I.J. Wireless and Microwave Technologies, **2025, 2, 41-50**
Published Online on April 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijwmt.2025.02.04

Novel Machine Learning Approaches for Identifying Attacks in IoT-based Smart Home Environment

Oyelakin A. M.*
Department of Computer Science, College of Information and Communication Technology, Crescent University, Abeokuta, Nigeria
Email: moruff.oyelakin@cuab.edu.ng
ORCID iD: https://orcid.org/0000-0003-2844-4837
*Corresponding Author

Sanni S. A.
Department of Computer Science, Faculty of Science and Engineering**,** University of Eswatini, The Kingdom of Eswatini
Email: ssanni@uneswa.ac.sz
ORCID iD: https://orcid.org/0000-0003-0072-137X

Adegbola I. A.
Department of Computer Science, Oyo State College of Education, Lanlate, Nigeria
Email: adegbolaia@oyscoel.edu.ng
ORCID iD: https://orcid.org/0000-0002-0268-1992

Salau-Ibrahim T. T.
Department of Cyber Security, Faculty of Computing, Federal University of Lafia, Nigeria
Email: taofeekat.tosin@cmp.fulafia.edu.ng
ORCID iD: https://orcid.org/0000-0002-0904-5673

Bakare-Busari Z. M.
Department of Computer Science, College of Information and Communication Technology, Crescent University, Abeokuta, Nigeria
Email: bakare.busari@cuab.edu.ng
ORCID iD: https://orcid.org/0009-0005-6574-0792

Saka B. A.
Department of Computer Science, College of Information and Communication Technology, Crescent University, Abeokuta, Nigeria
Email: badirat.saka@cuab.edu.ng
ORCID iD: https://orcid.org/0000-0001-5219-4570

Received: 14 October, 2024; Revised: 04 January, 2025; Accepted: 14 February, 2025; Published: 08 April, 2025

Abstract: Attackers keep launching different attacks on computer networks. Signature-based and Machine Learning (ML)-based techniques have been used to build models for promptly identifying these attacks in networks. However, ML-based approaches are more popular than their counterparts because of their ability to detect zero-day attacks. In the Internet of Things (IoT), devices are interconnected and this called for the need to guide such networks against intrusions. This study aims at building effective ML models from a recently released IoT-based Smart Home dataset. The study revealed patterns and characteristics of the IoT dataset, pre-processed it and then selected discriminant features using Binary Bat Algorithm (BBA). The pre-processing of the Smart Home IoT dataset for the study was carried out based on the issues identified during the exploratory analyses. The experimental evaluation carried out revealed that all the learning algorithms achieved promising classification results. For instance, Decision Trees recorded 98.60% accuracy, KNN produced 99.60% accuracy while Random Forest (RF) and AdaBoost-based models recorded 100.00% and 99.91% respectively. In all other metrics, RF-based attack classification model slightly recorded the best

results. The study concluded that the EDA, innovative data pre-processing, BBA-based feature selection improved the classification performances of the ML approaches used in this study.

Index Terms: Internet of Things, Attacks in IoT, Smart Homes, Meta Heuristic, Classification Model Performance

1. Introduction

Different kinds of cyber attacks are being launched in networks and the cyberspace by people with malicious intent, in recent times [1]. A good example of evolving networks where such attacks are launched is the Internet of Things (IoT). Devices in the Internet of Things (IoT) are widespread and thus securing these interconnected ecosystems against cyber threats has become necessary. According to [2] the number of attacks on the Internet of Things (IoT) platforms globally is more than 112 million in the year 2022 alone. Internet of Things (IoT) refers to the application of interconnected digital objects such as sensors and processing units (Boyes et al., 2018). It has to be mentioned that Internet of Things (IoT) networks have become popular as there are several use cases in both the domestic and industrial worlds.

IoT technology is used to collect, analyse, and comprehend data about the environment, allowing for modernisations that raise living standards. By making new types of communication between machines and people simpler, smart cities can be created [3]. IoT offers countless benefits and countless chances for the sharing of knowledge, innovation, and progress [4]. The exponential growth of the Internet of Things (IoT) devices provides a large attack surface for intruders to launch more destructive cyber-attacks. It has been identified that most of the studies that build machine learning models are too generic and did not establish the real scenarios which the dataset was created for.

A Smart Home Environment refers to a residence where devices and appliances are interconnected through the Internet of Things (IoT) technology to automate and control various household functions, often remotely, via smartphones or voice commands [5] It was reported by [6] that various technological approaches to design and deploy Smart Home Systems (SHS). The paper mentioned that the approaches include Wireless Sensor Network (WSN)-Based SHS, Multiagent System-Based SHS, Internet of Things (IoT)-Based SHS, Artificial Intelligence (AI)-Based SHS, Bluetooth-Based SHS and more. [7] provided insights into the vulnerabilities of IoT devices and the types of attacks they face and highlighted the prevalence of vulnerabilities that make these devices susceptible to a range of cyberattacks. Their report suggested that approximately 80% of IoT devices are prone to various types of attacks, including data theft, device hijacking, and denial-of-service (DoS) attacks. Similarly, [8] pointed out that globally, the end-user cyber security spending will grow by over 15% year-on-year over the course of 2025/ This means the spending will hit a new high of $212bn (£160.5bn).

[9] provided a survey on various datasets as well as some of the ML methods that have been used to classify attacks in IoT, generally. The study established that the security datasets in IoT are diverse and are prominent for intrusion detection studies. The study further argued that to build a promising ML-based attack detection in IoT, the researcher should engage in using innovative approaches to address issues in the datasets prior to using them to build intrusion classification models. As claimed by [9], ML approaches have been widely used to build such intrusion detection systems because they are more accurate when built from big and representative dataset. [10] equally argued that ML techniques can have significant impacts in the early detection of attacks in Internet of Medical Things (IoMT) where attacks have been on the increase in recent times.

Apart from this, some recent researches have established that ensemble learners have become popular in cybersecurity research, and have been found promising in building intrusion detection systems [11], [12] as against the single learners. This study proposed a Multi-Stage Attack Classification Model using EDA, Pre-processing, Binary BAT Feature Selection technique as well as single and ensemble learners as classification algorithms. Specifically, the study aims to use various EDA techniques to unravel the patterns in the Smart Home IoT dataset chosen, address issues identified through data wrangling, carry out feature selection based on a metaheuristic technique called Binary Bat Algorithm (BBA), and classify attacks using the ML algorithms. The choice of BBA in this study is due to the fact that the meta heuristic technique has been argued to be superior to other well-known algorithms such as genetic algorithm (GA) and particle swarm optimization (PSO) [13].

2. Related Work

[14] provided a detailed overview of UNSW-NB15(DS-1) and NF-UNSWNB15(DS-2) datasets used for intrusion detection studies. The researchers built models using the datasets. The models applied the MaxAbsScaler algorithm to implement a filter-based feature scaling strategy. The learning algorithms used in the study are Support Vector Machines (SVM), K-nearest neighbours (KNN), Logistic Regression (LR), Naive Bayes (NB), Decision Tree (DT), and Random Forest (RF). The authors pointed out that the accuracy tests for the multi-class classification scheme were improved from 60% to 94% using the MaxAbs Scaler-based feature scaling method.

[15] built an intrusion detection system to detect network anomalous traffic in IoT. The authors used the binary grey wolf optimizer (BGWO) heuristic algorithm. Thereafter, they applied recursive feature elimination (RFE) to select the most relevant feature subset for the target class. The study used a dataset with high class imbalance and thus synthetic minority oversampling technique (SMOTE) was used to oversample the minority class. The pre-processed data are then classified using XGBoost. The researchers argued that their proposed method achieved good results in accuracy, precision, recall, and an F1 score in the datasets used.

[16] presented a new smart home dataset that integrates network traffic, smart device data, and environmental sensor data into a single timeline. The researchers claimed that the dataset is comprehensive and allows for in-depth analysis of user behaviour to identify anomalies. It was also mentioned that the dataset contains activities from two individuals: the primary actor, who recorded activities over three weeks, and the secondary actor, who recorded activities over several days. Authors claimed that their work focuses on helping researchers create new machine learning models or improve existing ones to predict and detect anomalies in user behaviour.

[17] proposed machine learning models for the classification of attacks in an IoT network. Five supervised learning models were built, tested and evaluated using IoTID20 dataset. The chosen algorithms are shallow neural networks (SNN), decision trees (DT), bagging trees (BT), k-nearest neighbour (kNN), and support vector machine (SVM). The authors claimed that their experimental evaluation achieved an accuracy of 100% recorded for the detection in all cases. [18] built a deep learning-based ensemble classification model for the detection of malware in IoT devices. The model was built using a three-step approach. The data is pre-processed using scaling, normalization, and de-noising, whereas in the second step, features are selected and one hot encoding is applied followed by the ensemble classifier based on CNN and LSTM outputs for detection of malware. The study achieved an average accuracy of 99.5%.

Moreover, [19] proposed a Hybrid Intrusion Detection system using machine learning algorithms such as RF, XGboost, Decision tree, K-Nearest Neighbors, and misuse detection technique. The authors used NSL-KDD, CSE-CIC-IDS2018 datasets for the evaluation of the algorithms. The study reported promising results in all metrics used for the evaluation. However, it is observed that the datasets are not solely focused on IoT-based attacks. Similarly, [20] proposed methods for improving the effectiveness and deployment of anomaly-based intrusion detection system in real-world networks by exploring the application of machine learning models for anomaly-based intrusion detection. The study reviewed various neural network architectures such as Multilayer Perceptron (MLP), Autoencoders, and Deep Belief Networks and more. The approach proposed is of three steps. The stages include Data Augmentation, Hyperparameter Optimization: and Ensemble Learning. The ensemble was used to combine different models to improve detection accuracy, achieving 94.44% classification accuracy on KDD Cup 99 and 88.39% on NSL-KDD with low false-positive rates. The used datasets in this study are very old and may not be a true representative of current set of attacks in IoT scenarios.

3. Methodology

This study proposes to use a multi-stage approach in the application of machine learning algorithms for the identification of attacks in IoT-based Smart homes. The various stages in the proposed ML-based methods for attack identification in IoT are captured in Fig 1.

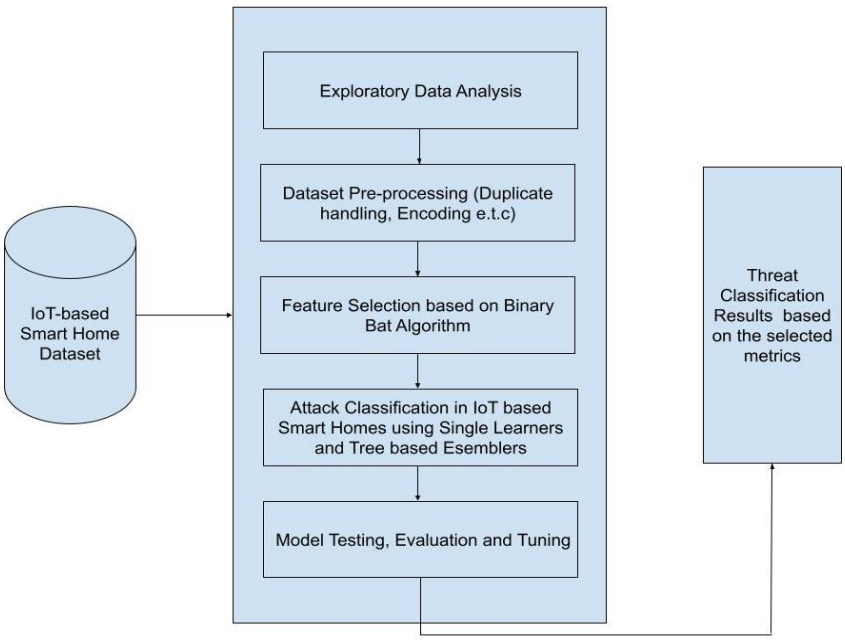

Fig. 1. Methodological Framework for Attack Identification in IoT-based Smart Homes

As shown in Figure 1, the key stages of building the classification models involve:

i. IoT-based Smart Home IDS dataset collection
ii. Exploratory Analysis and Visualisation of the Dataset
iii. Dataset Cleaning (Duplicate handling and Encoding) based on the issues identified in stage 2
iv. Feature Selection
v. Classification of Attacks in the dataset based on the chosen Tree and Non-Tree based Algorithms in the study
vi. Model Testing, Validation and Hyper Parameter Tuning.

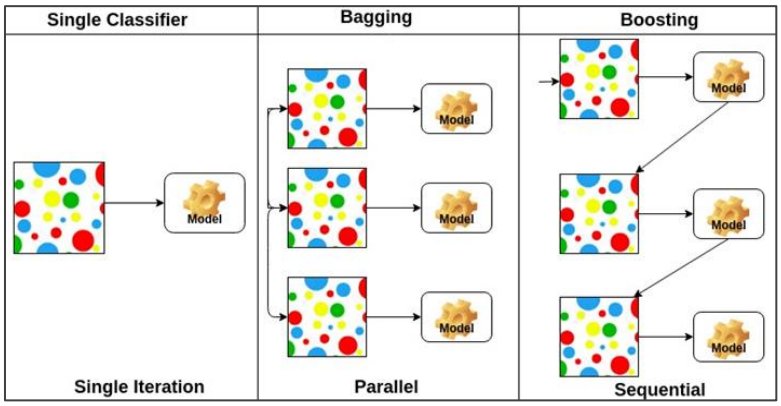

Fig. 2. Diagram Representing Models from Single and Ensemble Learners

Fig 2 is a representation of the grouping of learning algorithms as single as well as bagging and boosting that are classified as ensemble learners. The approach in this study is in line with the argument of [21] who demonstrated the importance of carrying out exploratory analysis in ML researches with a view to understanding patterns in security datasets while building better classification models.

3.1 Dataset Used in the Study

The dataset used in this study is available at https://www.kaggle.com/datasets/bobaaayoung/dataset-invade. The chosen dataset was released in the year 2024. The dataset is for IoT-based smart home scenarios and it contains 148516 instances and 24 rows (23 input attributes and the last being the target class). The choice of the dataset is because it contains features that are well suited for the identification of Cybersecurity threats in IoT-based Smart home scenarios. [22] argued that the dataset is a comprehensive one that consists of features for identifying Cybersecurity Threats in IoT Environments.

3.2 Data Labeling in Machine Learning Classification Problem

The problem being focused on in this study is a supervised classification problem. In the mathematical notations used for supervised and unsupervised learning is as shown in equations 1 and 2 respectively.

Labeled dataset

$$D: X = \{x^{(n)} \in R^d\}_{n=1}^N,\ Y = \{y^{(n)} \in R^d\}_{n=1}^N \quad (1)$$

Unlabelled dataset

$$D: X = \{x^{(n)} \in R^d\}_{n=1}^N\} \quad (2)$$

3.3 Key Activities at the Data Pre-processing Stage

Based on the issues identified in the dataset, the duplicate values were removed, and the categorical attributes were encoded. Label encoding was used for the three categorical input variables as well as the target class named "attack". This study equally selectively encoded the categorical data types in the dataset.

3.4 Binary Bat Algorithm for Feature Selection

There are different nature inspired algorithms that have been found useful in different real life problems [23, 24] and in the feature selection module of Machine Learning classification. [25] pointed out that Bat Algorithm (BA) was inspired by the echolocation property of microbats. The attack classification problem in this paper is a binary problem and thus we settled for Binary Bat Algorithm popularly called BBA. In building ML-based models, feature selection generally aims to find the most important features in a dataset, reduce training time, and build models that are less computationally expensive [26, 27].

[26] proposed a new nature-inspired feature selection technique that is based on bats behaviour, named binary bat algorithm. The binary Bat Algorithm is a well-known metaheuristic nature-inspired algorithm researched and developed by Dr. Xin-She Yang. The algorithm imitates the technique of echolocation that is employed by bats for perceiving their immediate surrounding environment and locating their prey and/or an obstacle. The BBA method used in this study works by associating a set of binary coordinates that denote whether a feature in the IoT-based Smart Home dataset belongs to the final set of features or not for each bat. In any given problem, the function to be maximized is the one given by a supervised classifier's accuracy.

The parameters that the BBA algorithm usually uses in searching activity are Position, Velocity, Frequency (of the bat), Pulse Rate, Initial Pulse Rate, Loudness, and Fitness.

3.5 Algorithms for the Attack Identification

The chosen algorithms in this study are:

i. Decision Tree (A Single Tree-based learner);
ii. K-Nearest Neighbors (KNN) (A Single non-tree based learner);
iii. Random Forest (A Tree-based ensemble); and
iv. AdaBoost (A Tree-based ensemble)

3.6 Selected Learning Algorithms

The first two learning algorithms used in this study are single classifiers while the last two are ensembles that are formed from the combinations of several single learners. In this study, the focus is on investigating how the different algorithms will behave when fed with the pre-processed dataset features.

A decision tree is a learning algorithm that uses a decision tree to make predictions. It follows a tree-like model of decisions and their possible consequences. The algorithm works by recursively splitting the data into subsets based on the most significant feature at each node of the tree. KNN is the type of ML algorithm useful for classification purposes. It classifies the data point on how its neighbour is classified.

The Random Forest algorithm was introduced by Leo Breiman as a learning algorithm that can be used for solving different classification and regression problems [28], Random forest works by averaging multiple deep decision trees, trained on different parts of the same training set, with the goal of reducing the variance. The AdaBoost algorithm was designed to boost weak learners by adjusting the weights of training samples [29]. It is a machine learning technique that combines multiple weak classifiers to create a strong classifier.

3.7 Dataset Test-Split Ratio Used

In each of the experimentations where the learning algorithms were used for the attack classification in the IoT-based Smart home dataset, eighty percent (80) and twenty (20) percent split ratio was used for the model validation.

3.8 Metrics for the Model Evaluation

Target Metrics used for the evaluation of the performances of the intrusion identification models are listed in equations 3, 4, 5, and 6.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{F1 score } = 2 \times \frac{\text{Prection} \times \text{Recall}}{\text{Prection} + \text{Recall}} \tag{6}$$

4. Results

4.1 Results of Exploratory Analysis

Based on the exploratory data analysis, the twenty-three input attributes and only target class in the dataset are: *duration, protocol_type, service, flag, src_bytes, dst_bytes, land, wrong_fragment, urgent, hot, logged_in, num_compromised, count, srv_count, serror_rate, rerror_rate, same_srv_rate, diff_srv_rate, srv_diff_host_rate, dst_host_count, dst_host_srv_count, dst_host_same_srv_rate, dst_host_diff_srv_rate, attack.*

The EDA results in Table 1 further revealed that there are mixed data types in the dataset. For instance, there are seven floating points, thirteen integer (int64) and four object (categorical) data types. The details are shown in Table 1.

Table 1. Features and Data Types in the Smart Home Dataset

#	Column	Non-Null Count	Dtype
0	duration	148517 non-null	int64
1	protocol_type	148517 non-null	object
2	service	148517 non-null	object
3	flag	148517 non-null	object
4	src_bytes	148517 non-null	int64
5	dst_bytes	148517 non-null	int64
6	land	148517 non-null	int64
7	wrong_fragment	148517 non-null	int64
8	urgent	148517 non-null	int64
9	hot	148517 non-null	int64
10	logged_in	148517 non-null	int64
11	num_compromised	148517 non-null	int64
12	count	148517 non-null	int64
13	srv_count	148517 non-null	int64
14	serror_rate	148517 non-null	float64
15	rerror_rate	148517 non-null	float64
16	same_srv_rate	148517 non-null	float64
17	diff_srv_rate	148517 non-null	float64
18	srv_diff_host_rate	148517 non-null	float64
19	dst_host_count	148517 non-null	int64
20	dst_host_srv_count	148517 non-null	int64
21	dst_host_same_srv_rate	148517 non-null	float64
22	dst_host_diff_srv_rate	148517 non-null	float64
23	attack	148517 non-null	object

The dataframe of the whole original dataset is shown in Table 2

Table 2. Dataframe of the whole Original Dataset

#	duration	protocol_ type	...	dst_host_ diff_srv_rate	attack
0	0	tcp	...	0.03	No
1	0	udp	...	0.60	No
2	0	tcp	...	0.05	Yes
3	0	tcp	...	0.00	No
4	0	tcp	...	0.00	No
...	...	...	...	...	
148512	0	tcp	...	0.06	No
148513	0	tcp	...	0.00	No
148514	0	tcp	...	0.00	Yes
148515	0	udp	...	0.01	No

The dataframe after the deletion of the duplicate values is captured in Table 3. After the deletion of the duplicate values, the dimension of the dataset is now 143760 by 24. The deletion is part of the data cleaning processes so as to make sure the data is in the best usable format for the learning algorithms.

Table 3. Dataframe of the duplicate values in the dataset

#	duration	protocol_ type	...	dst_host_ diff_srv_rate	attack
416	0	tcp	...	0.00	No
1786	0	tcp	...	0.00	No
2208	0	tcp	...	0.00	No
2332	0	tcp	...	0.00	No
2385	0	tcp	...	0.49	Yes
...	...	...	...	...	
148477	0	tcp	...	1.00	Yes
148488	0	tcp	...	0.00	Yes
148501	0	icmp	...	0.00	Yes
148511	0	icmp	...	0.00	Yes
148515	0	udp	...	0.01	No

The summary statistics of all the attributes in the intrusion dataset are as shown in Table 4
The statistics provided statistical information about each of the features in the dataset.

Table 4. Summary Statistics

	duration	Protocol _type	...	dst_host_ diff_srv_rate	attack
count	148517.0 00000	148517.0 00000	...	148517. 000000	148517.0 00000
mean	276.779305	1.055751	...	0.084103	0.481177
std	2460.683131	0.422304	...	0.194102	0.499647
min	0.000000	0.000000	...	0.000000	0.000000
25%	0.000000	1.000000	...	0.000000	0.000000
50%	0.000000	1.000000	...	0.020000	0.000000
75%	0.000000	1.000000	...	0.070000	1.000000
max	57715.0 00000	2.000000	...	1.000000	1.000000

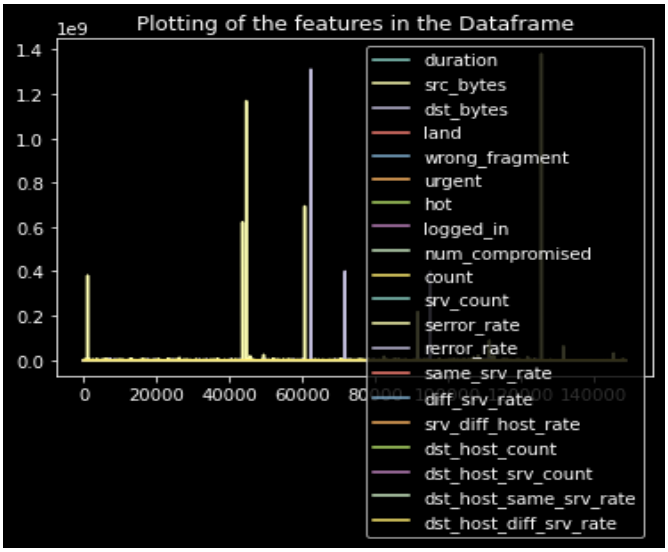

Fig. 3. Visualisation of the Features in the IoT Dataset

Fig 3 shows the pictorial representation of all the features in the IoT-based Smart Home dataset. The four categorical features were encoded and combined with the numerical ones for the attack classification model building and testing.

4.2 Non-Numeric Features in the Dataset were Encoded

As part of the steps in the dataset pre-processing the non-numeric features were encoded. This is to enable the chosen learning algorithms to make use of the intrusion classification in the Smart Home IoT environment.

4.3 Attribute Selection in the Dataset

At the attribute selection module, it was observed that the Binary Bat Algorithm helps in finding a better data representation in the dataset for better attack classification model building. The best of the features Nineteen (19) out of the original twenty-three (23) were used in building each of the models. The algorithm works by determining the best features at each of the iterations. The focus is to ensure that irrelevant and redundant features are not used for building the attack classification models.The feature selected were based on the working principles of the BBA algorithm.

4.4 Results of Attack Classification in IoT-based Smart Homes

The classification results of the selected algorithms in this study are as shown in Table 5.

Table 5. Classification Algorithm Performances

Learning Algorithm\Metric	Accuracy (%)	Precision	Recall	F1-Score
Decision Tree	98.600	0.987	0.987	0.987
K-Nearest Neighbors (KNN)	99.600	0.994	0.995	0.993
Random Forest	100.000	1.000	0.998	0.999
AdaBoost	99.910	0.997	0.998	0.997

5. Discussion of Findings

In this study, the pre-processing of the chosen dataset was strictly based on the problems identified with it during the Exploratory Data Analysis (EDA). For instance, the EDA revealed that there is no missing values in the dataset. It also revealed that there are duplicate values in the dataset and that the dataset contains mixed data types (numerical and categorical data). For this reason, this study handled the duplicate values as well as encoding the categorical features. The EDA equally showed that the dataset is very suitable for binary classification problem in security as the target class contains labels for attack and non-attack.

The features and data types in the Smart Home Dataset are captured in Table 1. The number of duplicate values in the dataset after the exploratory analysis was carried out is 4757. Thus, some records of the dimension 4757 by 24 were removed from the original dataset. The dataframe of the whole original dataset is shown in Table 2 while the summary of the duplicate values are as shown in table 3. The summary statistics of all the attributes in the intrusion dataset are as shown in Table 4. The classification results of the selected algorithms in this study are as shown in Table 5. The results in Table 5 showed that the EDA earlier carried out provided actionable insights on the kind of pre-processing steps that were applied on the IoT dataset.

Furthermore, it was observed that the use of BBA for optimal feature selection had impact on the enhancement of the classification performances of the four learning algorithms with the Random Forest having the highest classification results across the four metrics used for the evaluation. Train-Test Split method was used for the validation of each of the models each time. The Train test split is an ML model validation method where you can simulate how the model behaves when it is tested using new or untested data. Eighty percent was used as training set while twenty percent was employed as testing set in all scenarios for the maximal model performances. In each of the experimentations, the hyperparameters of the algorithms were tuned until the best results were obtained in each of the metrics used for the evaluation.

In this paper, the model that is rated second best in the classification of attacks is the AdaBoost. Specifically, the experimentations carried out revealed that all the algorithms achieved promising classification accuracy. For instance, Decision Trees recorded 98.60% accuracy, KNN produced 99.60% accuracy while Random Forest (RF) and AdaBoost-based models recorded 100.00% and 99.91% respectively. In all other metrics, RF-based attack classification model recorded the best results.

6. Conclusion

The paper first introduced the pervasiveness of attacks in general computer networks and IoT in particular. Then, the authors shifted attention to the importance of using the newly released IoT-based Smart Home dataset for intrusion investigation. In real-life scenarios, the models can be said to be efficient for the classification of unauthorised access, tampering and use of smart home resources or facilities. For this reason, the study focused on how to achieve improved classification of attacks based on the use of innovative ML approaches. The study described a multi-stage approach for building ML models for identifying attacks in the dataset. After the dataset acquisition, the following steps were followed in the study: initial exploratory data analysis to understand the dataset patterns, pre-processing of the dataset, selection of the most promising features using Binary bat algorithm and lastly the use of single and ensemble learners for the attack identification. Thereafter, we used two single and two ensemble learners to build the attack identification models from the dataset. It was observed that all the chosen learning algorithms achieved great performance based on the pre-processed data and selection of discriminant attributes with the use of BBA. It was, however, observed that in all scenarios, Random Forest (RF)-based model recorded the best overall performances in all the four metrics of Accuracy, Precision, Recall and F1-Score. It was also deduced that Adaboost-based model recorded the second best performances in all the metrics. All the models built in this study are less computationally intensive as attacks are readily classified from the dataset. This study argued that the performances of the four models are excellently promising when compared to recent similar studies.

Acknowledgement

The authors would like to thank all the anonymous reviewers who reviewed the manuscript and provided insightful feedback.

References

[1] O. Ibitoye, O. Shafiq, and A. Matrawy, " Analyzing adversarial attacks against deep learning for intrusion detection in IoT networks.," in 2019 IEEE global communications conference (GLOBECOM), 2019, pp. 1–6.

[2] Statista, "Annual number of Internet of Things (IoT) malware attacks worldwide from 2018 to 2022 (in millions)," 2023.

[3] A. Tasnim, N. Hossain, N. Parvin, S. Tabassum, R. Rahman, and M. I. Hossain, "Experimental analysis of classification for different internet of things (IoT) network attacks using machine learning and deep learning," in In 2022 International Conference on Decision Aid Sciences and Applications (DASA) (pp. 406-410). IEEE., IEEE, Mar. 2022, pp. 406–410.
[4] J. Alsamiri and K. Alsubhi, "Internet of things cyber-attacks detection using machine learning.," International Journal of Advanced Computer Science and Applications, vol. 10, no. 12, pp. 627–634, 2019.
[5] H. Touqeer, S. Zaman, R. Amin, M. Hussain, F. Al-Turjman, and M. Bilal, "Smart home security: challenges, issues and solutions at different IoT layers," Journal of Supercomputing, vol. 77, no. 12, pp. 14053–14089, Dec. 2021, doi: 10.1007/S11227-021-03825-1/METRICS.
[6] A. Chakraborty , M. Islam, F. Shahriyar, S. Islam, H. U. Zaman, & M. Hassan (2023). "Smart home system: a comprehensive review". Journal of Electrical and Computer Engineering, 2023(1), 7616683, 1-30, DOI:https://doi.org/10.1155/2023/7616683
[7] E. N. I. Bertino, "Botnets and internet of things security," Computer (Long Beach Calif), vol. 50, no. 2, pp. 76–79, Feb. 2017.
[8] A. Scroxton, "Global cyber spend to rise 15% in 2025, pushed along by AI," Computer Weekly, 2024.
[9] U. A. Adeniyi & A. M. Oyelakin, "A Survey on Promising Datasets and Recent Machine Learning Approaches for the Classification of Attacks in Internet of Things," Journal of Information Technology and Computing, vol. 4, no. 2, pp. 31–38, 2023.
[10] J. O. Olomu, A. M. Oyelakin, O. M. Ayinla, & H. A. Ibrahim, "A Review on Attack Landscape and Machine Learning Techniques for the Classification of Attacks in Internet of Medical Things (IoMT)," LAUTECH Journal of Computing and Informatics (LAUJCI), vol. 4, no. 1, Apr. 2024.
[11] Khraisat A., V. P. Gondal, and Kamruzzaman J., "Survey of intrusion detection systems: techniques, datasets and challenges," Cybersecurity, vol. 2, no. 1, 2019.
[12] Resende P.A.A. and Drummond A.C., "A survey of random forest based methods for intrusion detection systems," ACM Comput. Surv, vol. 51, no. 3, pp. 1–36, 2018.
[13] S. Mirjalili, S. M. Mirjalili, and X.-S. Yang, "Binary bat algorithm," Neural Comput Appl, vol. 25, no. 3–4, pp. 663–681, 2014.
[14] Milan Samantaray, Ram Chandra Barik, & Anil Kumar Biswal, "A comparative assessment of machine learning algorithms in the IoT-based network intrusion detection systems," Decision Analytics, vol. 11, no. 100478, 2024.
[15] B. Xu, Lei Sun, Xiuqing Mao, Ruiyang Ding, and Chengwei Liu, "IoT Intrusion Detection System Based on Machine Learning," Electronics (Basel), vol. 12, no. 20, p. 4289, 2023.
[16] Y. Majib , M. Alosaimi, A. Asaturyan & C. Perera (2023). Dataset for cyber–physical anomaly detection in smart homes. Frontiers in the Internet of Things, 2(1275080.), DOI: 10.3389/friot.2023.1275080
[17] A. A. Alsulami, Q. Abu Al-Haija, A. Tayeb, and A. Alqahtani, "An Intrusion Detection and Classification System for IoT Traffic with Improved Data Engineering. ," Appl. Sci., vol. 12, no. 12336, 2022.
[18] S. Riaz, S. Latif, S. M. Usman, S. S. Ullah, A. D. Algarni, and A. Yasin, "Malware detection in internet of things (IoT) devices using deep learning," Sensors, vol. 22, no. 23, p. 9305, 2022.
[19] F. Alghayadh and D. Debnath, "A Hybrid Intrusion Detection System for Smart Home Security Based on Machine Learning and User Behavior," Advances in Internet of Things, vol. 11, pp. 10–25, 2021.
[20] Maxime Labonne, "Anomaly-based network intrusion detection using machine learning," Institut Polytechnique de Paris, Paris, 2020.
[21] A. M. Oyelakin, A. O. Ameen, T. S. Ogundele , T. T. Salau-Ibrahim., U. T. Abdulrauf, H. I. Olufadi H.I,…, & Adeniji I. A., "Overview and Exploratory Analyses of CICIDS 2017 Intrusion Detection Dataset," ournal of Systems Engineering and Information Technology (JOSEIT), vol. 2, no. 2, pp. 45–52, 2023.
[22] Jacob, "Smart Home Intrusion Detection Dataset.", retrieved from https://www.kaggle.com/datasets/bobaaayoung/dataset-invade on 23rd September,2024
[23] Back T., Evolutionary algorithms in theory and practice. Oxford Univ.Press, 1996.
[24] I. Fister, X. S. Yang, S. Fong, and Y. Zhuang, "Bat algorithm: Recent advances," in 2014 IEEE 15th International symposium on computational intelligence and informatics (CINTI), Nov. 2014, pp. 163–166.
[25] X. Yang, "A New Metaheuristic Bat-Insspired Algorithm. Nature Inspired Cooperative Strategies for Optimization.," Studies in Computational Intelligence, pp. 65-74., 2010.
[26] R. Y. Nakamura, L. A. Pereira, K. A. Costa, D. Rodrigues, J. P. Papa, and X. S. Yang, "BBA: A Binary Bat Algorithm for Feature Selection.," in BBA: a binary bat algorithm for feature selection. In 2012 25th SIBGRAPI conference on graphics, patterns and images, IEEE, 2012, pp. 291–297.
[27] A. M. Oyelakin & R. G. Jimoh "A survey of feature extraction and feature selection techniques used in machine learning-based botnet detection schemes.," VAWKUM Transactions on Computer Sciences, vol. 9, no. 1, pp. 1–7, Sep. 2021.
[28] L. Breiman, "Random forests. ," Mach Learn, vol. 45, pp. 5-32., 2001.
[29] Y. Freund, "Boosting a weak learning algorithm by majority," Inf Comput, vol. 121, no. 2, pp. 256–285, 1995.

Authors' Profiles

Oyelakin A. M. is an academic, IT professional and technical author.He obtained National Diploma (Distinction Classification), B.Sc., M.Sc and PhD.in Computer Science. After graduation, he worked for some years in the IT industry in different capacities and later became a lecturer in the university on full-time basis. He has published over forty-six peer reviewed papers in journals and conference proceedings. His current areas of research interest are: Computer Networks, Cyber Security, Machine Learning, Intelligent Systems and Object Detection.

Sanni S. A. received his Masters and PhD in Information Science from the University of Malaya, Kwala Lumpur, Malaysia. He equally obtained his B.Sc.(Hons) Computer Science from University of Ilorin, Ilorin, Nigeria. He has to his credit several peer reviewed articles in journals and conference proceedings. His research focus is in the area of Information Sciences, Data Science and Machine Learning.

Adegbola I. A. holds a Bachelor's degree in Computer Science from the University of Ilorin (2006), a Master's degree in Computer Science from the University of Ibadan (2010), and a Ph.D. in Computer Science (2020), focusing on Spambot Detection using content classification. He is currently the Head of the Department of Computer Science at Oyo State College of Education, Lanlate, Nigeria.He has to his credit several peer reviewed articles in journals and conference proceedings. His research interests include IoT, Artificial Intelligence, Cybersecurity, and Community Informatics.

Salau-Ibrahim T. T. is a lecturer at the Department of Cyber Security, Federal University of Lafia, Nasarawa State. She received her Master's Degree in Computing and Information Systems from Queen Mary University, London and PhD from University of Ilorin, Nigeria. She has to her credit several peer reviewed articles in journals and conference proceedings. Her areas of interest are Artificial Immune Systems, Information Security and Cyber Security.

Bakare-Busari Z. M. is a promising university academic and currently lectures in the Department of Computer Science, Crescent University, Abeokuta, Nigeria. She is a First Class Graduate in the same department where she lectures and she is rounding off with her M.Sc. Programme at University of Ibadan, Ibadan, Nigeria. She is a dogged researcher and she is passionate about.continuous learning in ICT and research.

Saka B. A. received her bachelor's degree in Computer Science from University of Ilorin, Nigeria in 2005. She thereafter completed her master's degree in Computer science in 2009 from University of Ibadan, Nigeria. She is currently pursuing her PhD program and a lecturer in the Department of Computer Science at Crescent University, Abeokuta, Nigeria with over a decade teaching experience. Her research interest areas are: Machine Learning, Computational Intelligence and Security of Internet of Medical Things (IoMT).

How to cite this paper: Oyelakin A. M., Sanni S. A., Adegbola I. A., Salau-Ibrahim T. T., Bakare-Busari Z. M., Saka B. A., "Novel Machine Learning Approaches for Identifying Attacks in IoT-based Smart Home Environment", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.15, No.2, pp. 41-50, 2025. DOI:10.5815/ijwmt.2025.02.04