Simple Yet Powerful Machine Learning-Based IoT Intrusion System with Smart Preprocessing and Feature Generation Rivals Deep Learning

P. Sujatha[1], Bestha Umesh Chandra[2], Kusumala Divakar[3], Yeranakula Veerendra[4], Bannela Kuruva Mallesh[5]

[1,2,3,4,5]*Dept. of Computer Science and Engineering (Data Science), St. Johns College of Engineering and Technology, Yemmiganur, 518360, India*

Abstract: **Recent hype on Internet of Things (IoT) has led to the increased susceptibility of cyber-attacks and that, more than ever, has led to the timeliness of the need to have effective intrusion detection systems (IDS). To construct an IoT intrusion detection network, the paper will apply machine learning on the data of UNSW-NB15 dataset to preclassify the attacks into eight categories, including: DoS, Exploits, Fuzzers, Generic, Normal, Reconnaissance, Shellcode and Worms. Decision Tree (DT), Random Forest (RF), as well as XGBoost, LightGBM, AdaBoost, Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) are some of the machine learning models which have been trained and evaluated based on the classification accuracy. Optimization of the hyper parameter was carried out using both models and the outcome of the optimization is the random forest and the rate of 98.20 percent and the LightGBM whose rate is 98.13. The correctness of Decision Tree model and the XGBoost were 97.22 and 94.67, respectively. AdaBoost and ANN had a precision of 71.93 and 86.02 respectively. The above models were regarded as the most precise and the highest score of Relative Recall and F1-score with the highest score of most forms of attacks presumed to be the Random Forest and LightGBM models. It has been established that machine learning algorithms such as the random forest and lightgbm can be implemented to predict the network-based attacks on the network of the IoT devices and this can become a solution to the security of the IoT networks.**

Index Terms- Intrusion Detection, Machine learning, UNSW-NB15 Dataset, Attack Classification, DoS, Exploits, Fuzzers, Generic, Normal, Reconnaissance, Shellcode, Worms, Decision Tree, Random Forest, XGBoost, LightGBM, AdaBoost, Artificial neural network.

I. INTRODUCTION

The new possibilities of the Internet of Things (IoT) is an opportunity since it is more automation, and a threat since it is a high-risk vulnerability that devices that facilitate the establishment of an IoT are not offered with the safety. Among the cyber threats, there are Denial of Service (DoS), Exploits, Shellcode and Worms that are most likely to be weak in the devices. Even threats are difficult to detect, i.e., the conventional Intrusion Detection System (IDS) and the majority of the system based on machine learning can imply the extrapolation issue on how new types of attack can look. Consequently, it is a pressing requirement of a powerful scaled IDS that would allow it to help detect the attacks to an IoT system as they happen.

The given paper introduces an IoT intrusion detection system that presupposes the application of machine learning to track the presence of attack in the system based on the set of labeled data on UNSW-NB15 that includes eight attacks. The machine learning models mentioned in the study are Decision Tree (DT), random forest (RF), XGBoost, light GBM, Artificial neural networks (ANN) and convolutional neural networks (CNN) with the specification of the hyperparameter that should be made to maximize the performance.

Among them, one can single out the most crucial one that is the design of a more sophisticated machine learning model that can be used to detect the intrusion to the IoT, the overall comparison of the various models in terms of accuracy, precision, recall and F1-score, the design of the even more advanced

preprocessing algorithms, the design of the user-friendly web-interface with the assistance of Flask, HTML, CSS and JavaScript. The analysis indicates that the random forest model is the optimal model in terms of classification with score of 98.20 and once again the light GB model with score of 98.13 is also noted to be quite helpful in certain cases of the attacks. The conclusion made in this paper can be scaled to enhance security of the IoT network by ensuring that identification and categorization of attack are effectively done.

A. Objective of Project:

This project will also build a competent machine learning based intrusion detecting system (IDS) which shall be deployed on the Internet of Things (IoT) networks. The statistics of this IDS is found in UNSW-NB15, and they are trying to identify and classify the specific attacks of Iots systems, such as Denial of Service (DoS), Exploit, Shellcode and Worms. It will further compare and contrast the performance of the various machine learning models that will include Decision Tree (DT), Random Forest (RF), XGBoost, LightGBM, AdaBoost, Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) such that different attacks could be effectively detected. The best alternative to determine the known and unknown patterns of attacks in the dynamic IoT environments should be preprocessing techniques to obtain more accuracy and reliability of the models. The IDS will be integrated within an simplified web application that is to be developed using flask, HTML, CSS as well as JavaScript such that the user can register and log-in into the system and feed the traffic data in the network to monitor attacks in real-time. Lastly, the project will attempt to deliver scalable, precise and reachable internet of things network security service that will enhance the resiliency of the network to the threats that can occur and restrict sensitive data and essential procedures to cyber-attacks.

B. Problem Statement:

This has posed an immense problem with regard to locking the networked environments with the majority of them being inadequately safeguarded with respect to the absence of an advanced security characteristic that they are vulnerable to various cyber-attacks such as Denial of Service (DoS), Exploits, Shellcode and Worms. The attacks will be able to put businesses on their knees stealing valuable data and claim the inaccuracy of IoT networks and this should have a powerful intrusion detector system (IDS). The existing IDS is not applicable in a generalized way to the case of the IoT, signature based system is inSensible to identify new or emergent threat, machine learning based systems is inSensible to generalize and the traffic of the IoT can be dynamic. The other issue concerns that of enormous quantities of variety in equipments in other issues that ought to be identified in real time. Since the project is successful, the project will aim at creating a highly advanced machine learning-based IDS that will identify high amount of attacks to the IoT network. The proposed system will provide the consistent and versatile solution of the pitfalls the IoT infrastructures will experience within the framework of the advanced security to the majority of cyber threats with the UNSW-NB15 data and algorithms, including Decision Tree, Random Forest, XGBoost, LightGBM, AdaBoost, ANN and CNN.

II. RELATED WORKS

Current paper will assume a speculation of how machine learning algorithms can be used to increase the workability of intrusion detection system (IDS) based on UNSW-NB15 data. The models provided by authors are the ones of Random Forest, Decision Tree, and XGBoost and the matter of concern of authors is the hyperparameters optimisation. Their highest percentage was 98.20 percent on the Random Forest and the emerging understanding that the ML based IDS could be used in the process of categorizing the distinct types of IoT attacks including DoS, Exploits and Worms.[1]

The article provides a dynamic intrusion detection paradigm in internet-of-things. It is anchored on the execution of several machine learning models and testing them experimentally on both NSL-KDD and UNSW-NB15 data sets. The authors provide appropriateness to the appropriateness of feature selection and model adaptation that is more competent in detecting the various types of attacks such as the DoS and Exploits.[2]

The paper will also address the issue of the massive and disproportionate data employed in the process of identification of the IoT intrusion. All the authors employ numerous types of machine learning classifiers, and they all have feature extractors and are

evaluated using the UNSW-NB15 dataset. The results of the work, the algorithms used in the random forest and the AdaBoost, are suitable in terms of the use of the unequal classes, i.e., the DoS attack and the AdaBoost Reconnaissance attack.[3]

In the further paper, the issue of detecting an abnormal traffic within the IoT networks using the machine learning models will be given attention. The authors use the unsupervised learning techniques to reveal any breach of the typical network traffic pattern which demonstrates the fact that the proposed model can be expanded to the threat emergent of the IoT as well, by training it on the UNSW- NB15 dataset.[4]

In this paper, the author introduces a framework, HybridGuard, which further improves the process of identifying the minority-class attacks to an IoT network, especially, in edge scenario. The article combines both the classical machine learning model, as well as the deep learning model, to the development of the hybrid model that will overcome the issue of data imbalance. According to the model, the high scores of the UNSW-NB15 in the domains of attacks under service, which include Shellcode, are observed.[5]

In order to identify anomalies in an IoT network, the federated PCA will be proposed to be applied in this paper. The Grassmann manifold helps the authors to make the dimensionality reduction of the data feasible, through the PCA which increases the accuracy of the detection, in addition, the PCA does not infringe privacy of the data. The data of the UNSW-Nb15 and CICids was tested in the system and it is observed that the unknown patterns of attack were identified.[6]

The lightweight CNN-BiLSTM is the proposed intrusion detection device that is proposed by the article in small IoT. It is a CNN application on extraction of features and BiLSTM on sequential prediction. The authors demonstrate that a hybrid type of the model can be employed in the process of mitigating the high rate of detection at a high level of detection rate when the data set of the UNSW-NB15 has big computation load.[7]

The article identifies the application of LSTM networks in the detection of the intrusion in the IoT network in the form of the Long Short-Term Memory (LSTM). The authors design their own LSTM in order to offer a time association between attack traffic traiffs and assess their model on the UNSW-NB15 data that show better results on the top of the sophisticated attack traffic traffic patterns, than the traditional machine learning models do.[8]

The paper will introduce in this paper a plan of a parameter based intelligent intrusion detection system (IDS) which will be applied to any IoT network. The machine learning algorithms used within the system would be the random forest and the XGBoost depending on the nature of the network traffic selected. This model attempts and test UNSW-NB15 data and it is very useful in revealing DoS and Worm attacks.[9]

Such general overview will involve an argument on the modern methods of intrusion detection that involves the traditional methods of intrusion detection, machine learning based methods of intrusion detection and a hybrid approach of intrusion detection that involves the safety of the IoT. It also means that the issues of the existing datasets selection, feature extraction, and scalability of the models to the new applications of the new scenarios of the IoT are the current issue.[10]

Gives a realistic model of IoT IDS based on LSTM and evaluates the model by a series of benchmarks, including UNSW NB15. Its results demonstrate that versions of RNNs that are better developed also can be trained to understand the attack methods supposing that the network traffic (sequential) is successful.[11]

To manage the issue of limited resources of the devices of the IoT, this paper recommends to introduce an integrated ML + DL IDS, that the given model should be optimized with the help of feature engineering, data balancing and model tuning, and reduce the price of the offered model.[12]

The current article has utilized the CNN-based random forest feature selection through the purpose of reducing the dimension and acknowledgement of the anomalies within the network traffic. The models were also highly precise (approximately 99) and this element suggests the utility of the hybrid option of feature selection and deep learning.[13]

The paper is also founded on the hybrided advanced system that suggests the Generative Adversarial Networks as a balancing sampling system, optimization based feature selection and deep feature extractor. It comprises hybrid type but it is also focused on the improvement of power and accuracy in classifying the attack.[14]

The deep learning models (CNN, RNN, LSTM, MLP) were presented in this paper using the CIC IoT DIAD 2024 data set in order to evaluate the effectiveness of

the deep learning models in real-time identification. It is known that special deep architecture can be applied to enhance precision of detection to a high degree.[15]

III. PROPOSED SYSTEM AND WORKFLOW

The given system is based on the construction of an Intrusion Detection System (IDS) based on machine learning and Internet of Things (IoT)-specific network and is grounded on UNSW-NB15 data to identify and detect eight cyber-attack types, including Denial of Service and Worms. It has some models (the Decision Trees and the Convolutional Neural Networks) that it trains on a processing of the data, optimization of the hyperparameter, and the performance of the model in terms of accuracy and precision as a metric. The models are implemented and deployed on flask server where it can be utilized to analyze the traffic on the network in real-time and a web interface is provided to offer an interaction point with the user. In addition, it has user feedback process to provide unremitting enhancement, which can be scaled and streamlined to provide to the security internet of things net.

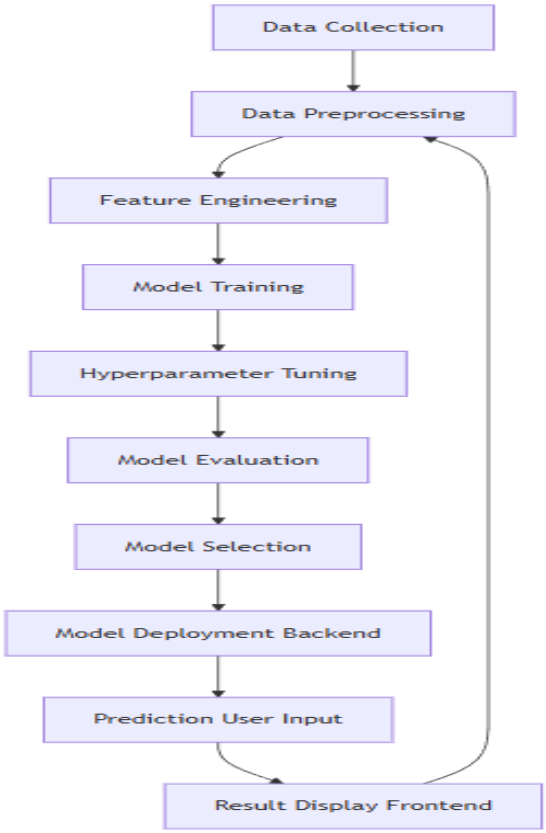

Figure 1 Flow Diagram

IV. METHODOLOGY

A. Dataset

The data that will be used in this project is the UNSW-NB15 which is a highly structured data that would be utilised in the modelling of the network traffic to identify intra intrusion of an IoT network. It contains 2.5 million entries and normal and malicious network traffic entries. Data set is categorized into 9 classes of which 8 of the classes belong to the attack type DoS (Denial of service), Exploits, Fuzzers, Generic, Normal, Reconnaissance, Shellcode and Worms and the remaining single class is normal traffic.

The fundamental data variables will be the size of the packets, connection time, type of protocol and flow time that will be highly helpful in identifying the nature of the network as well as separating the benign and malicious traffic. The attacks are identified and classified by the characteristics. The data is typical and attack traffic, which is equitable and would suit machine learning model training in detection of intrusion. It is based on this that the project would be guided to come up with a practical machine-learning based IDS that would not only serve the purpose of detecting attack on the IoT network effectively but would also be able to categorize the attacks.

B. Data Preprocessing

The largest pre-processing the data is the disincentive of the quality and efficiency of machine learning models that are utilized to determine the intrusions in the IoT networks. Several steps are worth noting in pre-processing of UNSW- NB15 data that will be used in this project. The missing or incomplete data are then cleaned in the data by imputing the values or dropping them to ensure that the data integrity of the data is preserved. The doubles are removed as well in a bid to remove some bias within the model. The second step will be to choose features to either filter or cut the irrelevant or redundant information and only choose the most relevant features that will render the model more efficient and accurate. Continuous numeric characteristics are then assumed to be brought to standard or normalized to some common scale such that no particular numeric characteristic gets the rank of the scale over the rest. The protocol types are categorical variables and may be one-hot encoded or label coded in the form of a number value expressed

by one-hot encoding or label encoding technique, that is why, it can be used in the machine learning algorithms. Lastly, additional manipulations as the creation of the data or the synthesis of the already existing ones are introduced to display more relevant trends in the network traffic. These preprocess will ensure that there are no errors in the data and that it is formatted and ready to be trained to induce an increase in the performance of the model.

C. Model Training

1. Decision Tree (DT): Decision Tree may be identified as a regressive and categorised learning algorithm, which is a supervised learning algorithm. It splits information of both nodes in terms of the most noteworthy quality thus yielding a tree-like structure that leads to the opportunity to make a decision in the situation (features). Recursive partitioning of the data is used to construct the tree using a feature that maximizes the data gain or minimizes the impurity. The leaves will be either to be classified (to be of class) or to be regressed.

Formula:

$$IG(S,A) = Entropy(S) - \sum \frac{|S_v|}{|S|} Entropy(S_v)$$

2. Random Forest (RF): Random Forest is an ensemble training algorithm of machine training, a construction of numerous decision trees through the training of which each decision tree is built on a random part of training data and attributes. The performance of the model is a final result of either the majority (when it comes to classification) or averaging (when it comes to regression) of the predictions of all the individual trees that assist in enhancing the model and over-fitting as well.

Formula:

$$\hat{y} = \text{mode}(y_1, y_2, \dots, y_n)$$

3. XGBoost: XGBoost (Extreme Gradient Boosting) is a well-constructed gradient boosting model, and it builds an accumulation of decision trees in sequential fashion. Just like a tree plays the role of mending the mistakes of the last trees. The model has been highly speed and performance optimized and it also implements regularization to prevent overfitting as well as improve the generalization ability of the model.

Formula:

$$\text{Objective} = \sum_{i=1}^{n} L\,(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega\,(f_k)$$

4. LightGBM: LightGBM (Light Gradient Boosting Machine) is a gradient boosting framework which uses algorithm in terms of histogram in an attempt to better cluster data as compared to the earlier gradient boosting frameworks. It can also be learnt within a shorter period of time especially with large datasets since it does not consume a lot of memory and less computation.

Formula:

$$\text{Objective} = \sum_{i=1}^{n} L\,(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega\,(f_k)$$

5. AdaBoost: AdaBoost (Adaptive Boosting) is an ensemble algorithm, which implies the usage of several weak classifiers to make a good one. It does it by biases the weights of the misclassified samples, so as to reduce the weight in successive classifier of the hard cases to classify. AdaBoost is used to improve the performance of insignificant classifiers e.g. decision stumps (shallow trees) repeatedly at the ensemble.

Formula:

$$\hat{y} = \text{sign}\left(\sum_{t=1}^{T} \alpha_t\, h_t(x)\right)$$

6. Artificial Neural Network (ANN): Artificial Neural Network (ANN) may be considered as a flock of network of neurons (nodes) that have their topology that resembles that of being in layers. This input information is then obtained as the result of calculation by an activation function and the weighted input into all the neurons is gained. The network is fed to learn weights of the connections in such a way to reduce the error in prediction. ANNs would particularly be used in the data modelling of non-linear correlation.

Formula:

$$y = \sigma\left(\sum_{i=1}^{n} w_i\, x_i + b\right)$$

7. Convolutional Neural Network (CNN): Convolutional Neural Network (CNN) is a deep learning, which is primarily applicable in processing of image and space data. It utilises convolutional layers to automatically extract features of the input data, e.g. an edge or a texture and builds more complex representations in the subsequent layers. CNNs have additionally been proved to be the best when it comes to classification of images, recognition of objects etc.

Formula:

$$y = \sum_{i,j} x_{i,j} \cdot w_{i,j} + b$$

V. RESULTS AND DISCUSSION

1. Decision Tree (DT) Results

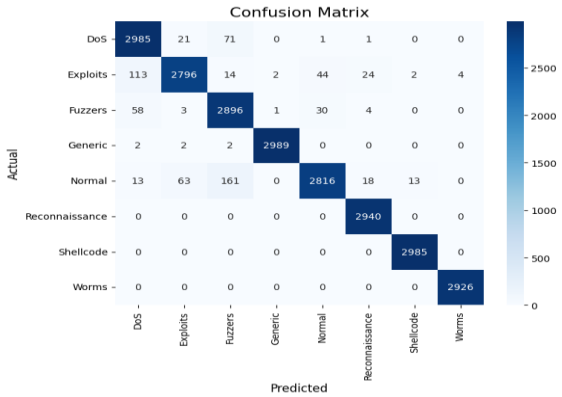

Figure 2 Decision Tree Confusion Matrix

The precision of the Decision Tree model was quite high to categorize the traffic of the IoT networks and was 97.22. The model helped mostly in uncovering the attacks including DoS, Shellcode, Reconnaissance and Worms because it was quite precise and never returned with a value of above 0.94. It also had the capability to identify this kind of attacks, it also showed total accuracy and memory to the generic, shellcode, and worms. However, it reacted to ordinary traffic slightly lesser owing to the fact that it not only plummeted in recall to 0.91 thus signifying that regular cases of regular traffic were misconstrued as assault. Overall, Decision Tree was a stunning model, which could be enhanced, in certain groups, including one called as "Normal" one.

2. Random Forest (RF) Results

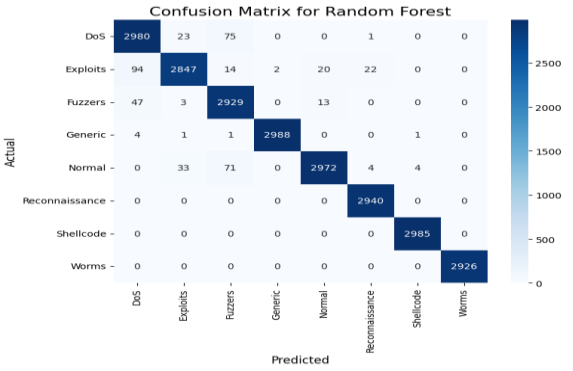

Figure 3 Random Forest Confusion Matrix

The best model that was used in this situation was the Random Forest model which had accuracy of 98.20. It was observed to be very specific and remembers most of the earlier attacks and then more specific the later ones, that is, Generic Shell code and Worms where it had the greatest recall precision. The model also achieved well attacks like DoS, Exploits and Fuzzers with high classification accuracy of recall value of 0.95. It also had a slight drop in recall with the normal traffic at 0.96 but it was doing well. The most reliable model that will be used during the research is the Random Forest because it can treat different forms of attacks and give a similar result.

3. XGBoost Results

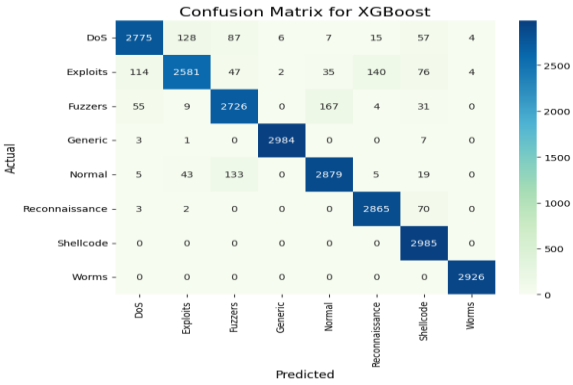

Figure 4 XGBoost Confusion Matrix

The accuracy of XGBoost was 94.67 which was less than the accuracy of the Random Forest and LightGBM, but it was also a good performance. It was discovered to be highly accurate especially in the instance of Generic and Worms where it had attained the ideal accuracy. Nevertheless, the recall of the attacks such as DoS and Exploits were a little less than

the optimal models with the value of 0.90 and 0.86 respectively. The XGBoost (shellcode) and XGBoost (Worms) had high recalls and recalls were perfect. It did not also do the best in the accuracy but its general performance in certain types of attacks was competitive.

4. LightGBM Results

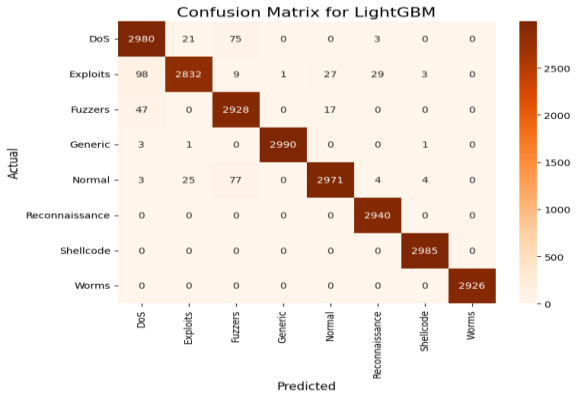

Figure 5 LightGBM Confusion Matrix

LightGBM measure was 98.13 and it is the second behind the random forest. It was quite accurate and could recall most of the attacks in the case of instance of Generic, in which it had the highest degree of recollection of a hundred percent that Shellcode and Worms had the same score. Other forms of attacks like DoS also had good precisions and recall of 0.95 and above. The call in the category of the Norman traffic was slightly small in comparison to the leading models but the peak was in 0.96. LightGBM was a highly prolific model when it comes to detection of IoT intrusion since it was not only able to detect all the variants of attacks, but was also fast and efficient that is, light.

5. AdaBoost Results

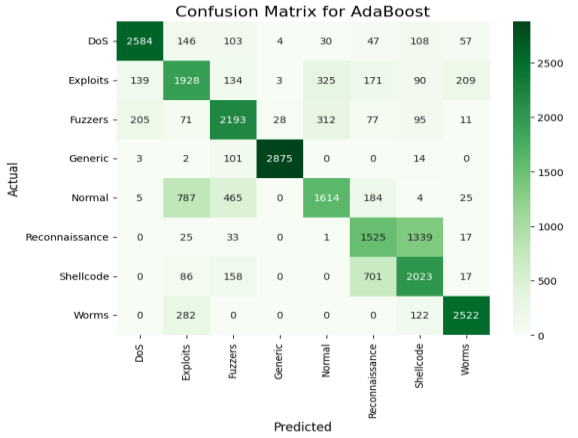

Figure 6 AdaBoost Confusion Matrix

The lowest performing AdaBoost was 71.93 that means that not all the type of attacks were classified. This was highly considered with the DoS and with the Worms, but it was relatively smaller in the other attacks, particularly with the normal and reconnaissance traffic where one could barely tell the difference between the normal and attack traffic. Accuracy of exploits and Shellcode was in actuality very poor which compensated the fact that the model is not responsive to more sophisticated attacks or less frequently used attacks. This is what constrained the overall performance of AdaBoost hence the reason why it was not the appropriate tool to be utilized in such a task as compared to other models.

6. Artificial Neural Networks (ANN) Results

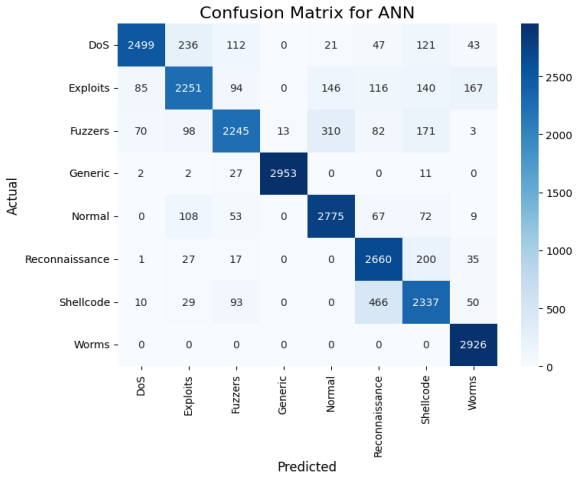

Figure 7 ANN Confusion Matrix

The model of Artificial Neural Network (ANN) has an accuracy of 86.02 per cent which is not very good as compared to the other models such as the random forest and the lightGBM model. The model worked well in DoS and Worms but failed in other attacks such as Exploits and Shellcode which had a low accuracy. Recall of the "Normal" traffic was quite high (0.90) and the recall of such attacks as DoS (0.81) and Exploits (0.75) was not that high and it proves that a certain portion of attacks of this type was not recognized by the model. Nevertheless, ANN worked in the process of detecting some types of attacks, and it might be improved further.

7. Convolutional Neural Networks (CNN) Results

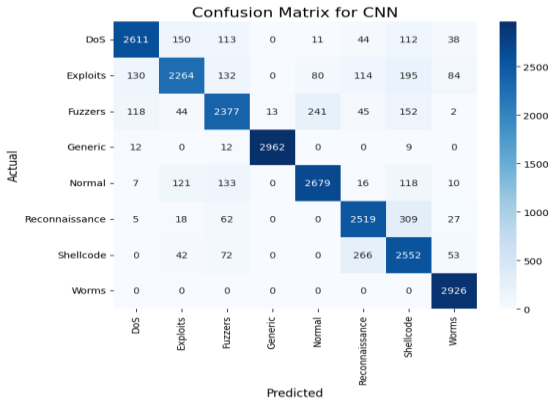

Figure 8 CNN Confusion Matrix

The CNN was not the best model of the accuracy as its score was 87.04 that was quite high, however, it could not compete with LightGBM or the Random Forest. CNN was quite accurate and even more accurate because it was Generic (1.00) and Worms (0.93) and difficult as Shellcode and Exploits had low precisions. The worms recall (1.00) was the most perfect and other attacks such as the DoS (0.85) and exploits (0.75) recall was not (1.00) but worse, thus the model failed to recall such attacks to such high level of precision. Generally CNN has not also performed as well as other models on recall and accuracy on some of the more complex attacks.

Discussion:

The results of the different machine learning models portray that the disparities in the degree of categorizing the attacks of the IoT networks are colossal. The best performance was 98.20 and 98.13 by the Random Forest and LightGBM respectively. These models had worked well in identifying most of the attacks especially the ones of the "Generic" category like Shellcode and Worms where they worked 100 percent. They can retain an enormous number of attack-patterns and the result of the attacks is replicable and it is one of the reasons contributing to their superiority regarding the detection of IoT intrusions. Conversely, the XGBoost was not that excellent (94.67) yet good in the identification of the Generic and Worms. AdaBoost, in its turn, is not as accurate as 71.93 because it fails to handle the attacks like the Exploits and Reconnaissance because it cannot be utilized in this task. ANN and CNN models were also rather good, but they did not match to the ensemble models in terms of the accuracy and recall. On the whole, one can assume that the Random Forest and LightGBM seemed to be the most appropriate models, and AdaBoost and ANN still need additional development to be capable of the classification in the IoT setting.

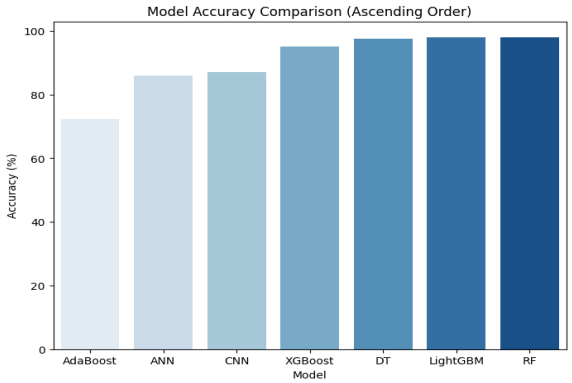

Figure 9 models accuracy comparison

VI. CONCLUSION

A comparison of certain forms of machine learning in the IoT intrusion detection of the UNSW-NB15 data has been made in this paper. These findings are in support of the premise that the ensemble approaches and specifically the Random Forest and LightGBM is more accurate, more precise and recalling in comparison with others. These models proved quite handy in cracking very huge number of attacks like DoS, Shellcode and Worms and the performance was never strict in most types of attacks. Other models like AdaBoost and Artificial Neural Networks (ANN) were less accurate and could not perform under some form of attack, this is why it is not so relevant to the task of intrusion in the IoT detection in the case under consideration.

It has been revealed that the IDS generated by machine learning could be deployed to ensure the safety of the IoT networks, and the most effective of them is the Random Forest and LightGBM. The second approach can be the enhancement of the performance of the non performing models, test the hybrid settings and use more datasets to gain additional progress towards the generalization of the system. In general, the paper is justified regarding the creation of a scalable and effective system of IoT intrusion detector.

VII. FUTURE ENHANCEMENT

The machine learning models of intrusion detection at the IoT are noteworthy in the advancement of features to optimize the detection of network traffic patterns and anomalies with specific features in the available capabilities. In feature engineering, raw packet data is transformed into significant metrics and this enables models to recognize patterns of attack. These techniques will reduce noise and overfitting since they reduce the features to a minimum with Principal Component Analysis (PCA) as a method. Scaling and regularization Continuous variables are of importance to feature-sensitive algorithms. The redundant features may be filtered with the help of the feature selection tool, such as mutual information or Recursive Feature Elimination (RFE) that enhances the performance of the model. Since attacks like DoS and DDoS can be detected based on temporal factors like day and session time. Moreover, network traffic aggregation can indicate the time trends of attacks that are beneficial in improving the quality of data and the ability of the model to recognize different IoT network attacks.

REFERENCES

[1] S. More, M. Idrissi, H. Mahmoud, and A. T. Asyhari, "Enhanced Intrusion Detection Systems Performance with UNSW-NB15 Data Analysis," Algorithms 2024, Vol. 17, Page 64, vol. 17, no. 2, p. 64, Feb. 2024, doi: 10.3390/A17020064.

[2] F. S. Alrayes, S. U. Amin, and N. Hakami, "An Adaptive Framework for Intrusion Detection in IoT Security Using MAML (Model-Agnostic Meta-Learning)," Sensors 2025, Vol. 25, Page 2487, vol. 25, no. 8, p. 2487, Apr. 2025, doi: 10.3390/S25082487.

[3] M. P. Sahu and A. Tamrakar, "DESIGNING AN INTELLIGENT INTRUSION DETECTION MODEL FOR IOT SYSTEMS USING MACHINE LEARNING," International Journal of Original Recent Advanced Research ISSN, vol. 02, p. 6, 2025.

[4] N. Sarwar, R. S. Alharthi, M. Aljohani, and M. A. Elhosseini, "Securing IoT networks: a machine learning approach for detecting unusual traffic patterns," Sci. Rep., vol. 16, no. 1, pp. 3397-, Dec. 2025, doi: 10.1038/S41598-025-33447-Z;SUBJMETA=166,639,705;KWRD=ENGINEERING,MATHEMATICS+AND+COMPUTING.

[5] B. Kar, U. Sahu, C. Thomas, and J. P. Sahoo, "HybridGuard: Enhancing Minority-Class Intrusion Detection in Dew-Enabled Edge-of-Things Networks," Computer Networks, vol. 276, Nov. 2025, doi: 10.1016/j.comnet.2025.111966.

[6] T.-A. Nguyen et al., "Federated PCA on Grassmann Manifold for IoT Anomaly Detection," IEEE/ACM Transactions on Networking, vol. 32, no. 5, pp. 4456–4471, Jul. 2024, doi: 10.1109/TNET.2024.3423780.

[7] M. Jouhari and M. Guizani, "Lightweight CNN-BiLSTM based Intrusion Detection Systems for Resource-Constrained IoT Devices," 20th International Wireless Communications and Mobile Computing Conference, IWCMC 2024, pp. 1558–1563, Jun. 2024, doi: 10.1109/IWCMC61514.2024.10592352.

[8] "(PDF) Intrusion Detection in IoT Networks Using LSTM Deep Learning Models with the UNSW-NB15 Dataset." Accessed: Feb. 06, 2026. [Online]. Available: https://www.researchgate.net/publication/391937093_Intrusion_Detection_in_IoT_Networks_Using_LSTM_Deep_Learning_Models_with_the_UNSW-NB15_Dataset?utm_source=chatgpt.com

[9] M. Luqman et al., "Intelligent parameter-based in-network IDS for IoT using UNSW-NB15 and BoT-IoT datasets," J. Franklin Inst., vol. 362, no. 1, p. 107440, Jan. 2025, doi: 10.1016/J.JFRANKLIN.2024.107440.

[10] M. M. Rahman, S. Al Shakil, and M. R. Mustakim, "A survey on intrusion detection system in IoT networks," Cyber Security and Applications, vol. 3, p. 100082, Dec. 2025, doi: 10.1016/J.CSA.2024.100082.

[11] "(PDF) IDS-IoT: Intrusion Detection System for the Internet of Things Using Enhanced Long-Short Term Memory." Accessed: Feb. 06, 2026. [Online]. Available:https://www.researchgate.net/publication/396750333_IDSIoT_Intrusion_Detection_System_for_the_Internet_of_Things_Using_Enhanced_Long-Short_Term_Memory?utm_source=chatgpt.com

[12] H. Fares, N. Aknin, G. Lazrek2, and M. Zeroual, "Intrusion Detection in IoT Environment Using Hyperparameters Tuned Machine and Deep

Learning Models on the CICIoT2023 Dataset," Informatica, vol. 49, no. 5, Aug. 2025, doi: 10.31449/INF.V49I5.8881.

[13] A. D. Vibhute, M. Khan, C. H. Patil, S. V. Gaikwad, A. V. Mane, and K. K. Patel, "Network anomaly detection and performance evaluation of Convolutional Neural Networks on UNSW-NB15 dataset," Procedia Comput. Sci., vol. 235, pp. 2227–2236, Jan. 2024, doi: 10.1016/J.PROCS.2024.04.211.

[14] R. Ghadami, "An intrusion detection system in the Internet of Things with deep learning and an improved arithmetic optimization algorithm (AOA) and sine cosine algorithm (SCA)," Sci. Rep., vol. 15, no. 1, pp. 38156-, Dec. 2025, doi: 10.1038/S41598-025-22074-3;SUBJMETA=117,166,639,705,987;KWRD=COMPUTER+SCIENCE,ELECTRICAL+AND+ELECTRONIC+ENGINEERING.

[15] M. A. Hossain, "Deep learning-based intrusion detection for IoT networks: a scalable and efficient approach," EURASIP J. Inf. Secur., vol. 2025, no. 1, pp. 28-, Dec. 2025, doi: 10.1186/S13635-025-00202-W/FIGURES/10.