FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 8 No. 3, June, 2024, pp 356 - 361
DOI: https://doi.org/10.33003/fjs-2024-0803-2448

COMPARATIVE ANALYSIS OF RANDOM FOREST AND ADABOOST LEARNING MODELS FOR THE CLASSIFICATION OF ATTACKS IN INTERNET OF THINGS

***[1]Adeniyi, Usman Adedayo, [2]Alimi, Maruf Olasunkanmi, [3]Oyelakin, Akinyemi Moruff and [1]Abdullahi, Samaila Musa**

[1]Cyber security Department, Faculty of Computing, Air Force Institute of Technology Kaduna.
[2]Computer Science Department, Faculty of Computing, Air Force Institute of Technology Kaduna
[3]Computer Science Department, College of Information and Communication Technology, Cresent University Abeokuta

*Corresponding authors' email: adedayousman1@gmail.com

ABSTRACT

Attacks are actions that attempt to break one of the following properties of the computer system: confidentiality, integrity, and availability. The immense increment in the amount of internet applications and the appearance of modern networks has created the need for improved security mechanisms. Internet of Things (IoT) is a system that uses the Internet to facilitate communication between sensors and devices. Several approaches have been used to build attacks detection system in the past. This study built two ensemble models for the classification of attacks using Random Forest and Adaboost algorithms respectively. Feature importance was used for selecting promising attributes from the IoT intrusion dataset. Thereafter, the results of the classification models were evaluated and compared. The models were evaluated based on when feature selection technique was applied and without respectively. For Random Forest-based classification model with feature selection, 99.0% ,0.95,0.88,0.82, were obtained for accuracy, recall, f1-score, and precision respectively while without feature selection 69.0%,0.86,0.76,0.64 were obtained respectively. For Adaboost-based classification model with feature selection 99.0%.0.69,0.61,0.66 were obtained for accuracy, recall, f1-score and precision respectively. Without feature selection the Adaboost model recorded 58.0%,0.58,0.48,0.50 respectively. The results showed that both models achieved high rates with feature selection technique used, with Random Forest performing slightly better, both learning models showed promised performances in classifying attacks in IoT environments. This study concluded that the use of the chosen feature selection method helped improve the performances of the two ensembles in the classification of attacks in the IoT dataset.

Keywords: Internet of Things, Machine Learning, Attacks in IoT, Security, Classification

INTRODUCTION

Attacks are activities that aim to violate one of the hardware or software system's confidentiality, integrity, or availability requirements. Attacks are in various forms due to the threats that are pervasive in networks and the cyber space (Ibitoye, Shafiq & Matrawy., 2019) In recent time machine learning (ML) algorithms are getting popular for classifying attacks in network (Oyelakin et al., 2020). The internet of things is a network of connected devices that can link without human involvement, thanks to the enormous growth in the number of online applications and the development of contemporary technology. IoT enables a large number of items with sensors (including bicycles, coffee makers, lights, and many more items).

IoT applications are transforming our work and lives by connecting to the internet in sectors such as healthcare, agriculture, transportation, etc. Additionally, it offers countless benefits and countless chances for the sharing of knowledge, innovation, and progress (Alsamiri & Alsubhi, 2019). IoT technology can collect, analyze, and comprehend data about the environment, allowing for modernizations that raise living standards. By making new types of communication between machines and people simpler, smart cities can be created (Tasnim, Hossain, Tabassum & Parvin, 2022). In the modern world, IoT technology is used in a variety of ways. Everything has become intelligent, including entry doors, window blinds, watches, TVs, fans, lightbulbs, and refrigerators. The amount of device engagement is growing daily. Their reliance is increasing as a result. Attackers may not directly hack the target system, but they can easily alter the behavior of other interdependent devices or the surrounding environment to achieve their objectives. Again Some compact IoT devices are missing a memory management unit (MMU). These devices employ a variety of complex encryption and authentication techniques, which consume excessive processing power and result in a noticeable delay, impairing normal operation and lowering performance, especially for real-time IoT devices. Because of this, it is easy for attackers to compromise these devices by taking advantage of memory flaw (Tasnim et al., 2022)

IoT devices have proliferated greatly in recent years, and it is anticipated that by 2020, there will be close to 30 billion of them on the market. The market competitiveness and technical constraints, however, make it difficult to increase the security of these devices. Even worse, default usernames and passwords are frequently left unchanged, which makes these devices a prime target for attack by attackers. A continual danger to the ever-expanding IoT world, new botnets like Hajime and Reaper demonstrate how adversaries are always changing their tactics to avoid detection. These IoT botnets can swiftly develop into a potent collection of weapons to seriously harm a number of stakeholders. They also exploit a manufacturer's default settings to scan the Internet for other devices (Shaikh, Bou-Harb, Crichigno, & Ghani, 2018).

IoT nodes are unlike other traditional networks in that they lack manual controls, have minimal capacity, and few resources. Additionally, IoT security challenges are becoming increasingly problematic due to the widespread use and rapid proliferation of IoT devices in daily life, necessitating the creation of network-based security solutions. While the existing methods do a good job of detecting some threats, it is still difficult to find others. There is no doubt that there is room for more advanced techniques to improve network security as network attacks rise in number and the amount of information present in networks multiplies dramatically (Alsamiri et al., 2019). Unauthorized individuals may take advantage of a network vulnerability in order to obtain sensitive data and harm the network (Alladi, Chamola, Sikdar & Choo, 2020). This study compered random forest and adaboost machine learning algorithms for classifying attacks in internet of things and focuses on how to achieve improved ensemble based attacks classification models in Internet of Things environment.

MATERIALS AND METHODS

The methodology used in this research is collection of data from kaggle online repository, https://www.kaggle.com/datasets/azalhowaide/iot-dataset-for-intrusion-detection-system-ids. pre-process and analyse it to improve machine learning results by combining two models. This approach allows the production of better predictive performance compared to a single model. Basic idea is to learn a set of classifiers.

Each step of the methodology is logically detailed corresponding with the activities to accomplish each objective of the study.

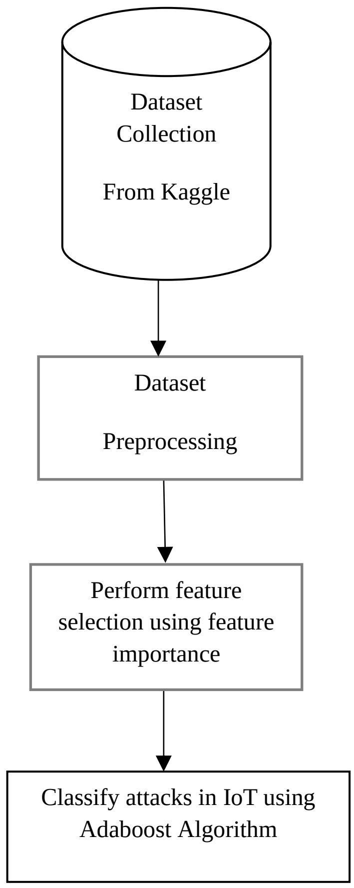

Figure 1: The Methodological Flow of the Proposed Random Forest and Adaboost Model

Figures 1 is used to illustrate the different stages in the machine learning-based classification of attacks in internet of things. Python Programming language was used for the implementation of various stages in the proposed model. The basic stages in the machine learning-based model classified in the implementation.

Evaluation Metrics

Evaluation of the two model Random Forest and Adaboost were perform using accuracy, precision, recall and f-measure**.**

The percentage of accurate predictions to all other guesses is known as accuracy. The percentage of all normal and attack data that are correctly classified serves as a measure of a model's overall effectiveness.

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

Precision The ratio of successfully predicted positive data to all anticipated positive data, including True Positive and False Positive Values, is what is meant by this term. The amount of successfully predicted positive data, including True Positive and False Positive values, is proportional to the total amount of anticipated positive data. It measures a model's overall effectiveness by counting how many of all attack scenarios are true.

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

The proportion of accurately predicted positive data to all anticipated True Positive and False Negative values is known as recall. Recall is defined as the proportion of correctly predicted positive data to all expected True Positive and False Negative values. Models that identify False Negative Values have an impact on the recall metrics

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

F-measure is an evaluation statistic used to describe how well a machine learning model (or classifier) is performing. It provides a model's precision and recall information in combination. This means that a high F1-score denotes a strong recall and precision value. F-measure is a classification evaluation metric that is defined as the harmonic mean of recall and precision. It is a metric used in statistics to assess how accurate a test or model is. It is expressed as follows in mathematical equation,

$$\text{F-measure}=\frac{2*Recall*Precision}{Recall+Precision} \quad (4)$$

Data Pre-processing

Missing Values was used as the pre-processing technique for this study. Missing values were handled by dropping rows with missing values. The specific columns affected by missing values are listed,

Table1: Identifying Rows With Missing Values

Missing Values	Number of Missing Values in the Dataset:
Flow duration	0
Header Length	0
Protocol Type	0
Duration	0
Rate	0
Srate	0
Drate	0
UDP	0
DHCP	0
ARP	0
ICMP	0
IPv	0
LLC	0
Tot sum	0
Min	0
Max	0
AVG	0
Std	0
Tot size	0
IAT	1
Number	1
Magnitue	1
Radius	1

Table 1: is used to depict the columns (feature) with missing values. The missing values were handled.

Table 2: Results of IoT Attack Classification Models

Models	Accuracy	Recall	F1-score	Precision
Random Forest Classifier with feature selection	0.99	0.95	0.88	0.82
Random without feature selection	0.69	0.86	0.76	0.64
AdaBoost Classifier with feature selection	0.99	0.69	0.61	0.66
AdaBoost without Feature selection Report	0.58	0.58	0.48	0.50

These outcomes offer valuable insights into the models performance in classifying attacks within IoT environments. Notably, the Random Forest Classifier and AdaBoost Classifier, both incorporating feature selection, achieved the highest accuracy rates. Conversely, models lacking feature selection demonstrated comparatively lower performance, especially in terms of precision and F1-score. These results underscore the substantial influence of judicious feature selection on the efficacy of models in the realm of IoT attack classification.

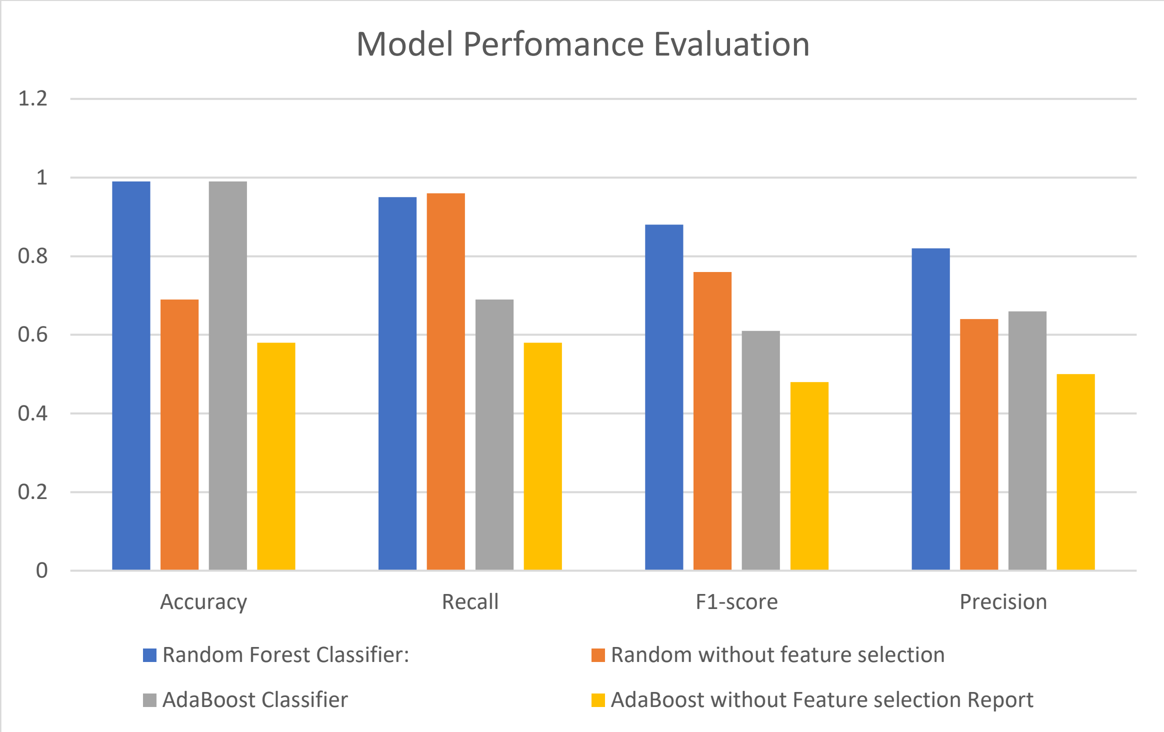

Figure 2: Performance of the Attack Detention Models

Summary and Findings

This study focuses on a comparative analysis of Random Forest and Adaboost learning models for classifying attacks in the Internet of Things (IoT). The objective is to investigate the performance of these algorithms in detecting different types of attacks prevalent in IoT networks. The IoT environment poses unique security challenges due to the interconnection of various devices with limited resources and memory capacity. Traditional security approaches may not be directly applicable in IoT due to these constraints. The research aims to collect and preprocess an IoT dataset, apply feature importance techniques for selecting relevant attributes, and then use Random Forest and Adaboost to classify attacks. The models' performance was evaluated using metrics such as accuracy, recall, and f1-score, precision. The study compared Random Forest and Adaboost machine learning models for classifying attacks in the Internet of Things (IoT). The models were evaluated using a dataset of IoT network traffic data, including normal and attack instances. Both models achieved high accuracy rates, with Random Forest performing slightly better in accuracy, and Adaboost showing higher precision and recall. Feature selection techniques significantly influenced model performance. Overall, both models showed promise in detecting and classifying attacks in IoT environments, with potential for future ensemble techniques to further enhance performance.

CONCLUSION

In this study, two ensemble machine learning algorithms are used for the classification of attacks in the IOT environment. Then, the results of the models were evaluated and compared. Performance measures were conducted to test the accuracy of the two models in the classification of different types of attacks that were found in the chosen dataset. The metrics used for the evaluation are accuracy, recall, f1-score and precision respectively. Though both models achieved promising results when feature importance was applied as attribute selection method, the results for without feature selection were not too good. For the former, it was found out that the Random Forest Classifier outperformed AdaBoost Classifier. Based on these findings, it can be said that the Random Forest Classifier for the targeted is more effective and trustworthy.

REFERENCES

Alhowaide A., Alsmadi, Tang J. (2021). Towards the design of reel-time autonomous internet of things network intrusion detection system, *Cluster Computing, 1-14.* doi.org/10.1007/s10586-021-03231-5

Alladi, T., Chamola, V., Sikdar, B., & Choo, K. K. R. (2020). Consumer internet of things Security vulnerability case studies and solutions. *IEEE Consumer Electronics Magazine*, *9*(2), 17-25. **DOI:** 10.1109/MCE.2019.2953740

Alsamiri, J., & Alsubhi, K. (2019). Internet of things cyber-attacks detection using machine learning. *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 12. DOI:10.14569/IJACSA.2019.0101280

Alzahrani, M. Y., & Bamhdi, A. M. (2022). Hybrid deep-learning model to detect botnet attacks over internet of things environments. *Soft Computing*, *26*(16), 7721-7735. doi.org/10.1007/s00500-022-06750-4

Al-Zewairi, M., Almajali, S., & Ayyash, M. (2020). Unknown security attack detection using shallow and deep ANN classifiers. *Electronics*, *9*(12). https://doi.org/10.3390/electronics9122006

Classen, J., Gringoli, F., Hermann, M., & Hollick, M. (2022). Attacks on Wireless Coexistence: Exploiting Cross-Technology Performance Features for Inter-Chip Privilege Escalation. In *2022 IEEE Symposium on Security and Privacy*

(SP) (pp. 1229-1245). IEEE. DOI**:** 10.1109/SP46214.2022.9833639

Churcher, A., Ullah, R., Ahmad, J., Ur Rehman, S., Masood, F., Gogate, M. & Buchanan, W. J. (2021). An experimental analysis of attack classification using machine learning in IoT networks. *Sensors*, *21*(2), 446. https://doi.org/10.3390/s21020446

Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20-28. doi.org/10.38094/jastt20165

de Souza, C. A., Westphall, C. B., Machado, R. B., Loffi, L., Westphall, C. M., & Geronimo, G. A. (2022). Intrusion detection and prevention in fog based IoT environments: A systematic literature review. *Computer Networks*, 109154.

Deogirikar, J., & Vidhate, A. (2017). Security attacks in IoT: A survey. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (pp. 32-37). IEEE.

Dissanayake, M. B. (2021). Feature Engineering for Cyber-attack detection in Internet of Things. *International Journal of Wireless and Microwave Technologies*, *11*(6), 46-54.

Fatai A. and Safiriyu I. E. (2012) Application of artificial intelligence in network intrusion detection. World applied programming, 158-166 ISSN: 2222-2510.

Guerrero-Sanchez, A. E., Rivas-Araiza, E. A., Gonzalez-Cordoba, J. L., Toledano-Ayala, M., & Takacs, A. (2020). Blockchain mechanism and symmetric encryption in a wireless sensor network. *Sensors*, *20*(10), 2798.

Haji, S. H., & Ameen, S. Y. (2021). Attack and anomaly detection in iot networks using machine learning techniques: A review. *Asian journal of research in computer science*, *9*(2), 30-46.

Ibitoye, O., Shafiq, O., & Matrawy, A. (2019). Analyzing adversarial attacks against deep learning for intrusion detection in IoT networks. *IEEE global communications conference (GLOBECOM)* (pp. 1-6). IEEE.

Ignacio P.R and Maria M R (2008) evaluation of current is intrusion detection system. *International journal LITH-ISY-EX*—08/4160---SE.

Jimoh, R. G., Oyelakin, A. M., Olatinwo, I. S., Obiwusi, K. Y., Muhammad-Thani, S., Ogundele, T. S. & Ayepeku, O. F. (2022). Experimental evaluation of ensemble learning-based models for twitter spam classification. *Information Technology for Education and Development (ITED)* (pp. 1-8). IEEE.

Kwon, S., Park, S., Cho, H., Park, Y., Kim, D., & Yim, K. (2021). Towards 5G-based IoT security analysis against Vo5G eavesdropping. *Computing*, *103*, 425-447.

Kaliyar, P., Jaballah, W. B., Conti, M., & Lal, C. (2020). LiDL: localization with early detection of sybil and wormhole attacks in IoT networks. *Computers & Security*, *94*, 101849.

Lekhraj M.M and Monika B.G (2014) An Effectual and Secure Approach for the Detection and Efficient Searching of Network Intrusion Detection System (NIDS). *International Journal of Computer Applications (0975 – 8887) Volume* 108 – No 15,

Mishra, S., & Tyagi, A. K. (2022). The role of machine learning techniques in internet of things-based cloud applications. *Artificial intelligence-based internet of things systems*, 105-135.

Mosavi, A., Sajedi Hosseini, F., Choubin, B., Goodarzi, M., Dineva, A. A., & Rafiei Sardooi, E. (2021). Ensemble boosting and bagging based machine learning models for groundwater potential prediction. *Water Resources Management*, *35*, 23-37.

Meng, Li, W., W., & Au, M. H. (2020). Enhancing collaborative intrusion detection via disagreement-based semi-supervised learning in IoT environments. *Journal of Network and Computer Applications*, *161*, 102631.

Oyelakin, A. M. (2021). An Investigation into the Performances of Supervised Learning Algorithms in Different Phishing Datasets. *Pakistan Journal of Engineering, Technology & Science*, *9*(2).

Oyelakin, A., & Jimoh, R. G. (2020). Towards Building an Improved Botnet Detection Model in Highly Imbalance Botnet Dataset-A Methodological Framework. *Volume*, *3*(03), 2020

Paricherla, M., Babu, S., Phasinam, K., Pallathadka, H., Zamani, A. S., Narayan, V., ... & Mohammed, H. S. (2022). Towards Development of Machine Learning Framework for Enhancing Security in Internet of Things. *Security and Communication Networks*.

Saheed, Y. K., Abiodun, A. I., Misra, S., Holone, M. K., & Colomo-Palacios, R. (2022). A machine learning-based intrusion detection for detecting internet of things network attacks. *Alexandria Engineering Journal*, *61*(12), 9395-9409.

Saheed, Y. K., Baba, U. A., Orje-Ishegh, T., & Longe, O. B. (2022). An Efficient Machine Learning and Deep Belief Network Models for Wireless Intrusion Detection System. *Research Square* https://doi.org/10.21203/rs.3.rs-2110380/v1

Sivasankari, N., & Kamalakkannan, S. (2022). Detection and prevention of man-in-the-middle attack in iot network using regression modeling. *Advances in Engineering Software*, *169*, 103126.

Sahu, A. K., Sharma, S., Tanveer, M., & Raja, R. (2021). Internet of Things attack detection using hybrid Deep Learning Model. *Computer Communications*, *176*, 146-154.

Shaikh, F., Bou-Harb, E., Crichigno, J., & Ghani, N. (2018, June). A machine learning model for classifying unsolicited IoT devices by observing network telescopes. *International Wireless Communications & Mobile Computing Conference (IWCMC)* (pp. 938-943). IEEE.

Sicato, J. C. S., Singh, S. K., Rathore, S., & Park, J. H. (2020). A comprehensive analyses of intrusion detection system for IoT environment. *Journal of Information Processing Systems*, *16*(4), 975-990.

Shah, Y., & Sengupta, S. (2020). A survey on Classification of Cyber-attacks on IoT and IIoT devices. In *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (pp. 0406-0413). IEEE.

Stiawan, D., Idris, M. Y. B., Defit, S., Triana, Y. S., & Budiarto, R. (2022). Improvement of attack detection performance on the internet of things with PSO-search and random forest. *Journal of Computational Science*, *64*, 101833.

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, *20*(1), 3-29.

Tasnim, A., Hossain, N., Parvin, N., Tabassum, S., Rahman, R., & Hossain, M. I. (2022, March). Experimental Analysis of Classification for Different Internet of Things (IoT) Network Attacks Using Machine Learning and Deep learning. *International* Conference on Decision Aid Sciences and Applications (DASA) (pp. 406-410). IEEE.

Tasnim, A., Hossain, N., Tabassum, S., & Parvin, N. (2022). *Classification and Explanation* of Different Internet of Things (IoT) Network Attacks Using Machine Learning, Deep Learning And XAI

Zagrouba, R., & Alhajri, R. (2021). Machine learning based attacks detection and countermeasures in IoT. *International Journal of Communication Networks and Information Security*, *13*(2), 158-167.