Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 4 | April-2026
ISSN (O) 3107-6696

HYBRID MACHINE LEARNING FRAMEWORK FOR IDENTIFICATION AND DETECTION OF UNAUTHORIZED WI-FI ACCESS POINTS

Mareswaramma Pilli[1], Jyothi Pravallika Reddy[2], Pujitha Chowdary Rama[3], Chandu Vallabhapuram[4], Syed Muhammad[5], Sadu Gnaneswar[6]

[1,2,3,4,5,6]*CSE(AI&ML) & Dhanekula Institute Of Engineering And Technology, Ganguru*

Abstract - Wireless networks are increasingly being compromised by Wi-Fi access points which execute man in the middle attacks, eavesdrop on conversations as well as stealing passwords. Traditional methods of detection which involve application of fixed rules or signatures tend to miss out the new or unknown threats. To detect the unauthorized Wi-Fi access points, this research paper provides a hybrid machine learning approach to analyze behavioral traffic. A preprocessing pipeline is leakage-free and it comprises of data cleaning, removal of suspicious features, feature encoding, scaling, and feature selection which increases model reliability. XGBoost, random forest and Linear Support Vector machine are used to classify access points as safe or rogue. In order to reduce the false negatives and to enhance the accuracy of detection, the predictions of such models are used in an ensemble voting process. Evaluation is done with the AWID Wi-Fi intrusion dataset that consists of 154 traffic characteristics. The proposed architecture is very accurate and recalls high in experiments. With the help of Flask web application, trained models would be deployed to enable detection of rogue access points in real-time.

Keywords— *Wireless Network Security, Rogue Access Point Detection, Machine Learning, Ensemble Learning, XGBoost, Random Forest, Support Vector Machine, Intrusion Detection System.*

I. Introduction

Wireless networks are necessitated in the contemporary communication systems to allow flexibility and practicality of internet in homes, workplaces and public places. The rapid development of wireless technology has however brought security weaknesses about. This issue is critical due to the potential man-in-the-middle attacks, password theft, and network data interception that the unauthorized or rogue Wi-Fi access points can execute. Network security demands that rogue access points be identified and these are access points that act as legitimate networks to steal confidential information.

Traditional ways of identifying malicious access points are based on signature or rules. Such techniques often fail to detect new or complicated threats, and can only detect known attacks. To circumvent these limitations, scholars have employed machine learning to determine illegal entry points by examining the network traffic and patterns of wireless traffic. Doyeon Kim et al. suggested a machine learning-based method to detect unauthorized access points, which enhances the information security by detecting abnormal network behavior [1]. In order to improve the accuracy of detecting rogue access points in wireless networks, Wang et al. used fine-grained channel information [2].

Various studies have used machine learning in improving the intrusion detection algorithms in network security. Gondal et al. network intrusion detection system diversity-based centroid approach improves the identification of harmful activity of a network [3]. Han et al. [4] proposed a timing-based technique of detecting rogue access points in a network by monitoring wireless transmission time. Kitisriworapan et al. [5] have developed a client-side technique of evil-twin attacks, which target a legitimate wireless network and steal user data.

Despite recent progress, many systems continue to experience high levels of false negative rates, computational complexity as well as not responding to new attack patterns adequately. To address such issues, the hybrid machine learning approach to the behavioral traffic analysis-based detection of unlawful Wi-Fi access points is proposed in this paper. The proposed approach utilizes ensemble voting, feature engineering, and a number of machine learning models to enhance the detection accuracy and reliability. It is suggested that the proposed approach can identify rogue access points and allow real-time deployment with the help of a web interface based on the test results performed on the AWID Wi-Fi intrusion dataset.

II. Literature Survey

To identify malicious actions, Gondal et al. designed the network diversity-centrogroid intrusion detection system. It is a tool that compares centroid based normal and malicious points of data with the perspective of identifying abnormal patterns of traffic. Compared to the method, which is more effective in enhancing detection, the method might need extra processing of vast quantities of network traffic. [6].

Buczak and Guven have written widely on machine learning and data mining to intrusion detection system in cybersecurity. This paper will mention the benefits and drawbacks of the supervised and unsupervised network attack detection classification and anomaly detection. According to the report,

machine learning, which is state of the art, can make intrusion detection system more accurate and tailored. [7].

The study by Mathur and Badone examined how machine learning can be applied in prediction and categorization. The research explains how this complex information should be analyzed and offers comparisons. In this study, it is demonstrated that correct algorithms and preprocessing techniques improve the proper classification of the data. [8].

Alotaibi and Elleithy defined and studied the rogue access points detection in wireless networks. The study entails signal, traffic and machine learning detection. The issues of the current rogue access point detection technologies are those of false positives, scalability, and real-time. [9].

Bhusari suggests that in the scenario of financial fraud, Hidden Markov Models will be applicable. The suggested approach employs sequence behavior characteristics in order to detect credit card abnormalities. Also in the search of suspicious patterns in big data, probabilistic algorithms and in financial fraud, the technique is involved [10].

Khan et al. investigated the possibility of detection and prevention of wireless ad hoc network intrusions. The article addresses the issue of network attack detection and dynamism network issues. The authors examine intelligent detection systems that develop under the circumstances of cyberattacks and networks. [11].

Reddy et al. came up with applying SVM-based discriminant function as a technique of enhancing intrusion detection classification. Normal and dangerous network data are detected by the SVM classification as well. As the experiments demonstrate, the proposed method is more rapid in detecting intrusion than the traditional methods. [12].

Asaju et al. applied randomizable filtered classifiers and K-Nearest Neighbor in the case of intrusion detection. The suggested system would make use of several classification methods to maximize the detection and minimize classification errors. The ensemble based approach is the fusion of machine learning in order to enhance intrusion detection. [13].

Hounsou et al. used the Soft computing techniques which are Genetic Algorithms and Self-organizing Feature Map to detect network intrusions. The technique entails the use of network traffic so as to detect anomalies and attacks. Evolutionary detection can however be computationally intensive. [14].

Hybrid intrusion detection will be used by Jahromy et al., which is based on an SVM where features are optimized by genetic algorithm to classify attacks. The hybrid method can be applied to enhance the network intrusion detection with the help of correct parameters and classification accuracy [15].

III.. Methodology

The solution proposed uses machine learning hybrid in order to detect the unauthorized Wi-Fi access points of a behavior analysis of the wireless traffic information. The loading and preparation of the Wi-Fi dataset involve all the data cleaning process to remove any suspicious identification, missing data, and the standardization of data through the encoding and scaling of information. The most meaningful data characteristics are then considered and noise is removed by feature selection and it increases the precision and effectiveness of the model. The processed data is divided into a set of training and testing data based on a group-based train-test split. XGBoost, Random Forest and Linear Support Vector Machine are then trained on processed feature set to distinguish between normal and rogue access points. The predictions of such models need to have ensemble voting in order to combine them. The Flask web application uses Wi-Fi packet analysis in real time and displays it as either a safe or rogue access point after the model training and preprocessing workflow.

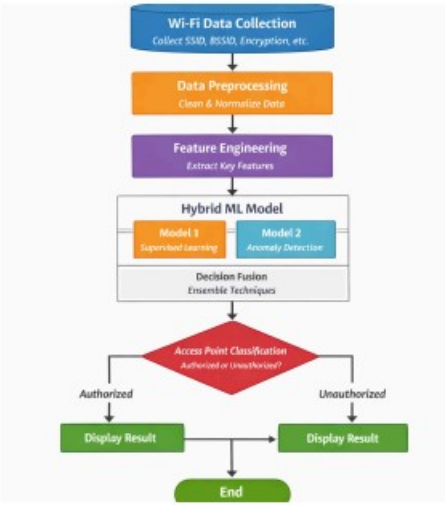

Fig.1. Proposed architecture

A. *Proposed Undertaking:*

The proposed approach uses machine learning in a hybrid mode within wireless networks to detect illegal access points of Wi-Fi networks. The system is used to check the strength of a signal, frame characteristics and encryption options by tracking Wi-Fi traffic. A leakage-free preparation process is used to purify the dataset, to remove identifier-based features and encode and scale them and to select the features to support an effective model training. This pre-training step improves generalization of machine learning models besides minimizing noise in the dataset.

The behavioral pattern of Wi-Fi traffic data is pre-processed and subsequently XGBoost, random forest and Linear Support Vehicle machine are trained to detect rogue access points. The result of the classification consists of the forecast of these models and their voting in large numbers. In network security application case, the method augments the detection as well as

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 4 | April-2026
ISSN (O) 3107-6696

minimizes false negatives. Flask-based web application includes the preprocessing workflow and training models to help give access point prediction and visualization in real time.

B. System Architecture:

Training and deployment are the two stages of the proposed system architecture. The information of the Wi-Fi traffic is imported and processed in the training stage so that the information is not leaked off. These include cleaning, encoding category attributes, scaling features and elimination of identifier-based features. The methods of feature selection are identified to single out the most significant features in the starting dataset and reduce the dimension count and maximize the performance of a model. After separation of the processed data into a training and a testing data, XGBoost, the Random Forest and Linear Support Vector Machine are then trained to determine the behavior of legal and illegal access points.

The deployed preprocessors and trained models are stored and integrated into a real-time prediction environment. The same is preprocessed on the data of incoming Wi-Fi traffic so that there is consistency of training. The trained machine learning models make decisions on whether the access point is a rogue or safe access point by the majority of the feature vectors being processed. Prediction outcome: A Flask-based web-based interface that would be used to monitor and detect unauthorized Wi-Fi hotspots in real time.

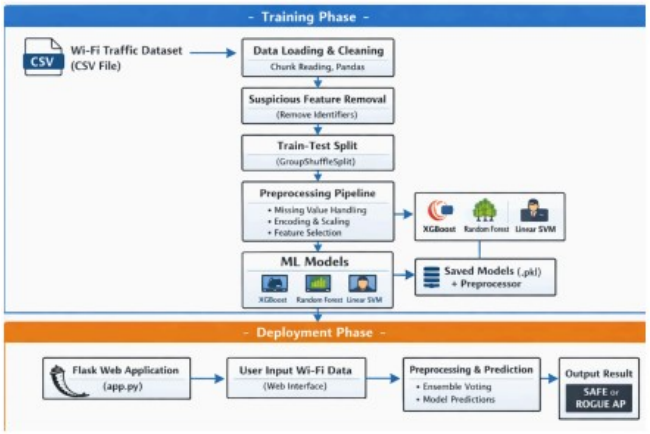

Fig.2. Proposed architecture

IV. Implementation

A. MODULES:

1. Data Collection Module

In the event of system testing and training, this module collects Wi-Fi traffic data on a network. To identify rogue access points, the AWID (Aegean Wi-Fi Intrusion Dataset) was used in this investigation, with wireless packet parameters of signal intensity, frame characteristics, encryption flags, and network IDs.

2. Data Preprocessing Module

This module is used to clean and prepare the data to be processed in machine learning. Removing identifiers based features, completing missing values, encoding categorical attributes, and scaling features, the dataset is brought in a state of accurate and reliable training of machine learning models.

3. Feature Selection Module

This module describes the most important dataset features with the statistical or information-based techniques. By selecting the most informative Wi-Fi traffic characteristics, model performance is improved, redundant attributes are removed, and the dimension of datasets is reduced.

4. Model Training Module

This module uses the processed dataset to train XGBoost, Random Forest and Linear Support Vector Machine. The models use patterns of training data to differentiate between authentic and fraudulent locations of Wi-Fi connection.

5. Ensemble Prediction Module

This lesson has used ensemble voting to combine the machine learning model predictions. The ensemble method decreases the false negative rates and increases the detection accuracy through a combination of the classifier strengths.

6. Real-Time Detection and Deployment Module

The last module involves giving the trained models and preprocessing pipeline with the help of Flask. The system predicts whether an access point is rogue or safe based on

statistics of Wi-Fi traffic, and displays the results in a dynamic web based interface.

B. ALGORITHMS:

1. XGBoost (Extreme Gradient Boosting)

XGBoost is the main classification model that can be used to classify illicit websites of the Wi-Fi access. Gradient boosting-based ensemble learning sequentially constructs multiple decision trees in order to reduce the error of prediction. The XGBoost will be trained to identify complicated patterns on the basis of the chosen Wi-Fi traffic data used to differentiate between fraudulent and legitimate access points. It identifies which features, including signal strength, frame characteristics and encryption flags, are most important by examining the significance of features.

2. Random Forest Algorithm

Random Forest in the ensemble machine learning technique trains a large number of decision trees and combines their results to enhance the accuracy of prediction. Training each tree on a random dataset and features reduces training overfitting and improves generalization of the model. To evaluate Wi-Fi traffic features and identify unwanted access points, this paper will use an extra model of detection, Random Forest.

3. Linear Support Vector Machine (SVM)

In the case of binary classification, SVM is a supervised learning algorithm. Linear SVM model identifies an optimal hyperplane to distinguish points by the classes. The data of Wi-Fi traffic feature vectors in this system is categorized into rogue and safe access point by applying linear SVM. It is able to examine the numerous characteristics of the AWID data due to its high-dimensional data processing.

4. Ensemble Voting Mechanism

Ensemble voting Classification with XGBoost, Random Forest, and Linear SVM. The decision is made based on majority or weighted voting following each of the models that have been predicted. A combination of multiple approaches enhances the credibility of detection and reduces the errors of classification.

V. Experimental Results

measure the suggested methodology with the AWID Wi-Fi infiltration dataset that contains various wireless traffic characteristics. Data consistency Preprocessing of the data, the features of identification, and features of scaled features were encoded and the scaled features were removed. Eighty percent of the processed dataset was composed of the training and testing sets. The training of the dataset was conducted with the help of XGBoost, random Forest and Linear Support Vector machine. Based on the variables of the traffic over Wi-Fi such as the signal strength, frame properties and encryption flags, the models were able to train the normal and abnormal access point traffic properties.

Data that were not observed was used in testing trained models and their results were used to determine their performance and generalization. We have experimented with the hybrid machine learning model and found that the hybrid model is very accurate in classification, precise and recalls. Most of the rogue access points are discovered with the use of the confusion matrix analysis by the wireless network security system with fewer false negative outcomes. Moreover, the network properties analysis has shown that network properties were responsible to detection of some properties. The final model was used to predict and indicate safe and malicious access points in real-time with Flask.

Accuracy: factor in the pros and cons of a test. Mathematics comes next.

$$Accuracy=\frac{(TN+TP)}{T}$$

Precision: Precision distinguishes between categorization accuracy or positive examples. The accuracy is determined by the use of the following:

$$\Pr e\,cision=\frac{TP}{(TP+FP)}$$

Recall: The ratio of corrects of the positive observations that a model predicts provides an insight into the power of the model in identifying each of the instances of the machine learning category.

$$Recall=\frac{TP}{(FN+TP)}$$

F1-Score: F1 was high, which means that machine learning model is correct. Precision and recall are supposed to be juxtaposed with the aim of improving the model accuracy. The accuracy is used to measure the predictive ability of a model using a collection of data.

$$F1=2\cdot\frac{(Recall\cdot \Pr e\,cision)}{(Recall+\Pr e\,cision)}$$

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 4 | April-2026
ISSN (O) 3107-6696

Fig 3 Training Confusion Matrix

According to the confusion matrix of the training stage, the model was very precise and less misclassified the regular and the rogue access points.

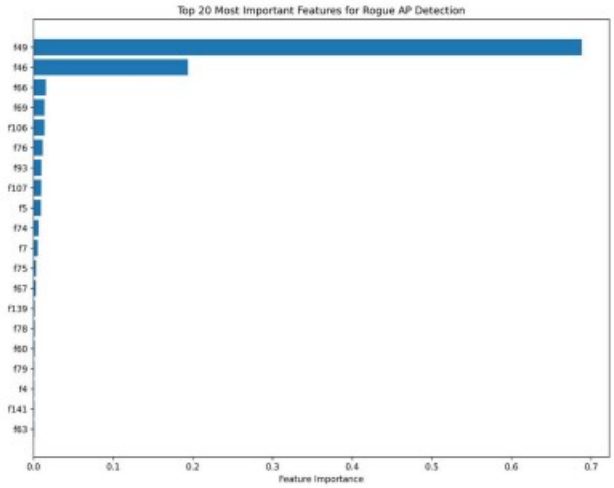

Fig 5 Feature Importance Analysis

The image shows the finest XGBoost properties in the identification of rogue Wi-Fi access points.

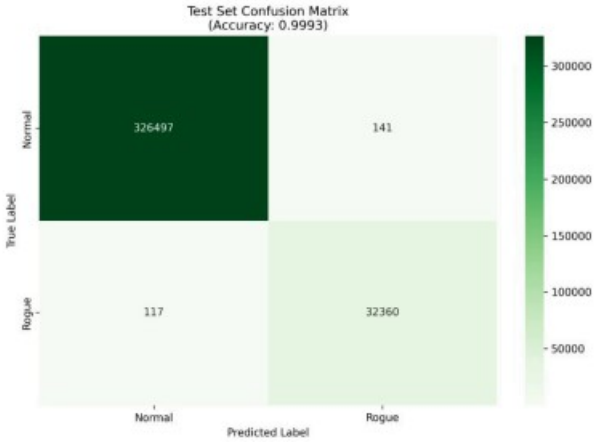

Fig 4: Test Confusion Matrix

In the identification of the rogue access points, confusion matrix testing dataset reveals that there is good model generalization with few false positives and false negatives.

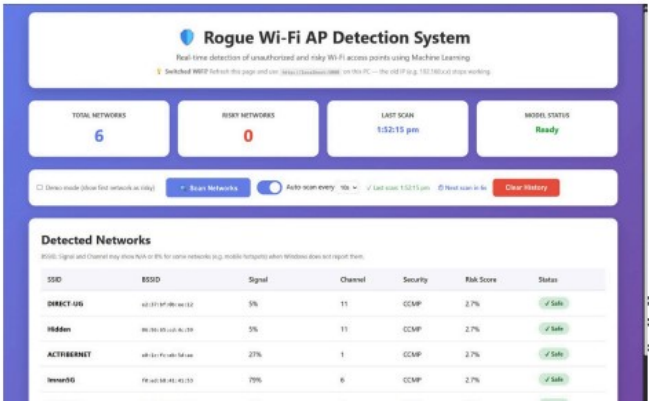

Fig 6 System Interface for Wi-Fi Network Detection

Here we may see the Flask based web application interface that shows the Wi-Fi networks that each access point sees, signal strength, type of security and the degree of risk.

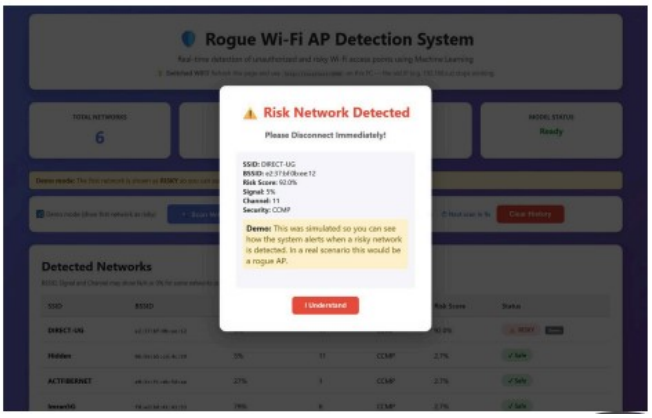

Fig 7 Rogue Network Detection Alert

The user is warned of the security threat in the case of a suspect or rogue Wi-Fi access point being discovered.

VI. Conclusion

In this study, hybrid machine learning system was devised to identify the unauthorized points of access of Wi-Fi network access. The suggested system implements leakage-free preprocessing pipeline, features selection and machine learning algorithms, such as XGBoost, Random Forest, and Linear Support Vector Machine, on Wi-Fi traffic patterns and detect rogue access points. The ensemble voting technique increases the predictive accuracy of the model by adding up the model forecasts.

According to experimental results of the AWID Wi-Fi incursion dataset, the suggested approach was found to be highly classified, precise, and recall. The system is able to detect rogue access points with minimal false negatives as is important in network security which is demonstrated by the confusion matrix. Such an approach, in its turn, suits the implementation of wireless networks because trained models, which are the component of Flask-based web application, can be used to track the undesired access points in real-time and identify them.

REFERENCES

[1] Doyeonkim, Dongil Shin, Dongkyoo Shin, "Unauthorized access point detection using machine learning algorithm for information protection, " IEEE International Conference on Big Data Science and Engineering.

[2] C. Wang, X. Zheng, Y. Chen and J. Yang, "Locating Rogue Access Point Using Fine-Grained Channel Information, " IEEE Transactions on Mobile Computing, vol. 16, no. 9, pp. 2560–2573, Sept. 2017.

[3] M. S. Gondal, A. J. Malik and F. A. Khan, "Network Intrusion Detection Using Diversity-Based Centroid Mechanism, " 2015 12th International Conference on Information Technology - New Generations, Las Vegas, NV, 2015.

[4] H. Han, B. Sheng, C. C. Tan, Q. Li and S. Lu, "A Timing-Based Scheme for Rogue AP Detection, " IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 11, pp. 1912–1925, Nov. 2011.

[5] S. Kitisriworapan, A. Jansang and A. Phonphoem, "Evil-Twin Detection on Client-side, " 2019 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Pattaya, Chonburi, Thailand, 2019.

[6] M. S. Gondal, A. J. Malik and F. A. Khan, "Network Intrusion Detection Using Diversity-Based Centroid Mechanism, " 2015 12th International Conference on Information Technology - New Generations, Las Vegas, NV, 2015.

[7] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection, " IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016.

[8] S. Mathur and A. Badone, "A methodological study and analysis of machine learning algorithms," International Journal of Advanced Technology and Engineering Exploration, vol. 6, pp. 45–49, 02 2019.

[9] B. Alotaibi and K. Elleithy, "Rogue access point detection: Taxonomy, challenges, and future directions," Wireless Personal Communications, vol. 90, pp. 5021–5028, 10 2016.

[10] V. Bhusari, "Application of hidden markov model in credit card fraud detection," International Journal of Distributed and Parallel systems, vol. 2, 11 2011.

[11] K. Khan, A. Mehmood, S. Khan, M. A. Khan, Z. Iqbal, and W. K. Mashwani, "A survey on intrusion detection and prevention in wireless ad-hoc networks," Journal of Systems Architecture, vol. 105, p. 101701, 2020. [Online].Available:http://www.sciencedirect.com/science/article/pii/S1383762119305089

[12] R. R. Reddy, Y. Ramadevi, and K. V. N. Sunitha, "Effective discriminant function for intrusion detection using svm," in 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2016, pp. 1148–1153.

[13] Asaju, L. Bolaji, P. B. Shola, N. Franklin, and H. M. Abiola, "Intrusion detection system on a computer network using an ensemble of randomizable filtered classifier, k-nearest neighbor algorithm," 2017.

[14] J. Hounsou, T. Nsabimana, and J. Degila, "Implementation of network intrusion detection system using soft computing algorithms (self-organizing feature map and genetic algorithm)," Journal of Information Security, vol. 10, pp. 1–24, 01 2019.

[15] B. Jahromy, A. Honarvar, M. Saif, and M. Jahromy, "A new method for detecting network intrusion by using a combination of genetic algorithm and support vector machine classifier," vol. 11, pp. 810–815, 01 2016.