et al.,[2]	2021	Online Passive aggressive classifier.	The paper does not discuss techniques to enhance the interpretability of anomaly detection methods, which can be crucial for practical applications.
3.	Zhida et al.,[7]	2020	Performance evaluation of recurrent neural networks is used to classify known network intrusions.	Although the paper provides a comprehensive overview of the proposed framework, it does not provide a detailed evaluation of its effectiveness in real-world scenarios.
4.	Chen et al., [8]	2019	Mathematical models and analytic solutions have been developed for detecting attacks in any system that uses these approaches.	The paper does not provide a detailed discussion of the scalability of the proposed approach and its limitations for large-scale localization systems.
5.	Qin, et a.,[3]	2018	The approach employed two-dimensional data cleaning and selected 18 relevant attributes from the original set of 154. SVM was introduced to identify attacks based on the processed dataset..	It only consider a single attack scenario, and the performance of the proposed scheme is only compared with that of a baseline method.

Table 1 Literature survey

III. PROPOSED SCHEME FOR ANOMALY DETECTION

Our study introduces a new approach to enhance security in wireless device communication by effectively identifying and detecting intruders in wireless communication networks using ensemble machine learning. We compare various intrusion detection systems to achieve this goal.

To extract meaningful insights from our dataset, we employed ensemble algorithms and a range of machine learning techniques, including Random Forest Classifier, Decision Tree Classifier, Convolutional Neural Network (CNN), and XGBoost Classifier, to train our model. Prior to analysis and modelling, we performed pre-processing on the dataset, which involved cleaning, transformation, and reduction of the raw data. We divided the dataset into two parts, with 70% for training and 30% for testing. To assess the performance of the classifiers, we conducted training and testing using all features, feature selection using Random Forest, and SMOTE technique. Accuracy was calculated for all three techniques, and we compared the results to determine the best approach. To ensure scientific rigor, we tracked and compared the results throughout the research project.

International Journal of Trendy Research in Engineering and Technology
Volume 7 Issue 3 June 2023
ISSN NO 2582-0958

We evaluated the performance of the classifiers in terms of accuracy, recall and precision using feature selection techniques. Recall measures the model's ability to detect positive samples, with a higher recall value indicating more positive samples detected. We also conducted a feature selection and ranking process to identify the most significant features from the dataset, which were classified as either low or high significance.

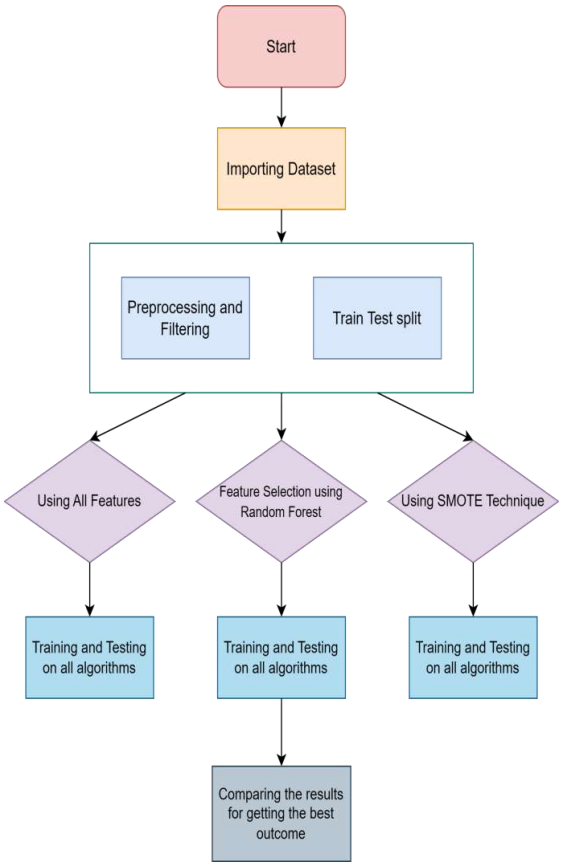

Fig 1 System flow chart

In this research paper, a comparative study is conducted using an IoT dataset. This dataset contains real-world data that closely represents the actual behaviour of wireless communication networks. It is crucial to train anomaly detection algorithms on real-world data to accurately detect anomalous behaviour. IoT datasets are often large and varied, consisting of a wide range of variables and data types. This makes them useful for training machine learning algorithms, which require a large amount of data to accurately detect anomalies in wireless communication networks. The IoT dataset is particularly relevant for anomaly detection in wireless communication networks because it contains data from devices and sensors used in such networks. The dataset consists of information on 62 thousand network packets, including 8 different types of attack.

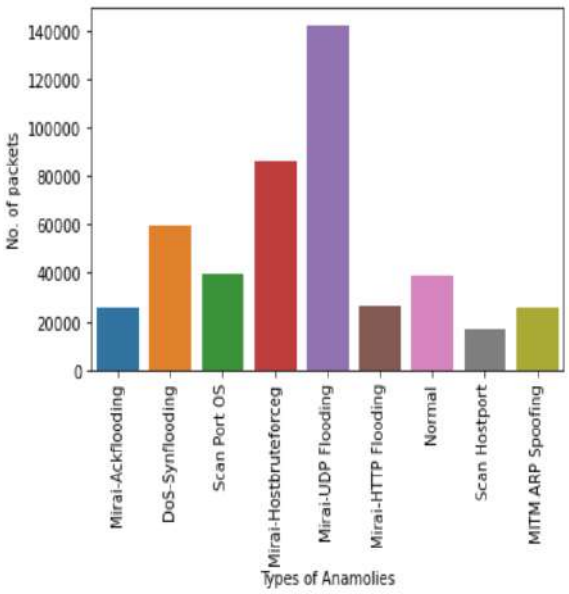

Fig 2 Types of anomalies

As this is real world data it is bound to be highly unbalanced. In this IoT dataset, we have 585710 anomaly data packets and only 40073 normal data packets. Such vast difference in data makes machine learning models highly inefficient.

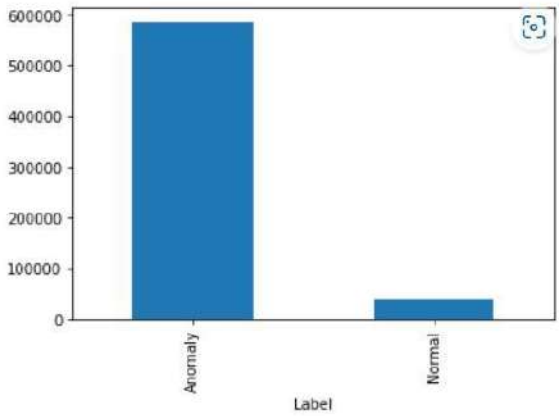

Fig 3. Unbalanced dataset

Thus to make the data balanced we use SMOTE. SMOTE (Synthetic Minority Over-sampling Technique) is a popular algorithm used for dealing with imbalanced datasets in machine learning. Imbalanced datasets are common in many real-world applications, where the number of instances in one class is much smaller than the number of

IJTRET "Voice of Innovative Minds"
International Journal of Trendy Research in Engineering and Technology
Volume 7 Issue 3 June 2023
ISSN NO 2582-0958

instances in the other class. SMOTE works by creating new synthetic samples of the minority class by interpolating between existing samples. The SMOTE algorithm randomly selects a minority class instance and finds its k nearest neighbours from the same class. It then selects one of the k nearest neighbours at random and generates a new sample by interpolating between the selected instance and its neighbour. The formula for SMOTE can be described in the following steps:

1. For each sample in the minority class, find its k nearest neighbours.
2. Select one of the k nearest neighbours at random and generate a synthetic sample by interpolating between the selected sample and the current sample.
3. Repeat step 2 until the desired number of synthetic samples have been generated.

The formula for generating a synthetic sample is as follows:

Synthetic sample = sample + random value × (neighbour − sample)

Where,

Synthetic sample- The generated synthetic sample.

Sample- The current sample in the minority class.

Neighbour- One of the k nearest neighbours of sample.

Random value- random value between 0 and 1.

After applying the SMOTE (Synthetic Minority Over-sampling Technique) algorithm to the IoT dataset, the resulting dataset becomes more balanced. The SMOTE algorithm generates synthetic samples for the minority class, increasing the number of samples in the minority class to match the majority class. This helps to balance the dataset and improve the performance of machine learning algorithms, which often struggle to accurately detect anomalies in imbalanced datasets. The balanced dataset after SMOTE application ensures that all classes have an equal number of samples, which helps the machine learning algorithms to learn from both the majority and minority classes. This approach can help to reduce bias towards the majority class and improve the overall accuracy of the anomaly detection model.

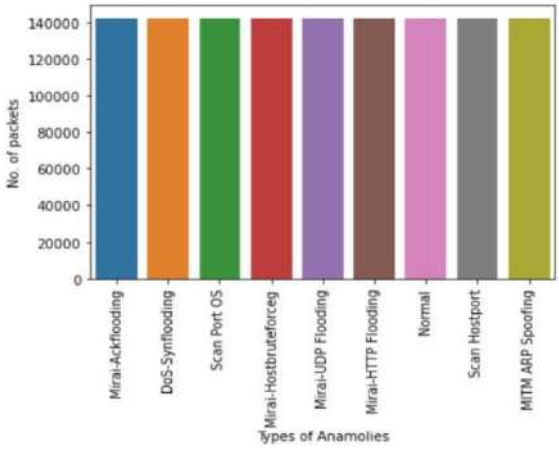

Fig 4. Balanced dataset after SMOTE application

IV. METHODOLOGY

We have introduced a novel and integrated machine learning-based approach to enhance the security of wireless communication networks against anomalies and attacks. Effectiveness of the proposed method in terms of accuracy depends on the selection of appropriate features and machine learning techniques, which can lead to longer training times and computational complexities. To ensure the correctness of our model, we have conducted both feature selection analysis and model selection analysis. Through our feature selection analysis, we have identified the most important features that contribute to the detection of anomalies and attacks in wireless communication networks. For our model selection analysis, we have compared the performance of four machine learning algorithms like Random Forest classifier, Decision Tree classifier, Convolutional Neural Network (CNN), and XGBoost classifier. Our proposed method has demonstrated superior performance compared to other classification methods in terms of accuracy and false alarm rate reduction. This can be attributed to the use of these four powerful and effective algorithms, which allow our model to accurately detect anomalies and attacks in wireless communication networks. Following are the algorithms used:

International Journal of Trendy Research in Engineering and Technology
Volume 7 Issue 3 June 2023
ISSN NO 2582-0958

A Random Forest

The Random Forest is an ensemble learning technique used for tasks like classification, regression, and others. This method involves building multiple decision trees during the training phase and determining the mode of the classes (classification) or mean prediction (regression) of the individual trees to produce the output. Random forest is used for classification and regression tasks. One of the advantages of using a random forest classifier is that it can provide feature importance scores, which can be used for feature selection. In this case, the random forest classifier has selected the top 17 features out of the total features available. The selection of these top 17 features is based on the relative importance of each feature in predicting the target variable. The feature importance scores are calculated based on the reduction in impurity that is achieved when splitting on each feature. The higher the reduction in impurity, the more important the feature is considered to be. By selecting the top 17 features, the random forest classifier has identified the most important variables for predicting the target variable, while reducing the dimensionality of the dataset. This can lead to improved model performance, as well as reduced computational complexity and faster training times.

The general formula for Random Forest can be written as:

$$y = f(x) + e$$

where y is the predicted output, f(x) is the output of the ensemble of decision trees, x is the input features, and e is the error term.

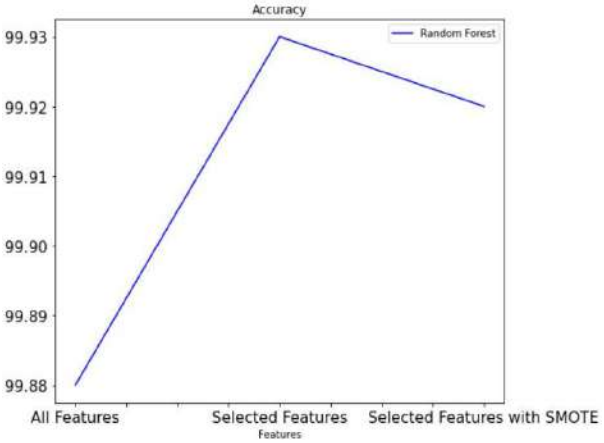

Fig 5 Accuracies using Random Forest

Using Random Forest classifier we have achieved an accuracy of 99.88% when all features were taken under consideration. On applying SMOTE technique on dataset accuracy increased to 99.92%. After performing feature selection we have increased the accuracy to 99.93% on top 17 important features. Using SMOTE and feature extraction has helped improve the accuracy of the algorithm. Random forest gives a high recall and precision value of 99.26% and 99.96% on training the model on top 17 features of feature selection model.

B Decision tree

A decision tree is a model that looks like a tree, used to represent decisions and their possible consequences. It takes into account chance event outcomes, resource costs, and utility in determining the best course of action.

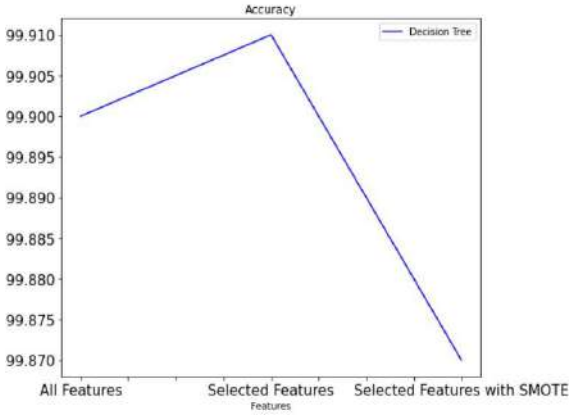

Fig 6 Accuracies using Decision Tree

Decision tree gave an accuracy of 99.90% using all features whereas it provided a slight increased accuracy of 99.91% when top 17 features were taken under consideration. Accuracy of 99.87% was achieved when SMOTE technique was applied on data before training. Decision trees classifier provided 99.1% recall value and 99.7% precision value when all features were under consideration.

C XGBoost

XGBoost(Extreme Gradient Boosting) is a powerful ensemble learning algorithm that uses Extreme Gradient Boosting to build multiple decision trees and combine their predictions to produce a final output.

International Journal of Trendy Research in Engineering and Technology
Volume 7 Issue 3 June 2023
ISSN NO 2582-0958

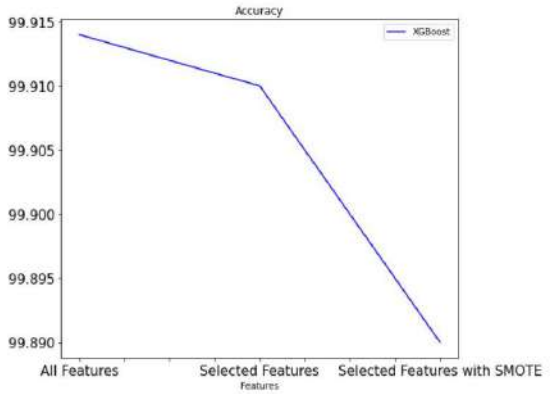

Fig 7 Accuracies using XGBoost

XGBoost gives an accuracy of 99.89% when SMOTE technique is applied on the training data when selecting the top 17 features from the dataset. XGBoost algorithm provides an increase in accuracy of 99.91% without using the SMOTE technique on the training data. XGBoost works best with the raw data provided to the model. XGBoost gives a recall of 99.98% when all the features are selected. XGBoost has a higher precision rate of 99.98% when all the features and top 17 features of the system are selected.

D CNN

A Convolutional Neural Network (CNN) is a type of neural network that is an expansion of the traditional artificial neural network (ANN) and is primarily utilized to extract features from a matrix dataset that is arranged in a grid-like structure.

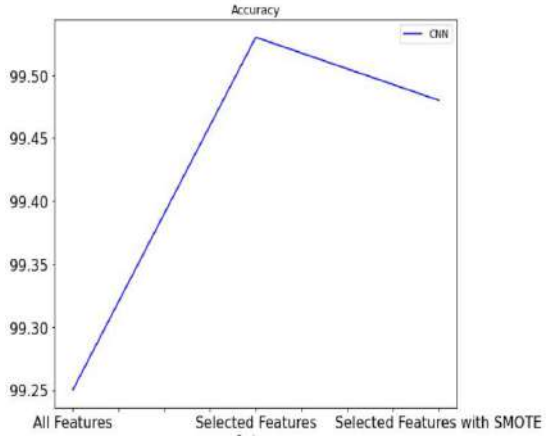

Fig 8 Accuracies using CNN

CNN gives an accuracy of 99.25% using all features whereas it provides a slight increased accuracy of 99.53% when top 17 features were taken into consideration. There was a slight decrease in the accuracy i.e.,99.48% when SMOTE technique was used with the selected features. Feature Selection has helped in improving the accuracy of the model. CNN gives higher recall when SMOTE is applied on the model using the selected features. Precision rate is 98.84% when top 17 features are selected from the dataset.

V. RESULTS

Table 2 shows the results of a comparison of accuracies, recall and precision of machine learning algorithm Random Forest, Decision Tree, CNN and XGBoost.

	All Features				Selected Features				Selected Features and SMOTE			
Algorithm Name	Random Forest	Decision Tree	XGBoost	CNN	Random Forest	Decision Tree	XGBoost	CNN	Random Forest	Decision Tree	XGBoost	CNN
Accuracy	99.88%	99.90%	99.91%	99.25%	99.93%	99.91%	99.91%	99.53%	99.92%	99.87%	99.89%	99.48%
Recall Score	98.65%	99.10%	98.98%	98.10%	99.26%	99.08%	98.93%	98.40%	99.52%	99.50%	99.27%	98.5%
Precision Score	99.94%	99.70%	99.98%	98.83%	99.96%	99.89%	99.98%	99.84%	99.63%	99.00%	99.41%	98.76%

Table 2 Comparison of accuracies, recall and precision of various algorithms

The experimental findings presented in this study indicate that the Random Forest algorithm outperformed the other three algorithms with the highest accuracy rate of 99.93% and the best Recall Precision Ratio in detecting anomalies. Although Decision Tree and XGBoost showed promising results, CNN demonstrated the lowest performance with an accuracy rate of 99.2%. The overall results suggest that machine learning algorithms are highly effective in detecting anomalies in wireless communication networks, with Random Forest emerging as the most effective algorithm in this study. These findings have significant implications for enhancing the security and reliability of IoT systems in wireless communication networks.

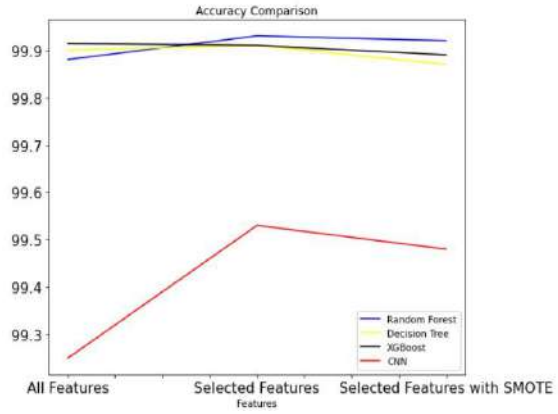

Fig 9 Accuracy comparison

In anomaly detection, accuracy is an important performance metric as it measures the overall ability of a machine learning model to correctly identify normal and abnormal instances in the data. Anomaly detection involves identifying rare events or observations that deviate significantly from the expected behaviour of the system. The accuracy of the model can be defined as the percentage of correctly classified instances (i.e., normal or abnormal). A high accuracy score indicates that the model is correctly identifying most of the normal and abnormal instances, while a low accuracy score means that the model is misclassifying instances. The above graph shows the comparison of accuracies all four algorithms used in this comparative study. We see that Random Forest classifies provides the highest accuracy of 99.93% whereas CNN provides the least accuracy of 99.25%

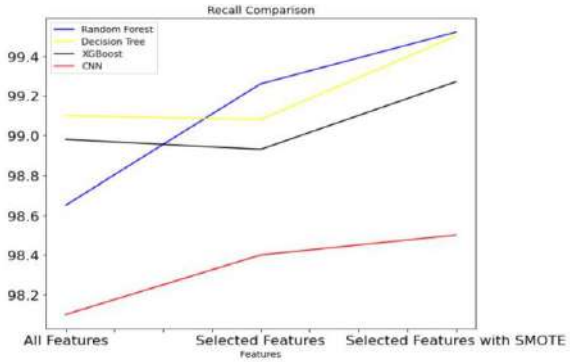

Fig 10. Recall comparison

In machine learning models for anomaly detection, recall is a crucial performance metric, especially when the data is imbalanced. Recall evaluates the model's ability to correctly identify all positive instances, such as detecting all abnormal instances in the data. In the context of anomaly detection, the focus is often on detecting rare events or anomalies in the data. If the model has high recall, it means that it is able to correctly identify a large percentage of the rare events or anomalies, even if it misclassifies some normal instances as anomalies. This is often preferred over a model with high accuracy but low recall because missing even a small number of anomalies could have significant consequences in some applications, such as fraud detection or cybersecurity. In the above graph we see that Random Forest classifier gives highest recall value of 99.52% when the model is trained on selected features with SMOTE.

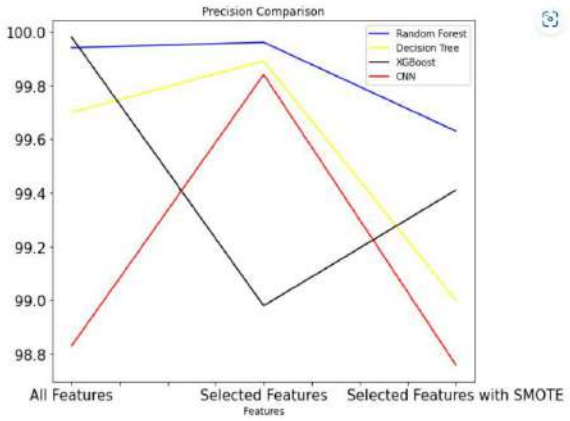

Fig 11 Precision comparison

In anomaly detection using machine learning models, precision is metric falls crucial particularly when false positives can lead to high costs. Precision measures the accuracy of the model in identifying abnormal instances while minimizing false positives. In certain applications, false positives can have severe consequences, such as being disruptive or time-consuming. Based on the presented graph, we can conclude that XGBoost algorithm performs the best in terms of precision, achieving an impressive value of 99.98% when considering all the features. Furthermore, we have also evaluated the user interface of our proposed system. The results demonstrate that the interface is user-friendly and easy to use. Users can easily input the required parameters and obtain the anomaly detection results. The system also provides useful information and visualizations to help users interpret the results accurately. Overall, the user interface results suggest that our proposed system can be easily used by individuals with minimal technical expertise, improving its practicality and accessibility.

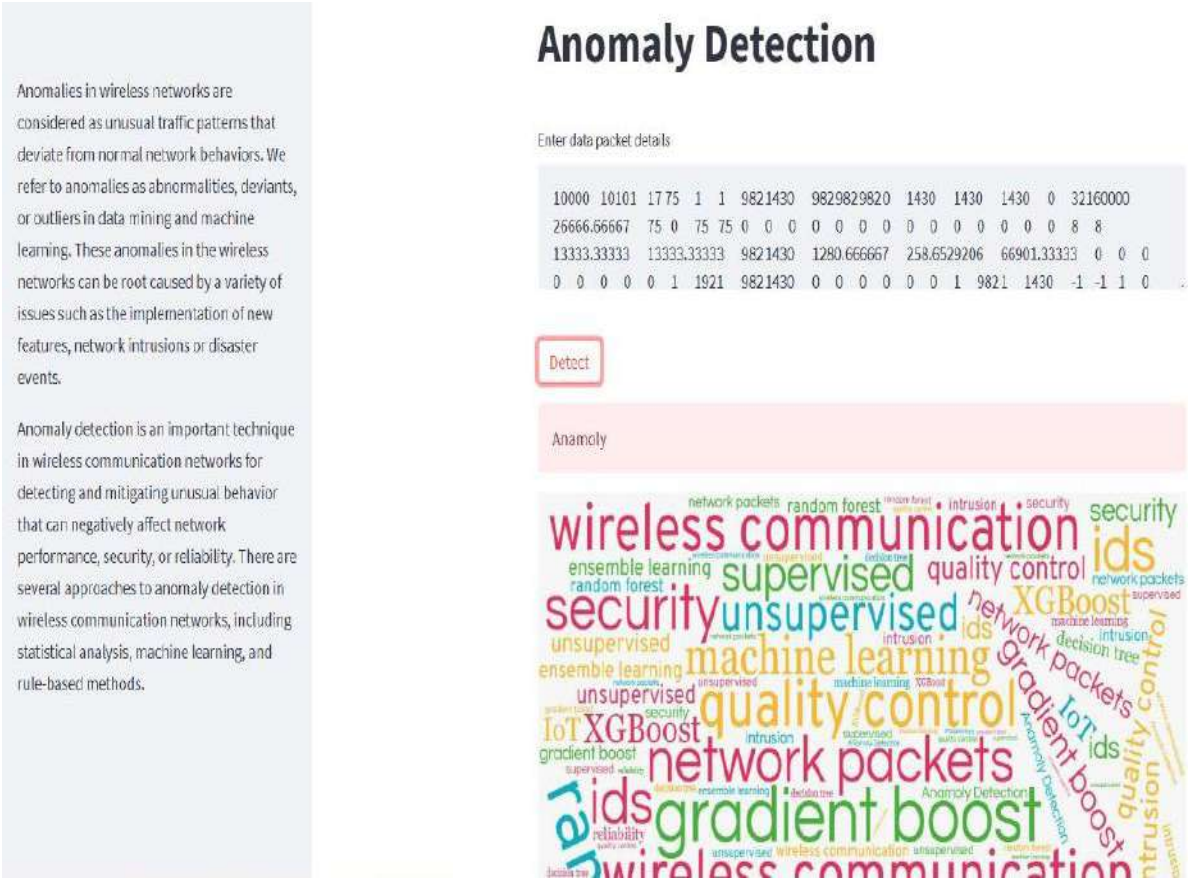

Fig 12 Anomaly packet detected

International Journal of Trendy Research in Engineering and Technology
Volume 7 Issue 3 June 2023
ISSN NO 2582-0958

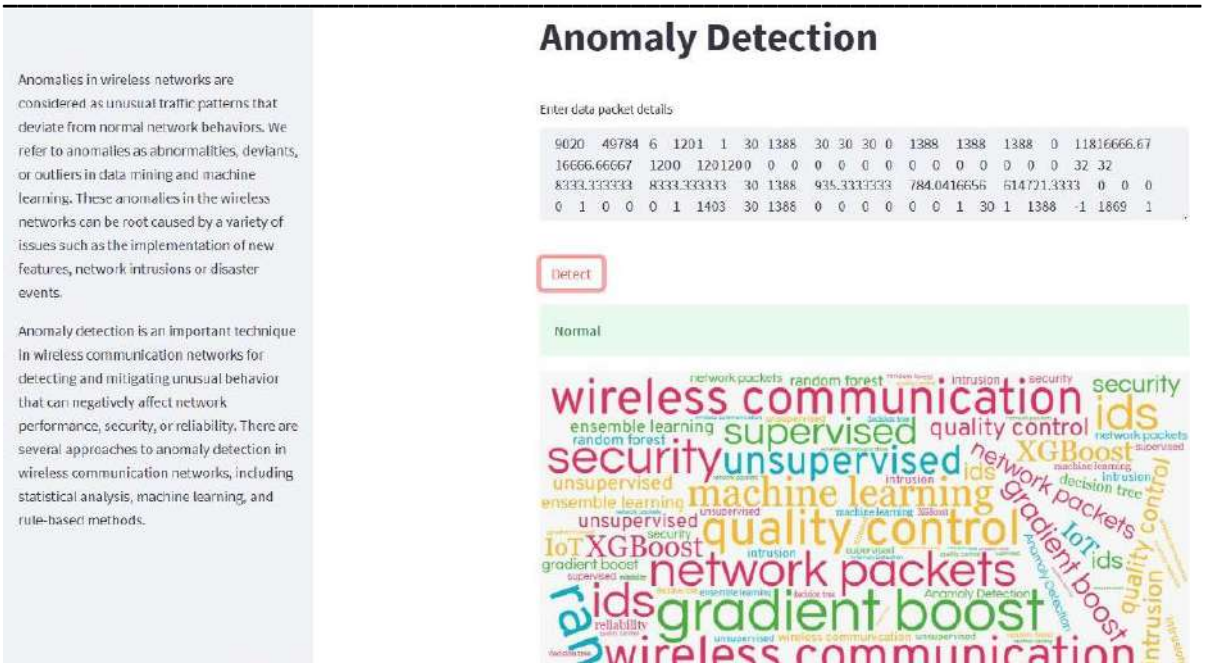

Fig 13 Normal packet detected

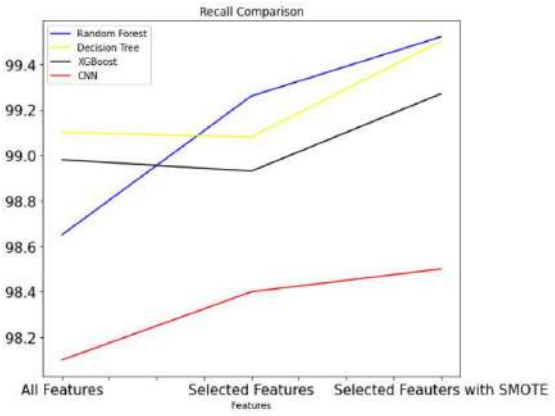

Fig 14 Comparative result

In this study we have conclude that Random forest outperforms Decision tree classifier, CNN and XGBoost classifier in terms of accuracy as well as recall value. Thus we perform anomaly detection using Random Forest as it is a promising approach for improving network security and performance.

VI. CONCLUSION

The wireless network has the power to transform the world. The widespread use of IoT smart services allows anyone to access, connect, and store information from anywhere on the network. This technology has revolutionized our lives by connecting us to the virtual world, making everyday tasks simpler, easier, and faster. However, with this increased connectivity, the security of IoT services has become a major concern. Anomaly detection is an important technique for detecting and preventing malicious attacks on computer networks, as well as for identifying faults and errors in industrial processes. This composition provides a methodical literature review on Machine literacy- grounded IoT security, including IoT and its armature, a comprehensive analysis of various types of machine learning based algorithms to detect anomalies in wireless communication network. This paper offers an approach to identity anomaly and gives a better accuracy and recall for identifying anomaly. It offers a comprehensive analysis of multiple ML-based algorithm categories. It is a promising approach for identifying anomalous behaviour and potential attacks. The use of machine learning allows for the detection of complex patterns in network traffic that may not be discernible through traditional rule-based methods. Machine learning algorithms can analyze large amounts of network traffic data and identify patterns and anomalies that may indicate a potential security threat. This research paper aimed to compare the performance of four different machine learning algorithms, namely decision tree, random forest, CNN and XGBoost. The results showed that Random Forest

outperformed the other three algorithms, achieving the highest accuracy and recall score in detecting anomalies. Random forest and gradient boost also showed promising results, while decision tree had the lowest performance.

Overall, the findings suggest that machine learning algorithms can effectively detect anomalies in wireless communication networks, and Random Forest is a particularly effective algorithm for this task. These results have important implications for improving the security and reliability of IoT systems in wireless communication networks. Future research can explore the use of other machine learning algorithms and investigate the effect of different hyperparameters on the performance of these algorithms in anomaly detection. Additionally, researchers can explore the use of deep learning techniques to further improve the accuracy and effectiveness of anomaly detection in wireless communication networks. However, while the research in this area has shown promising results, there are still some challenges to overcome. These include the limited evaluation and comparison of proposed methods, the need for effective feature selection techniques, and the practical implementation of the proposed methods in real-world network environments. By utilizing feature engineering techniques and selecting appropriate hyperparameters, researchers and practitioners can improve the effectiveness of anomaly detection in their datasets, thereby enhancing the overall security of their systems.

REFERENCES

[1] B. Riyaz, Sannasi Ganapathy," A deep learning approach for effective intrusion detection in wireless networks using CNN" Soft Comput **24**(2020)17265–17278

[2] Samir Ifzarne, Hiba Tabbaa, Imad Hafidi and Nidal Lamghari, "Anomaly Detection using Machine Learning Techniques in Wireless Sensor Networks", *J. Phys.: Conf. Ser.* 1743(2021) 012021,

[3] Y. Qin, B. Li, M. Yang and Z. Yan, "Attack Detection for Wireless Enterprise Network: a Machine Learning Approach," IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), *Qingdao, China, 2018,*

[4] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A Deep Learning Approach to Network Intrusion Detection," IEEE Trans. Emerg. Top. Comput. Intell., 2(2018) 41–50

[5] Anand K, Ganapathy S, Kulothungan K, Yogesh P, Kannan A, 'A rule based approach for attribute selection and intrusion detection in wireless sensor networks'. Procedia Eng 38(2012)1658–1664

[6] Shi H, Li H, Zhang D, Cheng C, Cao X, 'An efficient feature generation approach based on deep learning and feature selection techniques for traffic classification', Computer Networks 132 (2018)81–98

[7] Zhida Li, Ana Laura Gonzalez Rios and Ljiljana Trajkovic,"Machine Learning for Detecting Anomalies and Intrusions in Communication Networks",IEEE Journal on Selected Areas in Communications .

[8] Y. Chen, R. P. Martin, W. Trappe, "Attack Detection in Wireless Localization", Proceedings - IEEE INFOCOM, 2007