Application of the Learning Set for the Detection of Jamming Attacks in 5G Mobile Networks

Brou Médard KOUASSI[1], Vincent MONSAN[2], Abou Bakary BALLO[3], Kacoutchy Jean AYIKPA[4], Diarra MAMADOU[5], Kablan Jérome ADOU[6]

LaMI, Université Félix Houphouët-Boigny, Abidjan, CÔTE D'IVOIRE[1, 3, 4, 5, 6]
LMA, Université Félix Houphouët-Boigny, Abidjan, CÔTE D'IVOIRE[2]
UREN, Université Virtuelle de Côte d'Ivoire, Abidjan, CÔTE D'IVOIRE[4]
ImVia, Université Bourgogne Franche-Comté, Dijon, FRANCE[4]

***Abstract*—Jamming attacks represent a significant problem in 5G mobile networks, requiring an effective detection mechanism to ensure network security. This study focused on finding effective methods for detecting these attacks using machine learning techniques. The effectiveness of Ensemble Learning and the XGBOOST-Ensemble Learning combination was evaluated by comparing their performance to other existing approaches. To carry out this study, the WSN-DS database, widely used in attack detection, was used. The results obtained show that the hybrid method, XGBOOST-Ensemble Learning, outperforms other approaches, including those described in the literature, with an accuracy ranging from 99.46% to 99.72%. This underlines the effectiveness of this method for accurately detecting jamming attacks in 5G networks. By using advanced machine learning techniques, the present study helps strengthen the security of 5G mobile networks by providing a reliable mechanism to detect and prevent jamming attacks. These encouraging results also open avenues for future research to further improve the accuracy and effectiveness of attack detection in radiocommunication in general and specifically in 5G networks, thereby ensuring better protection for next-generation wireless communications.**

***Keywords*—Jamming attacks; 5G mobile networks; ensemble learning; XGBOOST-ensemble learning; attack detection**

I. INTRODUCTION

Radiocommunication is defined, according to the International Telecommunication Union (ITU), as telecommunication carried out using radioelectric waves, that is to say, an electromagnetic wave that propagates in space without an artificial guide and whose frequency is by convention less than 3000 GHz [1]. The fields of application of this form of communication are numerous, including Wi-Fi, mobile, satellite, IoT networks, wireless sensors, high-altitude platforms, smart cities, smart grids, connected vehicles, etc. All these applications use propagation of the useful signal emitted in all directions, often in environments with multiple and sometimes complex obstacles, thus undergoing all kinds of disturbances, including intrusions by jamming for denial of service (DoS) and distributed denial of service (DDoS) [2]. Jamming attacks consist of intentionally transmitting a signal [3] that covers the frequencies used by the communication system to degrade the quality of the signal received by a communication device. Jamming signals can be relatively weak intentional electromagnetic interference (IEMI) [4] that degrades the performance of radio communication networks without damaging them. With the proliferation of connected objects (IoT) and the convergence of networks, the number of devices used in everyday life and connected to the Internet by radio communication has increased considerably in recent years. According to statistics, this number has increased from 1 million in 1992 to more than 50 billion in 2020[5] globally. With this reality, telecommunications networks, particularly 5G mobile networks, have undergone significant transformations to adapt to this exponential growth of connected devices. Significant contributions include the evolution of the radio access network (RAN), the part of 5G that connects end-user devices. Architectures such as C-RAN, O-RAN, vRAN, etc., have been proposed to address these challenges. These new network architectures have improved the flexibility, capacity, and efficiency of 5G mobile networks. However, with this evolution, new security challenges have also emerged. Among these challenges, jamming attacks represent a significant problem that compromises network security. Despite the advantages offered by new network architectures, the security of C-RAN networks has been questioned due to their vulnerability [6] to malicious jamming attacks, especially regarding the use of radio resources. Even the most advanced C-RANs can be subject to all sorts of attacks on radio networks.

Different types of jamming attacks can be used against C-RAN, including random, reactive, deceptive, and constant jamming [7]. These attacks seriously threaten the proper functioning of C-RAN networks and can compromise the quality of services offered to end users.

In order to counter these jamming attacks, it is essential to put in place effective detection mechanisms. This study focuses on using machine learning techniques to detect these attacks in C-RAN networks. The effectiveness of Ensemble Learning and the XGBOOST-Ensemble Learning combination is specifically assessed, comparing their performance to other existing approaches.

By identifying and evaluating the different jamming attacks possible in C-RAN networks, our study aims to strengthen the security of these networks by proposing advanced detection mechanisms. These results will allow a better understanding of the characteristics of these attacks and the development of appropriate countermeasures to protect C-RAN networks from the harmful consequences of jamming attacks.

The issue of jamming attack detection in 5G mobile networks using machine learning (ML) techniques is addressed in this study. The approach developed here, uses a specific database (WSN-DS) to evaluate and compare the performance of different machine Learning algorithms. The goal is to determine the most efficient algorithm for detecting jamming attacks based on this data set. The present study has made the following contributions:

- The use of machine learning techniques, in particular ensemble learning and the XGBOOST-ensemble learning combination, to detect jamming attacks. This innovative approach leverages the capabilities of these advanced techniques to improve detection accuracy.

- Performance evaluation of ensemble learning and XGBOOST-ensemble learning in comparison with other existing approaches. This comparative evaluation highlights the superior effectiveness of the hybrid XGBOOST-Ensemble Learning method, which outperforms the other approaches studied and those described in the literature.

This paper is organized as follows. After the introduction, which sets out the problem addressed in this paper in Section I, Section II presents a literature review of previous work on intrusion detection attacks and methods (ML-IDS) in radio communications, particularly in 5G. The ML-IDS methodology adopted in this work (EL-IDS) is formulated and presented in Section III. The results obtained are presented in Section IV and discussed in Section V. Finally, Section VI focuses on the research objective and draws conclusions from this study.

II. Related Work

The new generation of wireless communication networks, the fifth generation (5G), guarantees a high transmission rate and low latency and maintains good connectivity between heterogeneous mobile devices. 5G cellular networks provide the key infrastructure to deliver emerging services. Security anomaly detection is increasingly important in protecting systems from malicious attacks. Several authors have conducted interference studies in the 5G network. F Wu et al. studied a mixed digital interference (MNI) recognition approach based on convolutional neural networks (CNN) [8]. The results of this work showed that the accuracy could reach 97% or more for different signal-to-noise ratios and fading channels. M. Usama et al. proposed a technique stimulated by recent advances in deep learning to exploit the rich information hidden in large volumes of data and tackle resource allocation problems [9]. Mughaid et al. built a simulator for NOMA and applied a drop attack to extract a dataset from the simulation model. The accuracy of detecting drop attacks using data extracted after applying ML algorithms is 95.7% for LR. Furthermore, their methodology for detecting wireless cyberattacks in 5G networks is based on applying ML and DL techniques such as Decision Trees, KNN, Multi-class Decision Jungle, Multi-class Decision Forest, and Multi-class Neural Networks. The proposed work is implemented and tested using a complete set of reference data on Wi-Fi networks [10]. The experiments yielded 99% accuracy for the KNN algorithm and 93% for DF and the neural network. L. Xiao et al. investigated MEC systems' attack patterns, focusing on mobile offloading and caching procedures. In this article, they propose security solutions that apply Reinforcement Learning (RL) techniques to provide secure offload to edge nodes against jamming attacks; also, lightweight authentication and secure collaborative caching schemes have been designed to protect data confidentiality[11]. The results of these reinforcement learning-based methods for mobile edge caching are relevant. Y. Wang et al. presented an anonymous jamming detection model for 5G and beyond based on critical signal parameters collected from the radio access network and core network protocol stacks on a test bench. 5G trial. The results of their approach give supervised instantaneous detection models an area under the curve (AUC) between 0.964 and 1 compared to time-based long-term memory models (LSTM), which reach an AUC between 0.923 and 1 [12]. Jamming and intrusion detection remain 5G's most important research areas of maintaining the trustworthiness of use cases and preventing user experience degradation by avoiding a severe infrastructure failure or a denial of service in critical applications within the company. Similarly, Marouane Hachimi et al. proposed machine learning-based intrusion detection in the 5G C-RAN network to enhance security [5]. Their approach was to classify the types of jamming attacks within a 5G network. Their experiment gave an attack classification accuracy of 94.51% with a false negative rate of 7.84%.

The work presented above indicates that the studies carried out by these authors have focused on interference recognition in Wi-Fi networks using Machine Learning and Deep Learning models.

Furthermore, these studies highlight the use of Machine Learning and Deep Learning algorithms for interference classification, but have not delved into comparative studies that evaluate the performance of different interference classification models in Wi-Fi networks. The present study uses Machine Learning techniques, in particular Ensemble Learning (Random Forest, KNN, Naïve Bayes, Logistic Regression) and the XGBOOST-Ensemble Learning combination (XGBOOST-Random Forest, XGBOOST-KNN XGBOOST-Naïve Bayes, XGBOOST-Logistic Regression) to detect interference attacks in the 5G network. It compares the performance of different interference identification techniques, to highlight their impact on the accuracy of interference classification. The use of these different approaches provides a better understanding of Ensemble-Learning classification methods and the XGBOOST-Ensemble Learning combination, highlighting the strengths and weaknesses of each technique in detecting interference in the 5G network.

III. Material and Method

A. Material

The database used for our study is the WSN-DS: a data set for intrusion detection systems in wireless sensor networks. It contains 374,661 simple connection vectors, each including 23 characteristics, and is labeled as normal or attack. The specific attack types are scattered into different attack categories, namely constant jamming, random jamming, deceptive jamming, and reactive jamming, in addition to the normal case (without attack).

The experiments used Python programming on a DELL desktop computer with an Intel(R) Core i7-10700 CPU clocked at 2.90 GHz, 32 GB of RAM, and a card NVIDIA Quadro P400 graphics.

B. Deployment Architecture

Fig. 1 presents the architecture. It divides base stations into Radio Remote Heads (RRH) and the Baseband Unit (BBU). RRH is the unit that provides the interface to the fiber and performs the digital processing, digital-to-analog conversion.

The traditional C-RAN architecture is based on Mobile Cloud Computing (MCC) principles with centralized BBUs in remote data centers. Resources provided to mobile users are typically located at the end of a long chain of nodes and across a mobile backhaul that can be congested at any time. However, more and more applications today operate almost in real time with requirements for very short transmission times. Thus, the performance of C-RAN systems is highly dependent on the physical proximity between mobile users and cloud servers.

The objective of this architecture is to concede the calculation and the storage to the H-RRHs near the mobile user to increase the processing capacity of the mobile terminals and allow the unloading of the greedy tasks in resources. All the added resources from the Cloud-RRH.

In addition, using cloud containers instead of virtual machines (VMs) at the RRH cloud level saves performance and processing time. A container is a collection of self-contained components ready to be deployed, and it can include libraries to be able to run the applications. Unlike VMs, multiple containers can share the same host operating system with its libraries and binaries. The containers are much lighter, translating into faster launch and easier migration from one machine to another.

Despite these proposals for attractive solutions introduced in this new C-RAN architecture, the radio interface remains a significant challenge in the face of jamming attacks using radio resources. For experimental results, a specialized dataset for Wireless Sensor Networks (WSN) was analyzed to classify jamming attacks; WSN-DS can have normal or malicious network traffic.

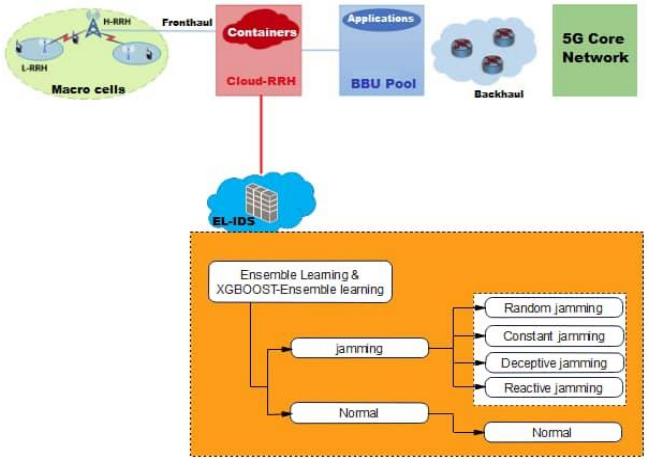

Fig. 1. Deployment architecture of EL-IDS in the CRAN – CRRH environment).

C. Learning Algorithms

The choice of the appropriate Learning algorithm [13] is crucial for the performance of a prediction model. In the present study, we opted for the use of Learning ensembles composed of the following algorithms: Random Forest, KNN, Naïve Bayes, Logistic Regression and XGBOOST. Each classification algorithm has its advantages and disadvantages, and we decided to combine them with XGBOOST for feature extraction to improve the performance of our classification model. When the data is insufficient, the learning set can use bootstrapping to train various classifiers using the different data samples, and also if the data is too large to train a single classifier then it is possible to partition the data into subsets for training purposes.

1) KNN (K Nearest Neighbor): The nearest neighbor (KNN) method [14] is a popular classification method in data mining and statistics due to its simplicity of implementation and significant performance in classification. However, traditional KNN methods cannot assign a fixed k value (even if set by experts) to all tested samples. Previous solutions assign different k values to different test samples by cross-validation, but they are usually very time-consuming [15]. The KNN method has been widely used in data mining and machine learning applications due to its simplicity of implementation and remarkable performance.

2) Naive bayes: Naïve Bayes is one of the most popular data mining algorithms. Its effectiveness relies on the attribute independence assumption, although this may be violated in many real-world datasets. Many efforts have been made to mitigate this assumption, among which feature selection is a critical approach [16].

The naive Bayes classifier has surprised machine learning researchers by performing well on various learning problems. The researchers sought to overcome the main weakness of naive Bayes attribute independence and improve the algorithm's performance [17]. The naive Bayes classifier simplifies learning by assuming that features are class-independent. Although independence is generally a bad assumption, naive Bayes often compete with more sophisticated classifiers in practice.

3) Logistic regression: Logistic regression is used to obtain the odds ratio in the presence of more than one explanatory variable. The procedure is similar to multiple linear regression, except that the response variable is binomial. The result is the impact of each variable on the odds ratio of the observed event. The main advantage is avoiding confounding effects by analyzing the association of all variables [18]. It is an algorithm based on a statistical model allowing the study of the relations between a set of qualitative variables, Xi, and a qualitative variable Y. It uses a generalized linear model on a logistic function as a link function. The probability of predicting an event with the logistic regression model is established or not from the optimization of the regression coefficients, and its result continuously varies between 0 and 1.

4) Random forest: Random forest is a supervised learning algorithm used for classification and regression. It combines multiple decision trees to produce more accurate predictions. Random forest is useful for datasets with categorical or continuous variables and can handle missing data. The RF algorithm randomly divides the data set into training data (in-bag) for training and validation data (out-of-bag) for testing. The level of learning and 2/3 of the data set is devoted to training data and 1/3 to validation data. Subsequently, many decision trees are randomly created using "bootstrap samples" from the dataset. The branching of each tree is determined by randomly selected predictors at node points [19].

5) XGBOOST: XGBOOST is an improved model of the Gradient Boost algorithm. This machine Learning algorithm solves common business problems while relying on a minimum amount of resources [20]. Extreme gradient boosting is a method that is used to reduce the number of errors in predictive data analysis. XGBOOST is an assembly of decision trees (weak learners) that predict residuals and correct errors of previous decision trees. The particularity of this algorithm lies in the decision tree used. It is a recently introduced machine learning algorithm, which has proven to be very powerful in modeling complex processes in other research areas.

D. Methods

The methodology used in this study is based on several well-defined steps, thus providing a solid and rigorous approach to achieving our objectives. The most advanced intrusion detection techniques are studied to enable the security system monitoring the network to analyze traffic in order to discover actions that disrupt network confidentiality, integrity and availability.

Here is a detailed description of these steps:

- Step 1 : Data preprocessing

In this first stage, data were pre-processed to prepare them for subsequent analysis. Two distinct groups of data are created: the independent variables and the dependent variable. The "Type of attack" column is designated as the dependent variable in our dataset. The transformation of this categorical variable into a numerical value to facilitate analysis is carried out.

- Step 2 : Feature extraction

In this step, the XGBOOST algorithm is used to extract the characteristics of the independent variables. This advanced method enables us to highlight patterns and significant information in the data. Next, our data are divided into three parts: training, validation and testing. This division enables us to measure the effectiveness of the model on separate data sets and ensure its generalizability.

- Step 3 : Training with cross-validation

In this crucial step, the model was trained using cross-validation with algorithms. This approach makes it possible to test different algorithms and select the best performing one for classifying instruction types, the specific task. The training and validation sets are used to adjust the model parameters and evaluate its performance.

- Step 4 : Testing and evaluation

The test phase is essential for evaluating the quality of our model and detecting attacks. The test dataset used is independent of the training and validation datasets, to assess the model's actual performance. The results obtained are carefully examined and compared with known attacks to measure the model's effectiveness in detecting attacks.

Following this well-structured methodology, an in-depth study is carried out on attack detection, pre-processing the data, extracting relevant features, training the model by cross-validation and rigorously evaluating its performance.

Fig. 2 below summarizes the methodology adopted:

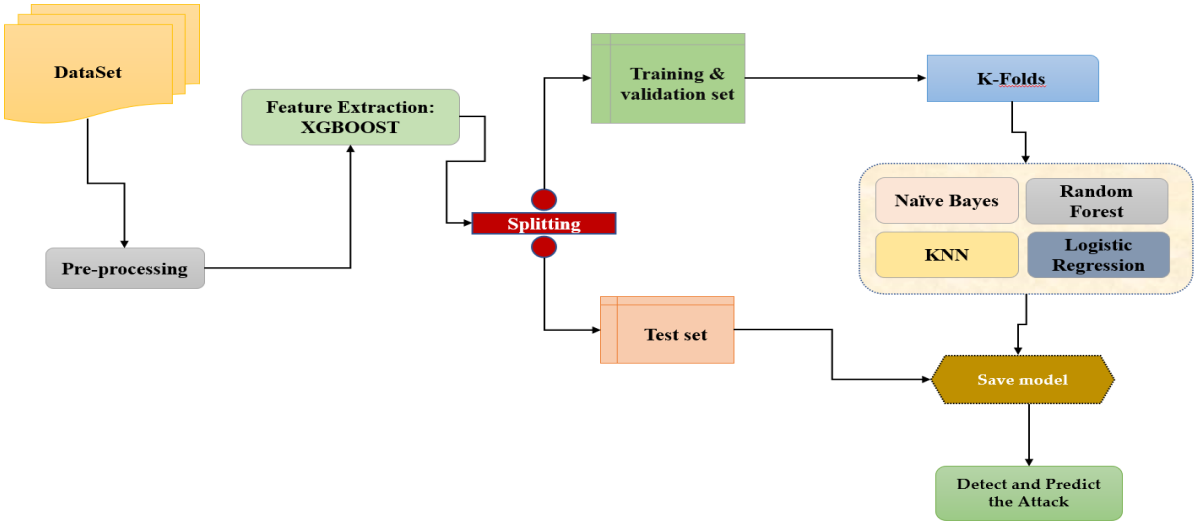

Fig. 2. Illustration of our methodology.

E. Evalutation Metrics

To evaluate the results of this study, several measures were used. Efficiency (MCC). The differential equations are: accuracy, precision, recall, F1 score and Matthew's correlation coefficient. The differential equations are as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precison = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Yi - \hat{Y}i)^2 \quad (6)$$

The Receiver Operating Characteristic (ROC) curve is a graph that shows the performance of a binary classification model. It describes the rate of true positives (sensitivity) as a function of the rate of false positives at different classification thresholds. An ideal ROC curve is approximately in the upper left corner of the graph, denoting high sensitivity and specificity.

The confusion matrix is a table that summarizes the results of the predictions of a classification model. It evaluates the model's predictions with the actual values from the dataset and classifies them into four categories: true positives, true negatives, false positives, and false negatives. The confusion matrix is used to compare a model's precision, recall, specificity, and overall accuracy.

IV. Results

The results are structured into two main parts: ensemble learning and XGBOOST Ensemble Learning.

The results of the machine learning models and XGBOOST-Ensemble learning have been presented separately. The machine learning models of the XGBOOST-Ensemble Learning combination outperform the machine learning models of ensemble learning, with a maximum accuracy of 99.72%.

V. Discussion

A. Case of Ensemble Learning

The Table I presents the metrics results:

The Random Forest model presents exceptional performances on all the criteria evaluated. It achieves high precision, a high F1 score, and high recall, all at 99.68%. The very low MSE of 0.007 indicates that the model predictions are close to the actual values. Moreover, the MCC of 98.18% indicates a robust correlation between the predictions and the actual observations. The execution time is reasonable at 356.15 ms.

The KNN model also performs well, although slightly lower than the Random Forest. Measurements of precision, F1 score, and recall are around 98%. The MSE of 0.063 indicates a slight average error of the predictions compared to the actual values. The execution time is longer at 1730.48 ms, which can be a drawback if efficiency is an important criterion.

The Naïve Bayes model has lower performance than the two previous models. Although precision and F1 score are reasonable at 88.10%, recall is relatively low at 84.83%. The high MSE of 0.358 indicates a more significant error of predictions against actual values compared to previous models. The MCC of 48.35% suggests a moderate correlation between predictions and actual observations. However, the execution time is very fast at only 2.62 ms.

The logistic regression model performs lower than other models. Precision and F1 scores sit at 86.24%, while recall is slightly higher at 88.21%. The high MSE of 0.511 indicates a significant prediction error compared to the actual values. The low MCC of 14.76% suggests a weak correlation between predictions and actual observations. Execution time is moderately fast at 185.65 ms.

The Random Forest model is the best among the four evaluated models regarding overall performance, with outstanding results on all measures. The KNN also shows good performance, although lower. Naïve Bayes models and logistic regression show relatively weaker performance, with more significant errors and less strong correlation between predictions and actual observations. Fig. 3 presents the histogram representing the performance of the models.

The Fig. 4 represents the ROC curve and the confusion matrix of the best model, namely the Random Forest.

TABLE I. Case of Ensemble Learning Metrics Results

Models	Accuracy (%)	Time(ms)	Precision (%)	F1 score (%)	MSE	Recall (%)	MCC (%)
Random Forest	**99.68**	**356.15**	**99.68**	**99.68**	**0.007**	**99.68**	**98.18**
KNN	98.23	1730.48	98.21	98.21	0.063	98.23	89.72
Naïve Bayes	84.83	2.62	88.10	88.10	0.358	84.83	48.35
Logistic Regression	88.21	185.65	86.24	86.24	0.511	88.21	14.76

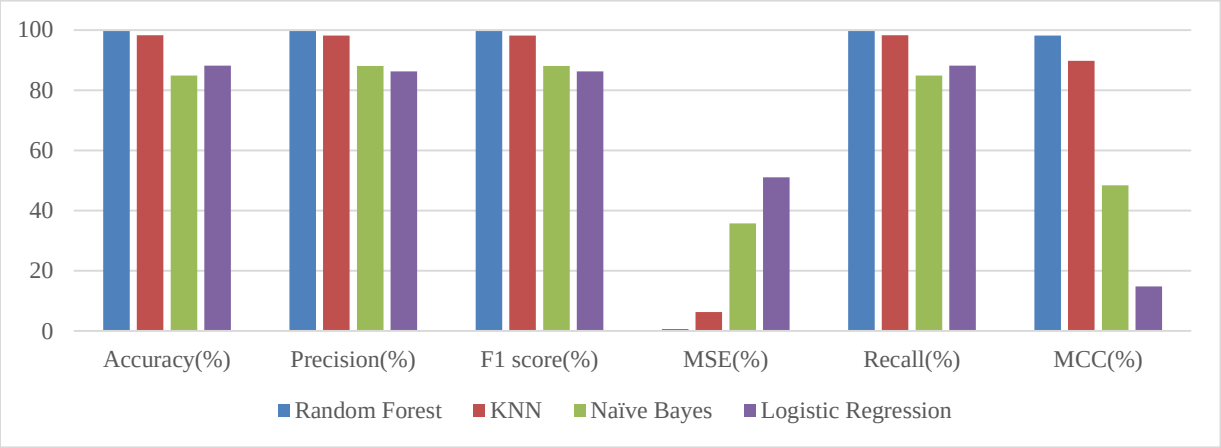

Fig. 3. Model performance histogram of case of ensemble learning.

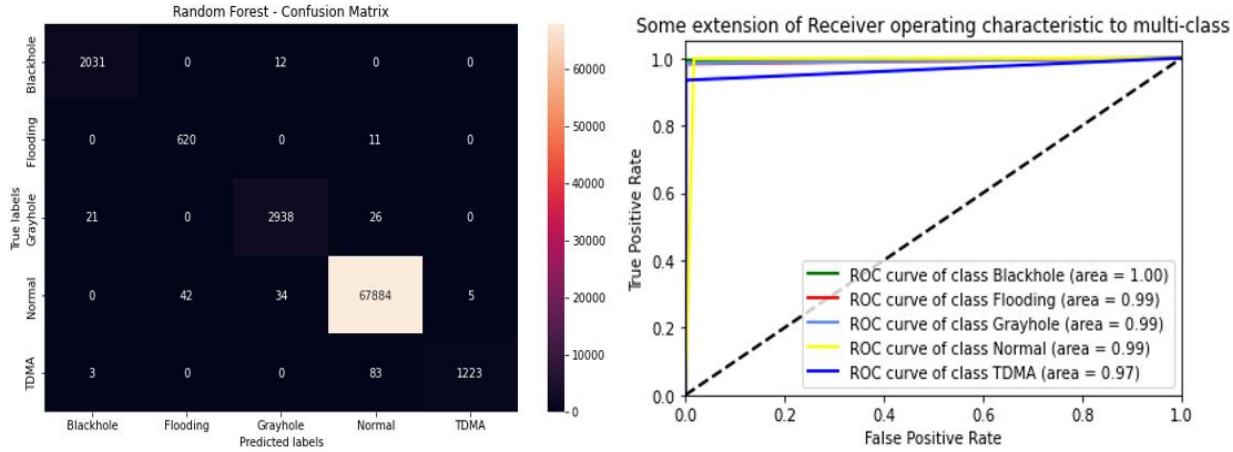

Fig. 4. ROC curve and confusion matrix of Random Forest.

B. *Case of XGBOOST- Ensemble Learning*

All of the models presented in Table II have high levels of accuracy, with scores ranging between 99.46% and 99.72%. This performance demonstrates their ability to classify the vast majority of samples accurately.

The execution time varies depending on the classification methods used. The XGBOOST-Naïve Bayes model is the fastest, with an execution time of only 1.369 milliseconds, while the XGBOOST-Random Forest model is the slowest, requiring 606.06 milliseconds. The other two models fall between these extremes regarding execution time. It is essential to consider these differences based on your specific application needs.

All models have an accuracy ranging from 99.46% to 99.72%, demonstrating their ability to classify most positive samples accurately. These models are, therefore, effective in avoiding false positives.

The F1 score, which combines precision and recall, presents high values for all models, ranging between 99.46% and 99.72%. This indicates a good balance between accuracy and the ability to recall positive samples.

The low MSE (Mean Squared Error) values obtained here indicate a low error in the predictions made. However, it should be noted that their interpretation may be limited in the context of classification.

Recall measures the ability of models to detect true positives among all truly positive samples. All models exhibit high recall values ranging from 99.46% to 99.72%, demonstrating their ability to identify positive samples.

The Matthews Correlation Coefficient (MCC) is a measure that considers the four categories of classification results. All models obtain high values of MCC, ranging from 96.98% to 98.41%, indicating a strong correlation between predictions and actual observations.

In conclusion, the models' performances based on XGBOOST and the other algorithms are globally compelling. The histogram represents the performance of these models visually.

Fig. 5 presents the histogram representing the performance of the models.

The ROC curve and the confusion matrix of the best model, namely the XGBOOST-Logistics Regression, are represented by the Fig. 6.

TABLE II. CASE OF XGBOOST-ENSEMBLE LEARNING METRICS RESULTS

Models	Accuracy (%)	Time(ms)	Precision (%)	F1 score (%)	MSE	Recall (%)	MCC (%)
XGBOOST-Random Forest	99.69	606.06	99.69	99.69	0.0066	99.69	98.25
XGBOOST-KNN	99.69	25.708	99.69	99.69	0.0064	99.69	98.26
XGBOOST-Naïve Bayes	99.46	1.369	99.47	99.47	0.0092	99.46	96.98
XGBOOST-Logistic Regression	**99.72**	**83.806**	**99.72**	**99.72**	**0.006**	**99.72**	**98.41**

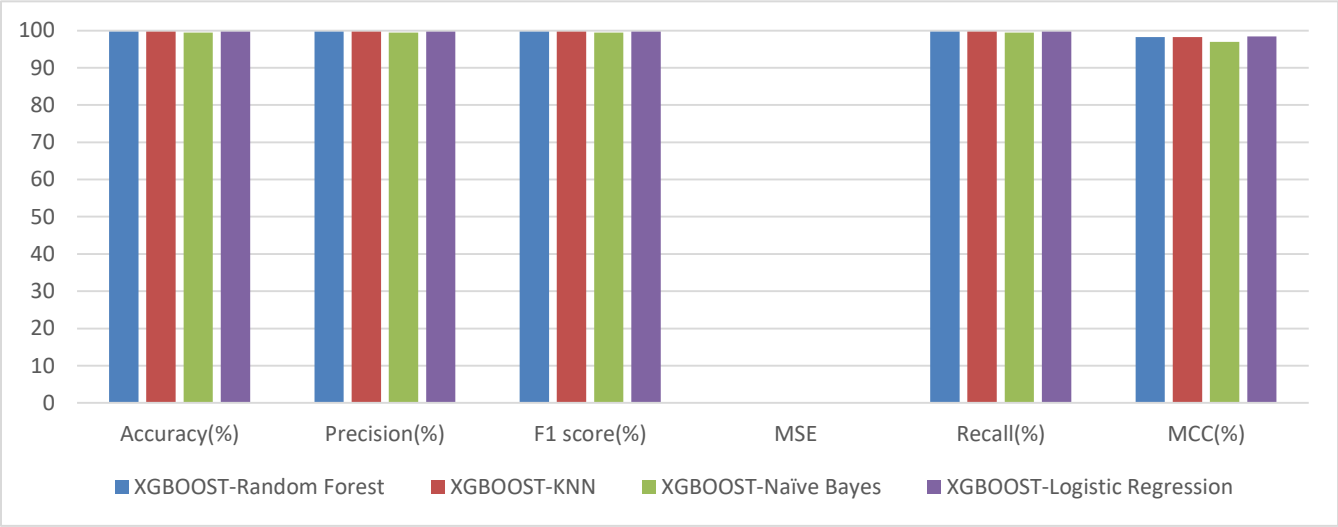

Fig. 5. Model performance histogram of case of XGBOOST- ensemble learning.

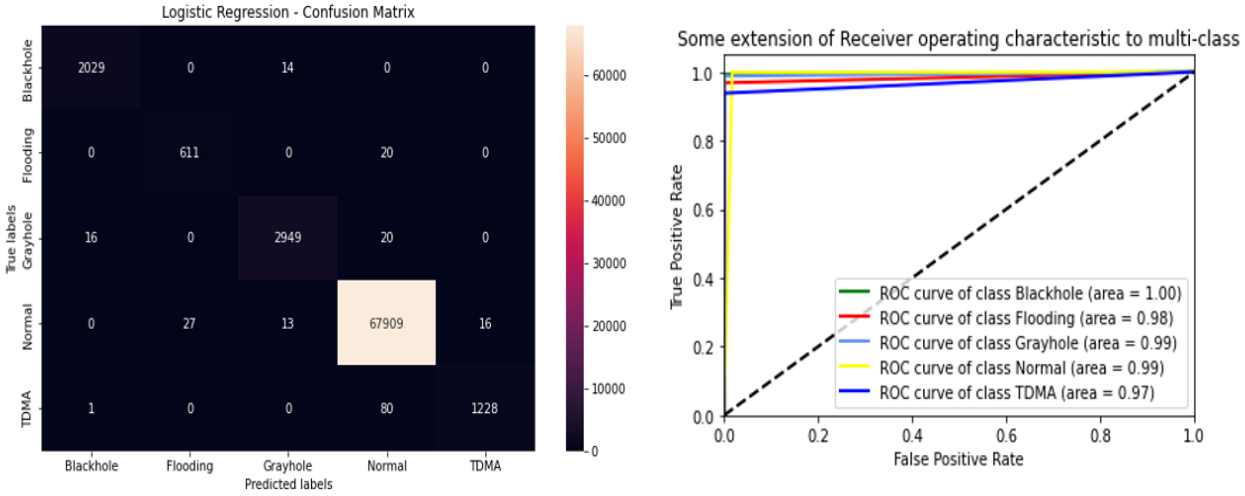

Fig. 6. ROC curve and confusion matrix of XGBOOST- ensemble learning.

C. Comparison with Existing Methods

The results of our experiments exceeded those of the state of the art. Table III shows the results. These results are also represented by the histogram, as shown in Fig. 7.

Fig. 7 shows the histogram comparing the results with those of the state of the art. A comparison was made between the proposed system and the methods used by other researchers in the field of intrusion detection in wireless sensor networks. The results showed that our system outperforms other authors' methods in the two approaches we performed.

TABLE III. COMPARISON WITH EXISTING RESULTS

Method	Accuracy (%)
Marouane Hachimi et al[16]	94,51%
Singh, N et al [17]	98,29%
Shaimaa Ahmed et al[18]	97,9%
Our method of scenario1	**99,68%**
Our method of scenario2	**99,72%**

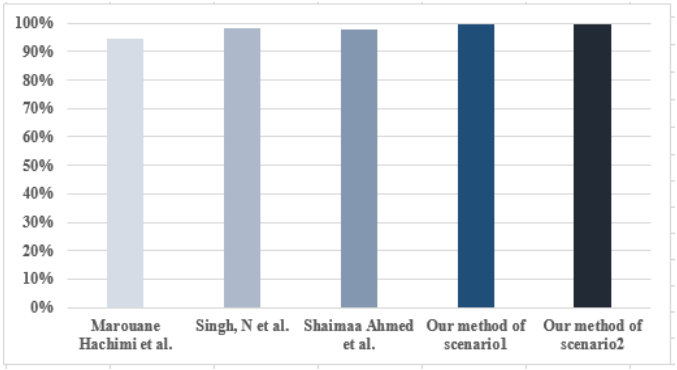

Fig. 7. Histogram of best scores of experiments.

In the first case, using the ensemble learning approach with the Random Forest algorithm, our system achieved an accuracy rate of 99.68%. This means that our method detected intrusions with high accuracy, surpassing the results obtained by other researchers. In the second case, using the XGBOOST-Logistic Regression approach as part of the Learning Ensemble, our system achieved an even higher accuracy rate of 99.72%. This remarkable performance highlights the effectiveness of our method of detecting intrusions in wireless sensor networks accurately.

These results demonstrate the superiority of our system compared to existing methods in terms of intrusion detection accuracy. The approach developed in this paper, based on advanced machine learning techniques, offers remarkable performance, strengthening the security of wireless sensor networks and guaranteeing more effective protection against intrusions.

Importantly, these results also demonstrate the importance of continued research in this area, as they pave the way for future improvements and new approaches for even more accurate detection of intrusions in wireless sensor networks.

VI. Conclusion

In conclusion, the present study has demonstrated that the use of machine learning techniques, in particular ensemble learning and the XGBOOST-Ensemble Learning combination, is promising for the detection of attacks in 5G networks. The results show that the hybrid method, XGBOOST-Ensemble Learning, outperforms all other approaches, including those described in the literature, with an accuracy between 99.46% and 99.72%. These results confirm the effectiveness of ensemble learning in detecting attacks in 5G networks. This study represents a significant advance in the detection of attacks in 5G networks using machine learning techniques. The promising results pave the way for further research and continuous improvements in 5G network security, helping to ensure the reliability and protection of next-generation wireless communications. Future works will explore other attack detection methods based on statistical analysis approaches, such as operational approach models. This will enable us to improve detection accuracy and develop more robust defense systems against attacks in 5G networks. Another avenue would be to integrate real-time detection techniques to enable a rapid and proactive response to potential attacks, thereby strengthening the security of 5G networks.

References

[1] « V.573 : Vocabulaire des radiocommunications ». https://www.itu.int/rec/R-REC-V.573-3-199006-S/fr (consulté le 29 mai 2023).

[2] J. Villain, V. Deniau, A. Fleury, E. P. Simon, C. Gransart, et R. Kousri, « EM Monitoring and Classification of IEMI and Protocol-Based Attacks on IEEE 802.11n Communication Networks », IEEE Trans. Electromagn. Compat., vol. 61, no 6, p. 1771-1781, déc. 2019, doi: 10.1109/TEMC.2019.2900262.

[3] « Détection de cyber attaques sur réseau Wi-Fi par classification de données spectrales - Archive ouverte HAL ». https://hal.science/hal-02315599v2 (consulté le 29 mai 2023).

[4] B. S. Chaudhari, M. Zennaro, et S. Borkar, « LPWAN Technologies: Emerging Application Characteristics, Requirements, and Design Considerations », Future Internet, vol. 12, no 3, p. 46, mars 2020, doi: 10.3390/fi12030046.

[5] M. Hachimi, G. Kaddoum, G. Gagnon, et P. Illy, « Multi-stage Jamming Attacks Detection using Deep Learning Combined with Kernelized Support Vector Machine in 5G Cloud Radio Access Networks », in 2020 International Symposium on Networks, Computers and Communications (ISNCC), Montreal, QC, Canada: IEEE, oct. 2020, p. 1-5. doi: 10.1109/ISNCC49221.2020.9297290.

[6] M. A. Ridwan, N. A. M. Radzi, K. H. M. Azmi, F. Abdullah, et W. S. H. M. W. Ahmad, « A New Machine Learning-based Hybrid Intrusion Detection System and Intelligent Routing Algorithm for MPLS Network », IJACSA, vol. 14, no 4, 2023, doi: 10.14569/IJACSA.2023.0140412.

[7] A. Cortés-Leal, C. Del-Valle-Soto, C. Cardenas, L. J. Valdivia, et J. A. Del Puerto-Flores, « Performance Metric Analysis for a Jamming Detection Mechanism under Collaborative and Cooperative Schemes in Industrial Wireless Sensor Networks », Sensors (Basel), vol. 22, no 1, p. 178, déc. 2021, doi: 10.3390/s22010178.

[8] F. Wu et al., « Mixed Numerology Interference Recognition Approach for 5G NR », IEEE Wireless Commun. Lett., vol. 10, no 10, p. 2135-2139, oct. 2021, doi: 10.1109/LWC.2021.3094928.

[9] M. Usama, I. Ilahi, J. Qadir, R. N. Mitra, et M. K. Marina, « Examining Machine Learning for 5G and Beyond Through an Adversarial Lens », IEEE Internet Comput., vol. 25, no 2, p. 26-34, mars 2021, doi: 10.1109/MIC.2021.3049190.

[10] A. Mughaid et al., « Improved dropping attacks detecting system in 5g networks using machine learning and deep learning approaches », Multimed Tools Appl, vol. 82, no 9, p. 13973-13995, avr. 2023, doi: 10.1007/s11042-022-13914-9.

[11] L. Xiao, X. Wan, C. Dai, X. Du, X. Chen, et M. Guizani, « Security in Mobile Edge Caching with Reinforcement Learning », IEEE Wireless Commun., vol. 25, no 3, p. 116-122, juin 2018, doi: 10.1109/MWC.2018.1700291.

[12] Y. Wang, S. Jere, S. Banerjee, L. Liu, S. Shetty, et S. Dayekh, « Anonymous Jamming Detection in 5G with Bayesian Network Model Based Inference Analysis », in 2022 IEEE 23rd International Conference on High Performance Switching and Routing (HPSR), Taicang, Jiangsu, China: IEEE, juin 2022, p. 151-156. doi: 10.1109/HPSR54439.2022.9831286.

[13] S. Zhang, X. Li, M. Zong, X. Zhu, et R. Wang, « Efficient kNN Classification With Different Numbers of Nearest Neighbors », IEEE Trans. Neural Netw. Learning Syst., vol. 29, no 5, p. 1774-1785, mai 2018, doi: 10.1109/TNNLS.2017.2673241.

[14] K. J. Ayikpa, K. J. Ayikpa, K. J. Ayikpa, D. Mamadou, P. Gouton, et K. J. Adou, « Experimental Evaluation of Coffee Leaf Disease Classification and Recognition Based on Machine Learning and Deep Learning Algorithms », Journal of Computer Science, vol. 18, no 12, p. 1201-1212, déc. 2022, doi: 10.3844/jcssp.2022.1201.1212.

[15] E. Frank, M. Hall, et B. Pfahringer, « Locally Weighted Naive Bayes », 2012, doi: 10.48550/ARXIV.1212.2487.

[16] « [PDF] An empirical study of the naive Bayes classifier | Semantic Scholar ». https://www.semanticscholar.org/paper/An-empirical-study-

of-the-naive-Bayes-classifier-Watson/2825733f97124013e8841b3f8a0f5bd4ee4af88a (consulté le 29 mai 2023).

[17] S. Sperandei, « Understanding logistic regression analysis », Biochem Med, p. 12-18, 2014, doi: 10.11613/BM.2014.003.

[18] « What is Random Forest? | IBM ». https://www.ibm.com/topics/random-forest (consulté le 29 mai 2023).

[19] H. Mo, H. Sun, J. Liu, et S. Wei, « Developing window behavior models for residential buildings using XGBoost algorithm », Energy and Buildings, vol. 205, p. 109564, déc. 2019, doi: 10.1016/j.enbuild.2019.109564.

[20] B. Pan, « Application of XGBoost algorithm in hourly PM2.5 concentration prediction », IOP Conf. Ser.: Earth Environ. Sci., vol. 113, p. 012127, févr. 2018, doi: 10.1088/1755-1315/113/1/012127.