Journal of Robotics and Control (JRC)
Volume 5, Issue 6, 2024
ISSN: 2715-5072, DOI: 10.18196/jrc.v5i6.22819

Comprehensive Study on Detecting Multi-Class Classification of IoT Attack Using Machine Learning Methods

Tamara Zhukabayeva [1], Lazzat Zholshiyeva [2*], Khu Ven-Tsen [3], Aigul Adamova [4], Nurdaulet Karabayev [5], Erik Mardenov [6]
[1, 2, 3, 4, 5, 6] International Science Complex "ASTANA", Astana, Kazakhstan
[1, 2, 5] L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
[1, 4] Astana IT University, Astana, Kazakhstan
[3] M. Auezov South Kazakhstan University, Shymkent, Kazakhstan
[6] Astana International University, Astana, Kazakhstan
Email: [1] tamara_kokenovna@mail.ru, [2] lazzat.zhol.81@gmail.com, [3] qbcba@bk.ru, [4] aigul.adamova@astanait.edu.kz, [5] punpuruwu@gmail.com, [6] emardenov@gmail.com
*Corresponding Author

***Abstract*—The proliferation of IoT devices has heightened their susceptibility to cyberattacks, particularly botnets. Conventional security methods frequently prove inadequate because of the restricted processing capabilities of IoT devices. This paper suggests utilizing machine learning methods to enhance the detection of attacks in Internet of Things (IoT) environments. The paper presents a novel approach to detect different botnet assaults on IoT devices by utilizing ML methods such as XGBoost, Random Forest, LightGBM, and Decision Tree. These algorithms were examined using the N-BaIoT dataset to classify multi-class botnet attacks and were specifically designed to accommodate the limitations of IoT devices. The technique comprises the steps of data preparation, preprocessing, classifier training, and decision-making. The algorithms achieved high detection accuracy rates: XGBoost (99.18%), Random Forest (99.20%), LGBM (99.85%), and Decision Tree (99.17%). The LGBM model demonstrated exceptional performance. The incorporation of the attack evaluation model greatly enhanced the identification of botnets in IoT networks. The paper displays the efficacy of machine learning techniques in identifying botnet assaults in IoT networks. The models generated exhibit exceptional accuracy and can be seamlessly integrated into existing cybersecurity systems**.

Keywords—Wireless Sensor Networks; IoT; Identification Attacks; Machine Learning; Botnets.

I. INTRODUCTION

To share and acquire data, all network devices connect. Today, wireless sensor networks (WSNs) are widely used and have become an important tool in various areas of life [1]. However, as their use becomes more widespread, the risk of attacks that threaten data security and user privacy also increases. As a result, security and data aggregation technologies in WSNs are becoming increasingly relevant [2][3]. The main challenges in WSNs remain security and power consumption. To prevent attacks on networks and protect data during transmission, it is necessary to ensure reliable security. Data aggregation helps reduce the number of messages in the network, which in turn lowers overall energy consumption [4]. This is particularly important because extending the lifespan of WSNs depends on the correct choice of a data aggregation algorithm.

In the Internet of Things (IoT) domain, there is a wide variety of commercial and consumer devices, making them vulnerable to various security threats. One of the most significant concerns is the potential access to sensitive information collected and transmitted by these devices. As the number of IoT devices grows, so does the risk of attacks from malicious actors. The widespread adoption of IoT creates unique security challenges [5]. The large amount of data and the ubiquitous presence of devices attract hackers. One of the most dangerous threats is botnets, which can cause significant damage by rapidly spreading and infecting other devices [6]. Botnet attacks, such as DDoS and DoS, often go undetected due to their distributed nature and impact on multiple devices simultaneously [7][8][9]. Because there are not as many sensor devices available, DoS attacks against IoT apps typically have a big impact. The more IoT-connected devices are used, the more botnets turn and the more powerful they become. Combating botnets is an important challenge for IoT cybersecurity. The conventional security techniques, such as encryption and key management, exhibit notable constraints in IoT ecosystem. Primary factors contributing to the issue include limitations in device resources, the ever-changing nature of networks, challenges in scaling up, and inadequate adaptability in defending against botnets. These methods frequently encounter difficulties in effectively updating and adapting to new threats. On the other hand, machine learning (ML) technologies provide more flexible and responsive solutions. These systems have the ability to analyze network data in real-time, identify abnormalities, and adapt to changing assaults, making them better suited for protecting intricate and developing IoT networks [10][11]. ML algorithms offer versatile solutions and continuously improve their performance [12]. ML algorithms efficiently classify attacks and perform data aggregation and forwarding tasks to the receiving node. ML algorithms can also be applied to identify and classify botnet attacks based on device type and stage of attack, which helps to gain a more detailed understanding of attack characteristics [13].

A. Motivation and Research Questions

The reason for writing this paper is due to the growing concerns about the threats posed by IoT and the need for robust attack detection mechanisms [14][15][16]. As the IoT evolves, the vulnerabilities and potential for large-scale botnet intrusions become increasingly apparent [17]. ML models have shown significant potential in overcoming this challenge [18][19]. Nevertheless, there is a need to methodically review and summarize the existing literature to gain an understanding of current methodologies, their effectiveness, limitations, and future research directions.

The research questions formulated for this study as follows:

1. What types of botnet attacks are most common in IoT networks, and how do different ML methods compare in their effectiveness for multi-class classification of these attacks?
2. What are the limitations of existing methodologies for botnet detection in IoT environments, and how can ML algorithms be optimized for more accurate detection of attacks on IoT devices?
3. What research areas show potential for improving the effectiveness of ML strategies in detecting and preventing attacks?

The objective of the study is to evaluate the effectiveness of ML models in detecting multi-class classification attacks in IoT networks, which provides improved detection performance and reduced processing time for the data used. The study also aimed at comparing different methods at each step of the machine learning workflow, including the selection of meaningful feature subsets, the impact of separating training and test data on model performance, and the performance of three supervised ML classifiers in terms of accuracy, recall, and F1-score. In the conducted study, four ML algorithms were considered, for attack detection in IoT: XGBoost, LGBM, Decision Tree and Random Forest, A publicly available N-BaIoT dataset was used to evaluate the performance of the selected algorithms, which reflects a wide range of attack features and anomalies in IoT. Based on the obtained results, the proposed multi-class classification attack model is developed.

B. Contributions and Organization

Threats in IoT are everywhere and the spread of IoT usage is driven by the successful application of smart homes and cities around the world. However, IoT devices operate on public networks with limited processing power, storage capacity and bandwidth, which makes them more vulnerable to attacks compared to other endpoint devices. To address these challenges, this study includes the following contributions:

– We have developed a new effective methodology to identify different types of botnet attacks on IoT devices. This approach is based on the use of sequential architecture and machine learning algorithms such as XGBoost, Random Forest, Light GBM and Decision Tree. This model takes into account the features of resource-constrained IoT devices and incorporates data preprocessing mechanisms to improve the accuracy of multi-class classification.

– The study proposed an integrated attack evaluation model for multi-class classification. This model significantly improves the attack identification and classification process, enhancing the overall performance of botnet detection systems in IoT networks.

The paper consists of the following sections: Section I presents the introduction, contributions and organization, motivation and research question. Section II describes research methodology and search strategy where research questions, background methodology, botnet, ML models and a complete review of related works are discussed and disclosed. Section III deals with experimental results and proposed method, where the application of machine learning techniques XGBoost, Random Forest, LGBM and Decision Tree which detects various types of botnets with dataset, experimental comparison of the performance of machine learning algorithms and model evaluation of them and discussion are proposed. In addition, the last IV section describes the conclusion and future work. The overall structure of the paper can be seen in Fig. 1.

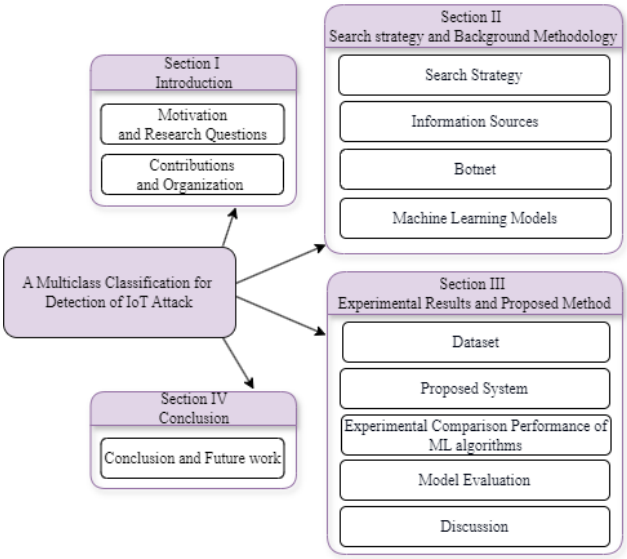

Fig. 1. Paper structure

II. Research Methodology and Search Strategy

This section describes the methodology and an overview of related research papers on botnet attack detection, and defence mechanisms against them using various security technologies and ML models for detecting attacks in IoT.

Researched the review articles with keywords from reliable research sources. A brief overview of botnets, IoT vulnerabilities, botnet malware, various methods to detect them, and the application of ML algorithms in the articles were considered.

A work based on research questions was conducted covering the metrics of IoT research released from 2020 to 2024, which investigates ML algorithms in the IoT domain, different types of attacks and attack models, methodologies, and evaluation criteria.

A. Search Keywords

To find answers to the research questions that were formulated during the study, the search queries utilized various keywords. By conducting the research, the keywords were combined with logical operators to create appropriate search queries that would help to get answers to the questions related to the topic under study:

RQ1- RQ4: (TITLE_ABS_KEY ("ML algorithms in IoT") OR TITLE_ABS_KEY(attack) AND TITLE_ABS_KEY (botnet) OR TITLE_ABS_KEY ("evaluation metrix and results") AND TITLE_ABS_KEY ("dataset in IoT") OR TITLE_ABS_KEY(methodology) OR TITLE_ABS_KEY (attack) AND TITLE_ABS_KEY (models) AND PUBYEAR > 2019 AND PUBYEAR < 2025 AND ("Mirai") OR ("IOT attacks") OR ("WSN Attack") OR ("Machine learning") OR ("WSN") OR ("DoS") OR ("BASHLITE") OR ("DDoS") OR ("Gafgyt") OR ("IoT")).

B. Information Sources

The research articles were selected from academic database sources such as Scopus, MDPI, Elsevier, Springer Link, IEEE Xplore, and Hindawi (Fig. 2).

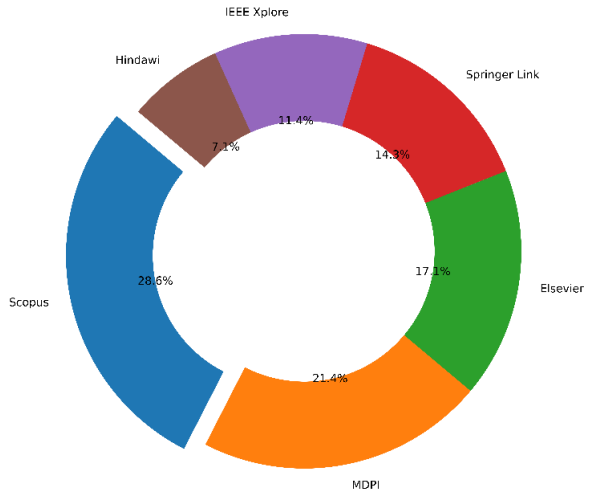

Fig. 2. Include papers from databases

From these databases, articles from journals and conferences from 2020-2024 were selected. Fig. 3 shows the distribution of included articles by year and the distribution of publications by identified topics.

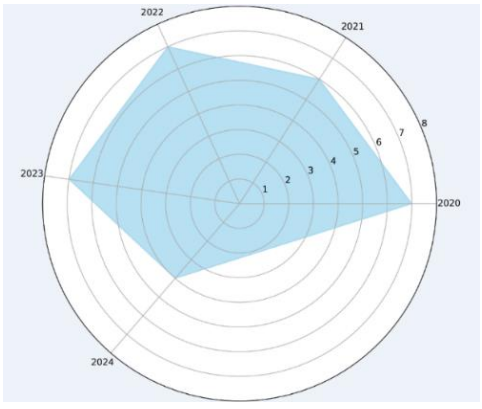

Fig. 3. Distribution of publication year

After researching the selected articles, papers focused on attacks in WSNs and IoT, ML methods, anomalies, and attacks are selected. Next, a selection of articles are selected by computer science and engineering fields. Fig. 4 illustrates the algorithm of article search and selection methodology.

From these databases, articles from journals and conferences from 2020-2024 were selected. Fig. 3 shows the distribution of included articles by year and the distribution of publications by identified topics.

As from Fig. 4, these articles were selected with our queries, resulting in 74 articles. To answer our research questions, we scrutinized all 74 primary studies. We examined the abstracts of the selected articles and filtered them according to our inclusion and exclusion criteria, double-checking the content of the articles if necessary. After analyzing the articles, 31 articles were included for further study. We extracted the following information for each study: full reference, short abstract and type of contribution, areas of application, integration with other testing methods, and evaluation details.

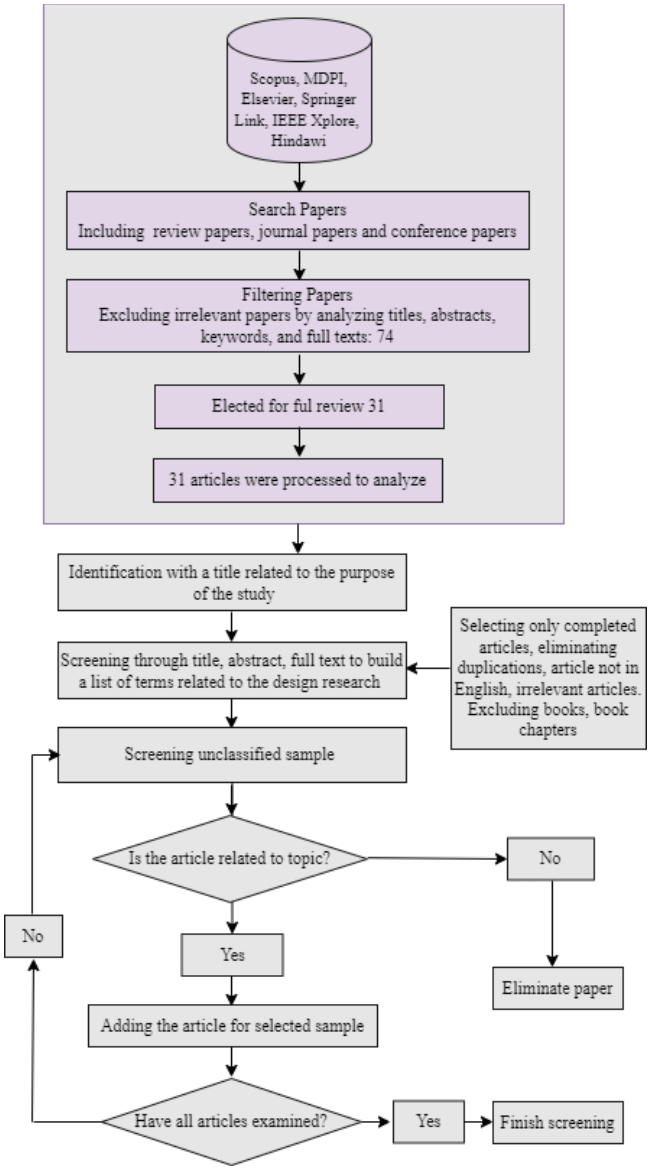

Fig. 4. Search methodology

We further analyzed the article abstracts, which show the general trend of sentiment in the IoT security research community and provide insight into how researchers perceive and discuss security issues related to IoT.

The Fig. 5 shows the evolution of average sentiment analysis scores obtained from annotations of IoT security publications from 2020 to 2024. According to Table 1, the subject of the statistical study is sentiment analysis in analyzing attacks in IoT. The investigation retrieved crucial data, such as complete references, types of contributions, areas of application, integration with other approaches, and evaluation specifics. Sentiment categorization was performed using ML models, and the findings were displayed in comparative tables to demonstrate the efficacy of various methods. Statistical analysis was used to quantitatively evaluate the trends revealed in the research and get insights into the perception and discussion of security risks in IoT. The analysis identified prevailing patterns and deficiencies in the current body of research, indicating potential avenues for future exploration. Utilizing charts and tables to visually portray facts significantly improves clarity and comprehension. By analyzing the publications, we found that various ML models are most commonly used to classify sentiment in IoT attacks. In the field of cybersecurity in IoT environments, the use of machine learning algorithms to detect breaches and intrusions has been the subject of numerous research studies. The results of the selected 31 articles are contained in Table I with methods and results of attack detection.

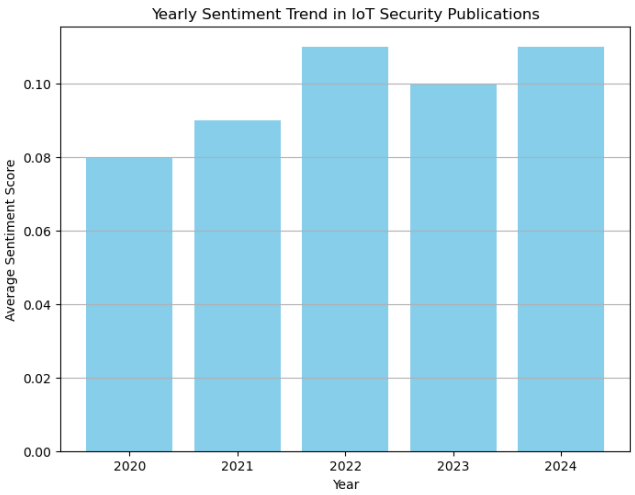

Fig. 5. Yearly sentiment trend in IoT security publications

Table I highlights the evolution of research in botnet detection, including their algorithms, and attack detection security. However, issues such as the dynamic nature of botnet attacks, the evolution of attack models, and the need for scalable detection mechanisms remain a challenge for future research. A review of 31 research papers on botnet detection highlights the approaches, methodology, and emerging trends in combating the botnet threat. IoT defense requires a multifaceted approach that includes deploying sophisticated mechanisms to detect anomalies and pinpoint intrusions. The research community has made significant strides in utilizing ML models to address these critical challenges (Fig. 6).

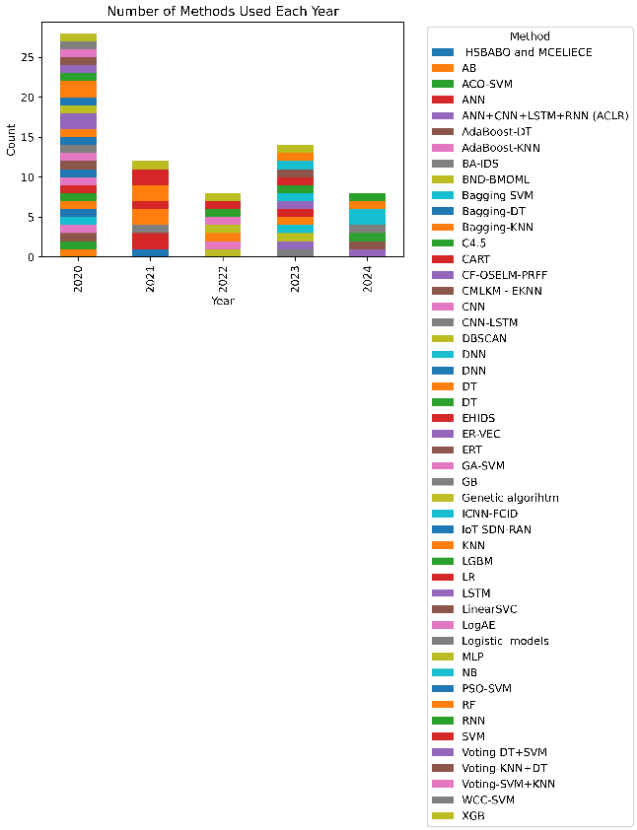

Fig. 6. Number of methods of each year

Authors [51] explore intrusion detection algorithms in WSNs and security concerns in the IoT. The main focus is on techniques for safeguarding data and assuring the dependability of networks. Additionally, it involves examining strategies for improving security measures and identifying potential dangers within these networks. Additionally, a systematic literature review to identify effective methods for detecting and preventing attacks on WSN. It includes an extended taxonomy covering attack types, datasets, detection methods, countermeasures, IDS is presented [52]. The use of ML discussed [53] that uses an ensemble of classifiers in eight variants that significantly improve the detection and prevention of botnet attacks, outperforming single classifiers in terms of accuracy. In a systematic review [54], the goal especially focuses on malware detection using permission analysis. The following studies discuss ML defence mechanisms [55][56][57], they also discuss ML algorithms by reducing the cost of securing WSN in several areas [58].

The authors [59] showed a comprehensive survey of ML algorithms used to support WSNs is provided considering WSN-specific constraints including security. The authors of [59] compared different ML algorithms in terms of anomaly detection. The research [60] used WSNs that utilize several ML methods are discussed. Among the machine learning methods used in practice, XGBoost is one of the most effective methods in many applications. The research [60][61] used a new algorithm considering data sparsity and weighted point sketch for approximate tree learning is described and proposed. An intrusion detection model based on XGBoost is proposed.

TABLE I. COMPARATIVE ANALYSIS OF IOT ATTACK DETECTION

Author	Year	Dataset	Method	Accuracy (%)
Bijalwan *et al.* [20]	2020	Botnet Dataset	AdaBoost-DT AdaBoost-KNN Bagging-DT Bagging-KNN Bagging SVM Voting-KNN+DT Voting DT+SVM Voting-SVM+KNN	98.36% 94.65% 95.30% 94.77% 75.99% 95.47% 85.06% 94.65%
Lakshmi Prasanna *et al.* [21]	2024	BoT-IoT training dataset	NB Logistic models RF KNN Multi-class Fast Parallel DT and Multi-Class Feature rank Gaussian Kernel	74% 75% 76% 79% 98%
Zhou *et al.* [22]	2020	NSL-KDD ARWID SIS-IDS2017	CFS-BA: Ensemble C4.5 RF Forest PA	99.8% 98.8% 99.1% 98.7%
Verma *et al.* [23]	2020	CIDDS-001 UNNSW-NB15 NSL-KDD	RF CART MLP AB GBM XGBoost ETC	94.94% 96.74% 82.76% 97.94% 99.53% 98.76% 82.99%
Sarwar *et al.* [24]	2023	IoTID20 MedBIoT UNSW-NB15 N-BaIoT	Extra Tree Random Voting Ensemble Classifier (ER-VEC)	99.99% 99.91% 95.64% 100%
Zixu *et al.* [25]	2022	UNSW BoT-IoT	GANs + AE	97.11%
Nadem *et al.* [26]	2021	NSL-KDD Selected sub-features of the dataset	SVM	95.98% 87.74%
Kumar *et al.* [27]	2021	NSL-KDD BoT-IoT DS2OS	RF KNN XGB	99%
Farahmand-Nejad *et al.* [28]	2020	N_BaIoT	WCC-SVM PSO-SVM GA-SVM ACO-SVM	95% 93 % 87% 88%
Tikekar *et al.* [29]	2024	Botnets	NB	90.62%
Liaqat *et al.* [30]	2020	Bot-IoT	CNN-cuDNN LSTM DNN-GRU LTSM-GRU	99.99% 99.96% 99.98%
Wani A. and Revathi S. [31]	2020	-	IoTSDN-RAN IoT-SVM	97.91% 97.48%
Huč *et al.* [32]	2021	DS2OS	DT	98%
Devprasad *et al.* [33]	2022	NSL KDD UNSW-NB15	DT SVM	98.77% 89.43%
Vishwakarma *et al.*[34]	2022	ToN-IoT	DNN	69.53%
Karthik *et al.* [35]	2021	-	HSBABO, MCELIECE	94%
Mohamed *et al.* [36]	2023	UNSW-NB15, ToN-IoT	EHIDS CF-OSELM-PRFF ABA-IDS ICNN-FCID	96.47%, 95,36% 94.70%, 91.31% 90.88%, 90.03% 92.38%, 92.26%
Awajan [37]	2023	Observed data	DNN, SVM	93.71%
Alrayes *et al.* [38]	2022	N_BaIoT	BND-BMOML	99.32%
Kim *et al.* [39]	2020	N_BaIoT	RNN, LSTM	97%
Sharma *et al.* [40]	2023	UNSW-NB15	CNN	84%
Rani *et al.* [41]	2023	DS2OS	LGBM-IDS	99,42%
ALMahadin *et al.* [42]	2022	UNBS-NB 15 and KDD99	SVM	99.62%
Mustafa *et al.* [43]	2023	N_BaIoT	DBSCAN	96.66%
Jain *et al.* [44]	2022	NSL-Botnet UNSW-NB15	LSTM	99.4% 93%
Çtin *et al.* [45]	2022	CICIDS2017	Genetic algorithm	91%
Raju *et al.* [46]	2023	CICIoT2023	DT	99.17%
Ali *et al.* [47]	2024	UNSW-NB15	ANN+CNN+LSTM+RNN (ACLR)	96.98%
Alkahtani Hasan *et al.* [48]	2021	N-BaIoT	CNN-LSTM	90.88%
Ullah *et al.* [49]	2021	BoT-IoT	CGANs + FNN	77.01%
Chu *et al.* [50]	2023	ToN-IoT	GANs	98.53%

The currently proposed botnet detection methods can be categorized based on the specific stage of work to be detected and the approach to detecting attacks. An anomaly detection autoencoder to protect nine IoT devices from botnet attacks is proposed. N-BaIoT is the first dataset used to build an autoencoder. BASHLITE (Gafgyt) and Mirai attacks are common botnets attacking IoT devices [62]. Collectively, the findings highlight the diversity of strategies, models, and technologies used in the research to counter botnet attacks, enhance the security of the IoT, and effectively address evolving cyber threats, especially in the areas of botnet detection and protection of IoT applications.

C. Botnet

Today, there is an active proliferation of botnets that identify potential IoT victims by scanning the network for open ports and subsequently infiltrating using exploits or weak credential leaks. Such attacks are characterized by their simplicity and ability to automatically propagate through the network by using worm-like mechanisms. This indicates a lack of user awareness of IoT security and that appropriate protection measures are not always in place. Mirai and Gafgyt have evolved into an entire family of botnets. Its diverse variants have reached the level of implementing distributed denial of service (DDoS) attacks, vulnerability scanning, command execution, and dynamic malware download and launch [63]. The Mirai Botnet is a network of compromised IoT devices that exploits security vulnerabilities caused by weak passwords. It is used to perform Distributed Denial of Service (DDoS) attacks by flooding target servers with excessive traffic [64]. In the Gafgyt network, administrators use its infrastructure to manage the range of attack directives provided by users, respond to queries, and facilitate collaborative discussions. The Gafgyt Botnet, like Mirai, possesses the capability to collect data and execute supplementary malicious activities. It exploits weaknesses in IoT devices and use propagation techniques like as SQL injections [65]. Gafgyt operates as an IoT attack, using several smart routers as both bot nodes and targets. Typically, once infected, the IoT device hosting Gafgyt initiates a network-wide scan to identify the responding nodes and then attempts to breach their defences by password mining or exploiting vulnerabilities. This modus operandi facilitates the spread of the botnet as infected devices are transformed into additional bot nodes, which aids in its propagation. Notably, Gafgyt favours smart routers among IoT devices due to their ubiquitous presence, extensive vulnerability landscape, and weak management practices [66].

According to SonicWall's mid-year update of its 2023 Cyber Threat Report, global IoT malware grew 37% in the first six months of 2023 [67]. The biggest culprits are the Mirai, NyaDrop and Gafgyt botnets. These malware families still account for 66% of the attack payload, creating botnets from infected IoT devices. The research also shows that cybercriminals are targeting outdated vulnerabilities: 34 of the 39 most popular IoT exploits specifically target vulnerabilities that are more than three years old. Geographic differences in IoT malware attacks. The landscape of IoT malware attacks in 2023 shows significant geographic variation. While North America saw a moderate decline in attacks, regions such as Asia and Latin America saw significant growth. This uneven distribution emphasizes the different levels of vulnerability in different regions. In countries with rapid digitalization, especially in Asia and Latin America, the IoT is being adopted at a pace that outpaces the development of appropriate cybersecurity measures. Regions with less developed cybersecurity systems are more susceptible to attack.

In December 2022, the number of IoT attacks worldwide exceeded 10.54 million (Fig. 7) [67]. Nevertheless, in the corresponding month of 2021, the quantity of documented IoT assaults decreased to nearly six million. The peak monthly attack volume was registered in June 2022, reaching almost 13 million attacks.

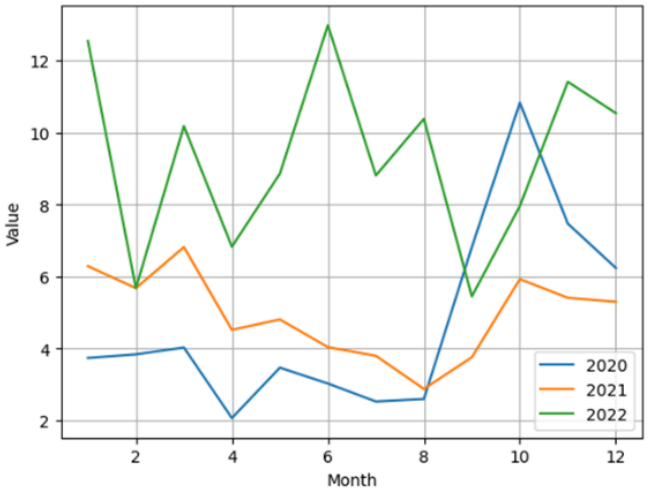

Fig. 7. Number Global IoT attacks 2020-2022 years [52]

D. Machine Learning Models

For the research on detecting IoT botnets, four ML algorithms: XGBoost, RF, LGBM and DT were chosen based on their established efficacy and capacity to adapt to the unique difficulties present in IoT settings. Given the over 10.54 million IoT attacks globally in December 2022, choosing these ML methods is essential. Effective detection techniques are needed to address the diverse IoT devices and their vulnerability to various attacks, including complex botnets.

E. XGBoost Algorithm

XGBoost (eXtreme gradient boosting) is an optimized model that can run on its own as a standalone algorithm; however, it has several features that outperform other algorithms. One of the features is regularization, which is used to prevent overfitting and enhance the generalization of the model. This feature is useful for dealing with large datasets and high-dimensional spaces. It incorporates cross-validation without the use of external libraries. Because of this capability, it is possible to stop at an early stage preemptively. XGBoost is renowned for its rapidity, scalability, and precision, incorporating inherent regularization to mitigate overfitting, rendering it well-suited for IoT datasets with a high number of dimensions [68]. The XGBoost is very accurate, quick, and versatile; it may be used for a variety of tasks, including classification issues. This algorithm minimizes computing time and enhances the gradient-boosting method of determining the objective

function [69]. Big data research challenges are answered efficiently and precisely during the training phase thanks to parallel computing [70]. Applying, the enhanced XGBoost model offers the best balance between performance and training time [71]. XGBoost was able to minimize the regularized objective function (L1 and L2). The training process uses iterative methods where new trees are added to predict the errors and residuals of previous trees and then combined with the previous trees to produce the final predictions. This method is called gradient boosting because it uses the gradient descent algorithm to minimize the loss when adding new models.

F. LGBM Algorithm

LightGBM (LGBM) is a framework from Microsoft, the main advantage of which is the speed of training on large data sets. It is based, as in the case of CatBoost'a and XGBoost'a, on the algorithm of gradient-based decision tree bousting. The LGBM algorithm is specifically designed to be efficient and fast, particularly when dealing with huge datasets. It achieves this by employing a histogram-based method, which helps reduce memory consumption and training time [72]. The preparation of data for predictions with LGBM is done in the same way as for the linear regression model for XGBoost. LGBM uses a novel gradient-based one-way sampling (GOSS) technique to filter the data instances to find the separation value, while XGBoost uses a pre-sorted algorithm and a histogram-based algorithm to compute the best separation. LGBM is another efficient machine learning algorithm which is also used for classification and regression tasks, similar to XGBoost but with some differences in architecture and speed [73].

G. RF Algorithm

The essence of the random forest (RF) method is to apply a set (ensemble) of decision trees (DT), each of which individually gives a residually low quality of classification, but in the aggregate due to their large number a higher result is obtained. It achieves high accuracy by using many decision trees and is particularly successful when dealing with noisy IoT data [74]. This method is used for classification tasks, in which case a decision is made by majority voting, and in regression, the answers of trees are averaged. The RF method is based on the so-called wisdom of crowds. The performance of a random forest is determined by the following rule: "A large number of relatively uncorrelated trees working together will outperform any of their components" [75]. Some of the trees may be incorrect, but the majority will be correct and as a result, the population of trees may follow the correct direction. The prerequisites for successful prediction can be considered to be some meaningful signal in the features (so that the models are more accurate than random guessing), and a weak correlation between the predictions (and errors) of individual trees.

H. DT Algorithm

DT is a simple yet powerful method, Decision Trees offer interpretable models that are easy to visualize. They are effective for classification tasks and handle both numerical and categorical data [76]. They are effective for classification tasks and handle both numerical and categorical data. They are effective for classification tasks and handle both numerical and categorical data. The DT method is based on the process of recursive partitioning of the initial set of objects into subsets previously assigned to the specified classes. Decision rules are used to perform the partitioning, and attribute values are checked according to a given condition. There are two main elements of the structure - nodes and leaves. Nodes contain decisive rules and subsets of observations satisfying them. Leaves contain observations classified by the tree. Each leaf belongs to one of the classes and the object is assigned the corresponding class label. The nodes specify the rules that partition the observations it contains, and the leaves are in turn labeled with the class label of the class whose objects fall into that leaf. If the class defined by the tree matches the target class, the object is recognized, otherwise it is unrecognized. The topmost node is called the root node, it contains all training and working datasets. The DT is a linear classifier; objects are partitioned in two-dimensional space by lines (in multidimensional space - by planes) [77].

III. EXPERIMANTAL RESULTS

This section provides an explanation of the confusion matrix and the evaluation metrics used for comparison and describes the results of the experiment. This is followed by a discussion of the results.

Throughout the model-building and testing procedures, the system used in the experiments remained unchanged. The model was trained and evaluated using ML algorithms XGBoost, RF, LGBM and DT. The multi-calss classification attack detection accuracy rate was used as a benchmark metric for evaluating the algorithm preference.

A. Dataset

The use of the N-BaIoT in this study is due to the difference in real data. The dataset [78][79] was used to classify: benign, g-jank, g-combo, g-scan, g-tcp and g-udp (Fig. 8).

This dataset is designed to address the lack of published botnet datasets for IoT. It used real data collected from 7.5 GB datasets for different types of common Internet of Things devices. The dataset characteristics are described in Table II.

TABLE II. CHARACTERISTICS OF DATASET

Name	Characteristics	Name	Characteristics	Name
Benign (Class_1)	A secure class, designated as Class_1, covers network traffic devoid of any malicious intent or action	Benign (Class_1)	A secure class, designated as Class_1, covers network traffic devoid of any malicious intent or action	Benign (Class_1)

Trained and optimized a deep autoencoder on 2/3 of its robust training dataset. This was done to track common network traffic patterns. Each device's test data included the remaining 1/3 of the secure data plus all malicious data.

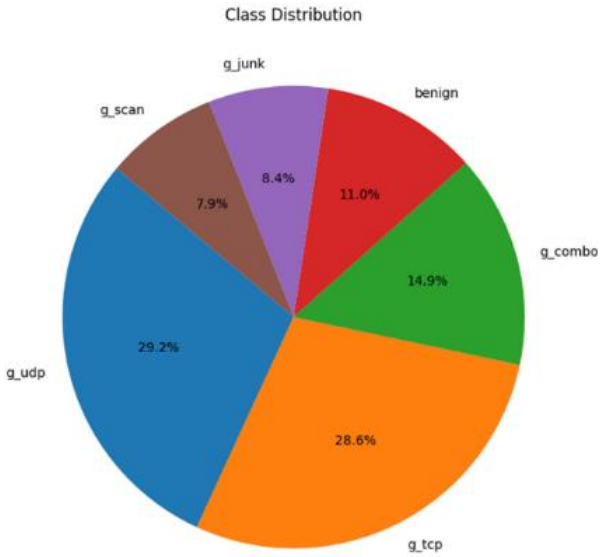

Fig. 8. Botnet attacks class distribution diagram

B. Proposed Model

This section details the development of the detection model designed to recognize attack behavior. The design phase consists of several sub-phases, and the data used in this research is in the public domain and can be retrieved from the dataset. This dataset specifically targets the scarcity of publicly available datasets on botnets, particularly those related to the IoT. The data provided is derived from 9 commercially available IoT devices that were genuinely infected by the Mirai and BASHLITE malware, resulting in actual traffic data [79]. Fig. 9 depicts an algorithm describing the complete attack detection process.

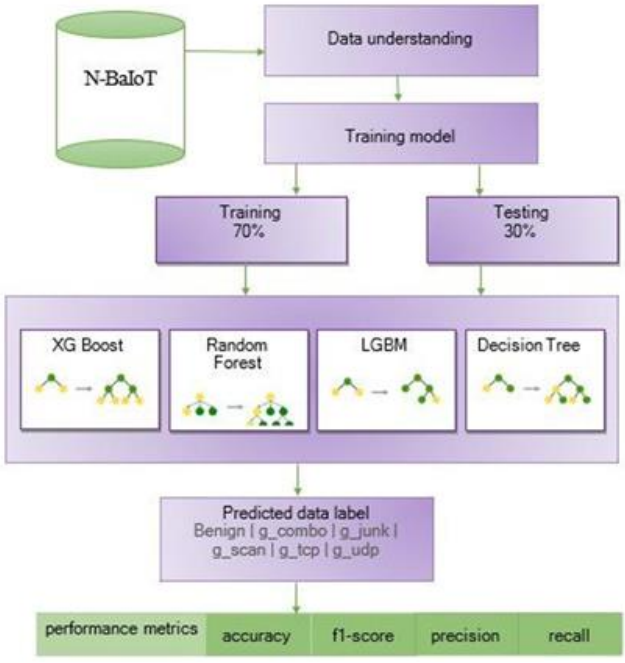

Fig. 9. Proposed model

The main objective is to develop an intrusion detection model capable of detecting attacks in IoT-based intelligent environments. The model includes three steps: data preparation and preprocessing, classifier training, and decision-making. All processes are performed based on a practical dataset. An illustration of the operation of the proposed multi-classification detection model is shown in Fig. 9. We presented a new efficient model to detect different types of botnet attacks on IoT devices. This method is based on sequential structure and application of machine learning algorithms such as XGBoost, Random Forest, LGBM and Decision Tree.

The attribute "type of attack" was chosen as the target variable; accordingly, the other attributes will act as independent variables (predictors). Before training the models, the data set was divided into two samples: training and test. The first sample is for training the classification models and the second sample is for evaluating the quality of performance of the classification models. The method "sklearn.model_selection.train_test.split()" was used to split the data, taking as parameters the dependent and independent variables, also the size of the test sample. The model is tested on training and test samples of 70% and 30% respectively.

C. Experimental Comparison of Performance of Machine Learning Algorithms

During model training and testing, a confusion matrix was created for each set of devices for testing and validation. The estimation of the threat detection rate can be effectively represented using the confusion matrix. Fig. 10 shows the confusion matrix, which estimates the classification accuracy by dividing the total number of observations by the predicted and actual values. It identifies model defects: vertical columns represent predictions and horizontal rows represent actual data.

Fig. 10 shows the confusion matrix for multi-class classification detection performed by XGBoost, RF, LGBM, and DT algorithms. Demonstrates the best performance in multi-class classification detection achieving an accuracy of 99.18% for XGBoost, 99.20% for RF, 99.85% for LGBM and 99.17% for DT. LGBM performed the highest in all other metrics.

D. Model Evaluation

In this experimental study, we evaluate four algorithms for multi-class classification of attack detection. A high f1-score indicates that the model performs well in detecting intrusions and minimizing false alarms. Another key metric in this area is recall, which reflects the model's ability to detect all intrusion occurrences. A high recall indicates that the model finds almost all intrusions, even if this results in a certain number of false positives. It is critical to consider both of these metrics when detecting intrusions, as missing even a single intrusion can have serious consequences. Therefore, models in this area should aim for high values of both f1 and recall to find a balance between accuracy and completeness of intrusion detection.

The evaluation metrics are computed and shown in Fig. 11 to Fig. 14. In the field of intrusion detection, the f1 metric plays an important role in evaluating the overall performance of a model. It shows how successful the model is in recognizing real intrusions and reducing false alarms.

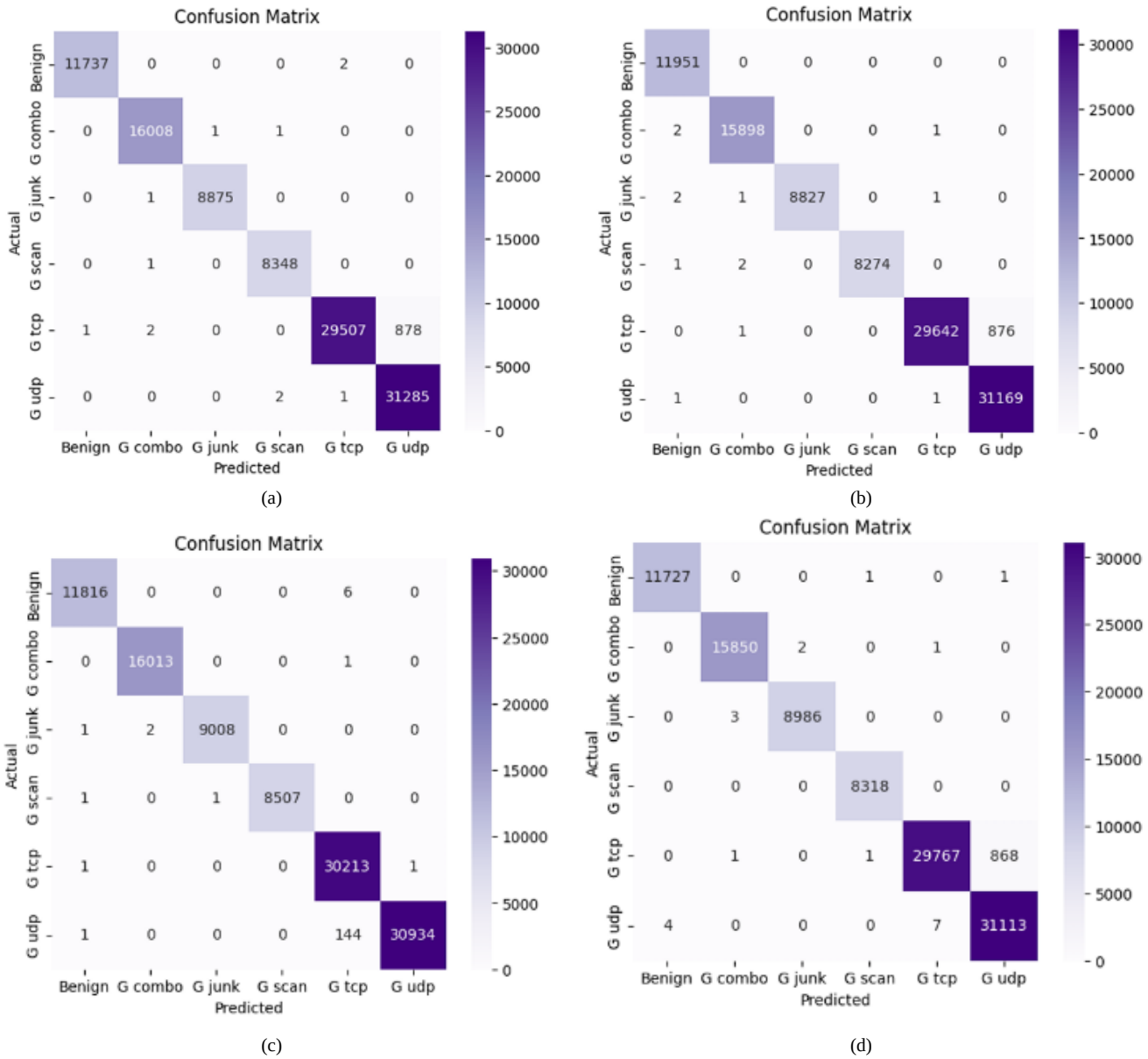

Fig. 10. Confusion matrix of XGBoost (a), RF (b), LGBM (c), DT (d)

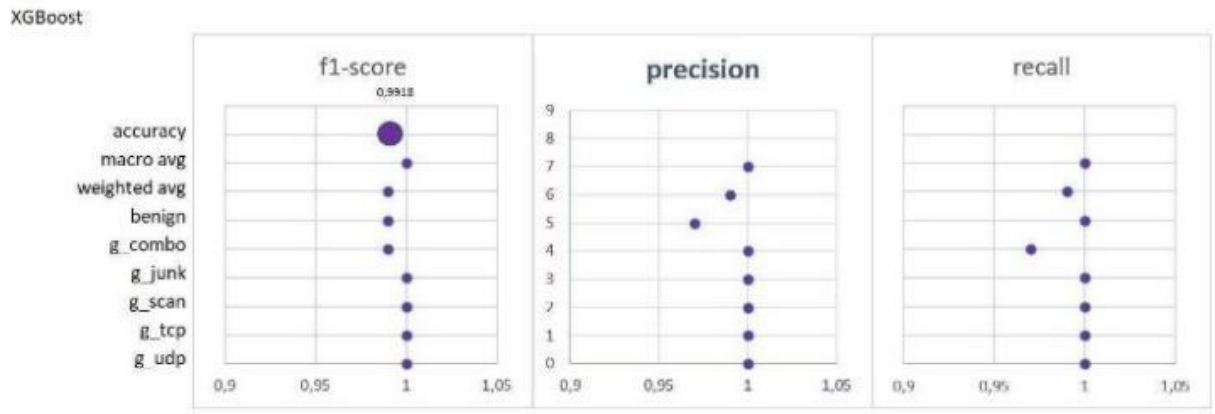

Fig. 11. Example of a figure caption

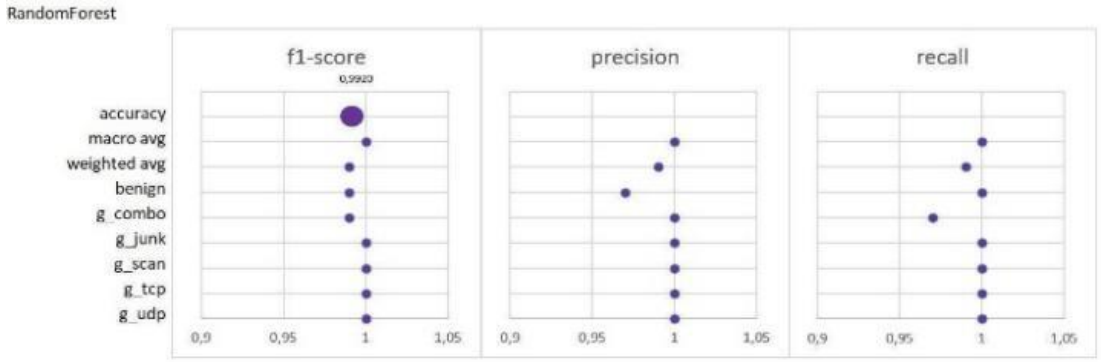

Fig. 12. Classification report of XGBoost

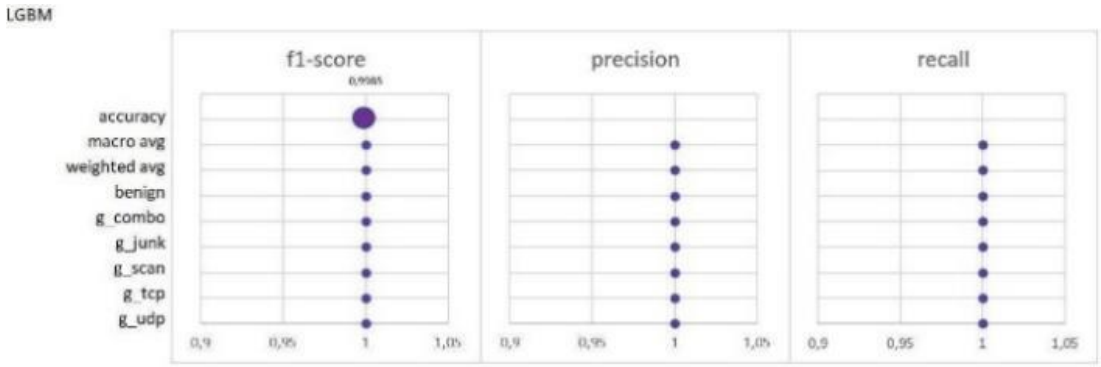

Fig. 13. Classification report of RF

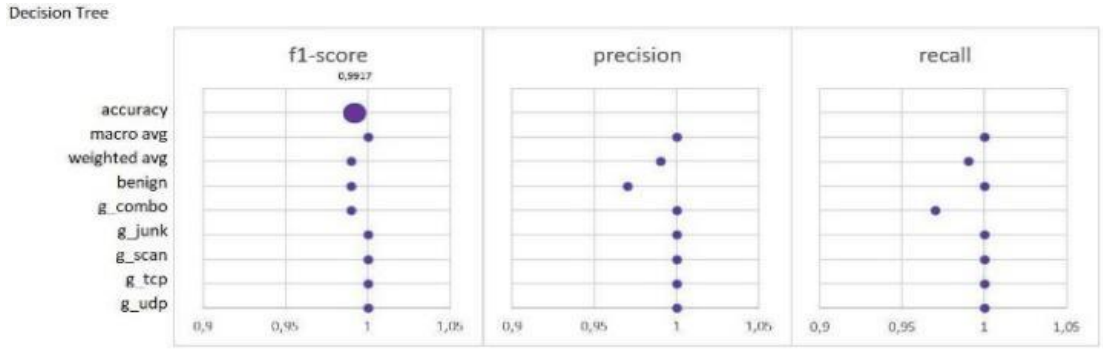

Fig. 14. Classification report of LGBM

The precision-recall curves (Fig. 15 a-d) show the high performance of the trained classifiers. Fig. 15 shows several Precision-Recall curves, one for each class, with precision values on the y-axis and recall values on the x-axis. The area under each curve is annotated, with values ranging from 0.97 to 1.00, indicating high precision and recall for the respective classes.

Overall, judging from the Precision-Recall curves, the XGBoost, RF and LGBM models outperform the DT model in terms of accuracy and recall for multi-class classification detection.

E. Discussion

We used the N-BaIoT dataset containing real data collected from network-connected IoT devices infected by botnets such as Gafgyt (BASHLITE), as shown in Figure 8. The training and test datasets are separated in the ratio of 70% and 30% respectively. The structure of the proposed model for anomaly detection and feature extraction for multi-class classification detection is shown in Fig. 9. The dataset contains normal data and attack samples. The results of the study on N-BaIoT dataset shows about the effectiveness of the proposed methods in terms of accuracy.

In Fig. 16, we compare our proposed model results with other literature from Table I, which shows the comparative analysis of IoT attack detection, which utilizes a multi-class classification model. Fig. 16 shows higher accuracy compared to other models for botnet detection. This indicates the effectiveness and feasibility of using the proposed model for attack detection.

This approach is tailored to the unique constraints of IoT devices, ensuring efficient and effective detection. Data preprocessing enhances the accuracy in multiclass classification, whereas multiclass classification enables simultaneous detection of multiple types of attacks. Detection achieved an accuracy of 99.85% using the LightGBM algorithm. This underscores the efficacy of the model in detecting and classifying botnet attacks, which is crucial for enhancing IoT cybersecurity.

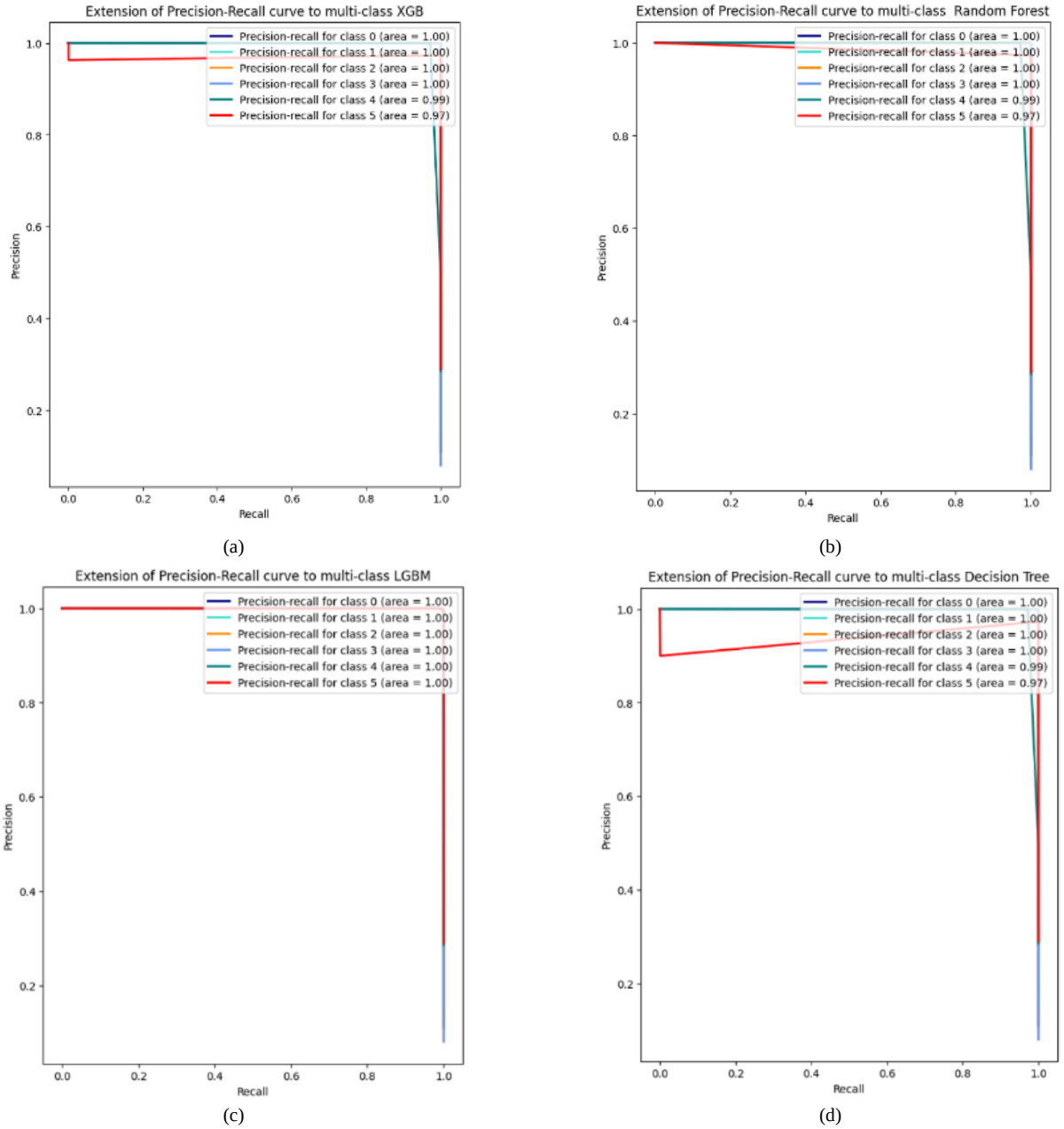

Fig. 15. Classification report of DT

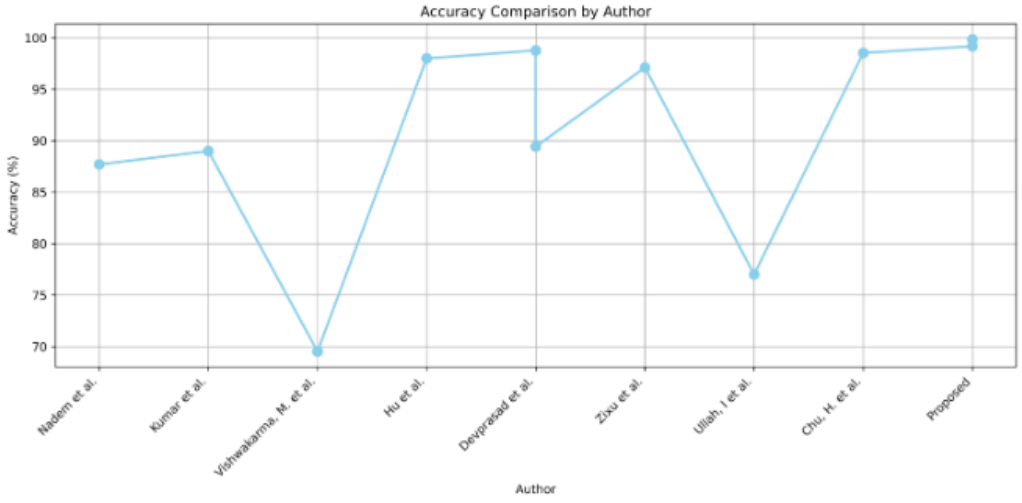

Fig. 16. Comparison of the effectiveness of multi-class classification attacks

IV. CONCLUSION

This paper presents the comprehensive study of botnet detection methodologies in IoT networks using ML methods. We have developed a new efficient methodology to identify different types of attacks on IoT. Our proposed model takes into account the features of IoT devices with limited resources and incorporates data preprocessing mechanisms to improve the accuracy of multi-class classification.

The proposed model results demonstrated the superiority of LGBM algorithms over other algorithms on multi-class classification of attack detection. XGBoost, Random Forest, LGBM and Decision Trees algorithms, showed high accuracy rates of 99.18%, 99.20%, 99.85% and 99.17% respectively.

Some of the main constraints are the performance may vary when using different datasets other than N-BaIoT. -Concentrating just on particular attack types may overlook growing risks. The limitations of resources in IoT devices are impacting their capacity to scale effectively. Future research should use adaptive algorithms to effectively address emerging threats. Evaluate the efficacy of the model in real-world situations and varied datasets [80]. The developed model classifiers exhibit a high degree of accuracy and are suitable for integration into complex attack detection systems. The practical applications of these results are significant: the models proposed here can contribute significantly to the architecture design of robust attack detection systems in IoT networks. However, despite the current advances in classification accuracy, further research is needed to further improve the performance and adaptability of these models to different environments and datasets. Future research is expected to explore the possibility of applying additional algorithms to improve the detection process.

ACKNOWLEDGMENT

This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. No AP19680345).

The authors would like to thank the "IoT Research Lab" for their valuable contributions to this study.

REFERENCES

[1] S. W. Nourildean, M. D. Hassib, and Y. A. Mohammed, "Wireless sensor network based on Internet of Things: a review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 1, p. 246, Jul. 2022, doi: 10.11591/ijeecs.v27.i1.pp246-261.

[2] J.-Y. Yu, E. Lee, S.-R. Oh, Y.-D. Seo, and Y.-G. Kim, "A Survey on Security Requirements for WSNs: Focusing on security-related characteristics," *IEEE Access*, vol. 8, pp. 45304-45324, 2020, doi: 10.1109/access.2020.2977778.

[3] N. Chandnani and K. N. Khairnar, "Analysis of architecture, structure, security and challenging aspects for data aggregation and routing techniques in IoT WSNs," *Theoretical Computer Science*, vol. 929, pp. 95-113, 2022, doi: 10.1016/j.tcs.2022.06.032.

[4] M. Kaur and A. Munjal, "Data aggregation algorithms for wireless sensor networks: A review," *Ad Hoc Networks*, vol. 100, p. 102083, Apr. 2020, doi: 10.1016/j.adhoc.2020.102083.

[5] T. Sasi, A. H. Lashkari, R. Lu, P. Xiong, and S. Iqbal, "A comprehensive survey on IoT attacks: Taxonomy, detection mechanisms and challenges," *Journal of Information and Intelligence*, vol. 2, no. 6, pp. 455-513, Dec. 2023, doi: 10.1016/j.jiixd.2023.12.001.

[6] H. Owen, J. Zarrin, and S. M. Poore, "Botnet overview, problems, threats, methods, detection and prevention," *Journal of Cybersecurity and Privacy*, vol. 2, no. 1, pp. 74-88, Feb. 2022, doi: 10.3390/jcp2010006.

[7] M. Asad, M. Asim, T. Javed, M. O. Beg, H. Mujtaba, and S. Abbas, "DeepDetect: Detection of Distributed Denial of Service Attacks Using Deep Learning," *The Computer Journal*, vol. 63, no. 7, pp. 983–994, Jul. 2019, doi: 10.1093/comjnl/bxz064.

[8] A. Tekerek, "A novel architecture for web-based attack detection using convolutional neural network," *Computers & Security*, vol. 100, p. 102096, Jan. 2021, doi: 10.1016/j.cose.2020.102096.

[9] V. Adat and B. B. Gupta, "Security in Internet of Things: issues, challenges, taxonomy, and architecture," *Telecommunication Systems*, vol. 67, no. 3, pp. 423–441, Jun. 2017, doi: 10.1007/s11235-017-0345-9.

[10] U. Tariq, I. Ahmed, A. K. Bashir, and K. Shaukat, "A Critical Cybersecurity Analysis and Future Research Directions for the Internet of Things: A Comprehensive Review," *Sensors*, vol. 23, no. 8, p. 4117, Apr. 2023, doi: 10.3390/s23084117.

[11] P. Sinha, V. K. Jha, A. K. Rai, and B. Bhushan, "Security vulnerabilities, attacks and countermeasures in wireless sensor networks at various layers of OSI reference model: A survey," 2017 *International Conference on Signal Processing and Communication (ICSPC)*, pp. 288-293, Jul. 2017, doi: 10.1109/cspc.2017.8305855.

[12] A. Ehsan, C. Catal, and A. Mishra, "Detecting Malware by Analyzing App Permissions on Android Platform: A Systematic Literature Review," *Sensors*, vol. 22, no. 20, p. 7928, Oct. 2022, doi: 10.3390/s22207928.

[13] A. Nazir et al., "Advancing IoT security: A systematic review of machine learning approaches for the detection of IoT botnets," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 10, p. 101820, Dec. 2023, doi: 10.1016/j.jksuci.2023.101820.

[14] M. Ghiasi, T. Niknam, Z. Wang, M. Mehrandezh, M. Dehghani, and N. Ghadimi, "A comprehensive review of cyber-attacks and defense mechanisms for improving security in smart grid energy systems: Past, present and future," *Electric Power Systems Research*, vol. 215, p. 108975, Feb. 2023, doi: 10.1016/j.epsr.2022.108975.

[15] U. I. Okoli, O. C. Obi, A. O. Adewusi, and T. O. Abrahams, "Machine Learning in Cybersecurity: A Review of Threat Detection and Defense Mechanisms," *World Journal of Advanced Research and Reviews*, vol. 21, no. 1, pp. 2286-2295, Jan. 2024, doi: 10.30574/wjarr.2024.21.1.0315.

[16] A. Yazdinejad, A. Dehghantanha, R. M. Parizi, G. Srivastava, and H. Karimipour, "Secure Intelligent Fuzzy Blockchain Framework: Effective threat detection in IoT networks," *Computers in Industry*, vol. 144, p. 103801, Jan. 2023, doi: 10.1016/j.compind.2022.103801.

[17] N. Mishra and S. Pandya, "Internet of Things applications, security challenges, attacks, intrusion detection and future prospects: A Systematic Review," *IEEE Access*, vol. 9, pp. 59353-59377, 2021, doi: 10.1109/access.2021.3073408.

[18] J. Kim, M. Shim, S. Hong, Y. Shin, and E. Choi, "Intelligent detection of IoT botnets using machine learning and deep learning," *Applied Science*, vol. 10, no. 19, pp. 7009, 2020, doi: 10.3390/app10197009.

[19] S. Khan and A. B. Maileva, "Botnet detection in IoT sensor networks: A lightweight deep learning system with hybrid self-organizing maps," *Microprocessors and Microsystems*, vol. 97, pp. 104753, 2023, doi: 10.1016/j.micpro.2022.104753.

[20] A. Bijalwan, "Botnet Forensic Analysis Using Machine Learning," *Security and Communication Networks*, vol. 2020, pp. 1–9, Feb. 2020, doi: 10.1155/2020/9302318.

[21] B. L. Prasanna and M. S. Reddy, "A Parallel Rank Based Multi-Class Ensemble Classification Framework on ISOT Cyber Threat Detection," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 10s, pp. 556–566, 2024.

[22] Y. Zhou, G. Cheng, S. Jiang, and M. Dai, "Building an efficient intrusion detection system based on feature selection and ensemble classifier," *Computer Networks*, vol. 174, p. 107247, Jun. 2020, doi: 10.1016/j.comnet.2020.107247.

[23] A. Verma and V. Ranga, "Machine Learning Based Intrusion Detection Systems for IoT Applications," *Wireless Personal Communications*, vol. 111, no. 4, pp. 2287–2310, 2019, doi: 10.1007/s11277-019-06986-8.

[24] A. Sarwar *et al*., "IoT networks attacks detection using multi-novel features and extra tree random - voting ensemble classifier (ER-VEC)," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 12, pp. 16637–16651, 2023, doi: 10.1007/s12652-023-04666-x.

[25] T. Zixu, K. S. K. Liyanage, and M. Gurusamy, "Generative Adversarial Network and Auto Encoder based Anomaly Detection in Distributed IoT Networks," *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1-7, 2020, doi: 10.1109/globecom42002.2020.9348244.

[26] P. Hadem, D. K. Saikia, and S. Moulik, "An SDN-based Intrusion Detection System using SVM with Selective Logging for IP Traceback," *Computer Networks*, vol. 191, p. 108015, May 2021, doi: 10.1016/j.comnet.2021.108015.

[27] P. Kumar, G. P. Gupta, and R. Tripathi, "Toward Design of an Intelligent Cyber Attack Detection System using Hybrid Feature Reduced Approach for IoT Networks," *Arabian Journal for Science and Engineering*, vol. 46, no. 4, pp. 3749–3778, Jan. 2021, doi: 10.1007/s13369-020-05181-3.

[28] A. Farahmand Nejad and S. Noferesti, "A real-time botnet detection model based on an efficient wrapper feature selection method," *International Journal of Security and Networks*, vol. 15, no. 1, p. 36, 2020, doi: 10.1504/ijsn.2020.10028190.

[29] P. C. Tikekar, S. S. Sherekar, and J. Kumar, "An Approach for Detection of Botnet Based on Machine Learning Classifier," *SN Computer Science*, vol. 5, no. 3, Mar. 2024, doi: 10.1007/s42979-024-02636-4.

[30] S. Liaqat, A. Akhunzada, F. S. Shaikh, A. Giannetsos, and M. A. Jan, "SDN orchestration to combat evolving cyber threats in the Internet of Medical Things (IoMT)," *Computer Communications*, vol. 160, pp. 697–705, Jul. 2020, doi: 10.1016/j.comcom.2020.07.006.

[31] A. Wani and S. Revathi, "Ransomware protection in IoT using software defined networking," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 3, p. 3166, Jun. 2020, doi: 10.11591/ijece.v10i3.pp3166-3175.

[32] A. Huč, J. Šalej, and M. Trebar, "Analysis of Machine Learning Algorithms for Anomaly Detection on Edge Devices," *Sensors*, vol. 21, no. 14, p. 4946, Jul. 2021, doi: 10.3390/s21144946.

[33] K. D. Devprasad, S. Ramanujam, and S. B. Rajendran, "Context adaptive ensemble classification mechanism with multi-criteria decision making for network intrusion detection," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 21, Jun. 2022, doi: 10.1002/cpe.7110.

[34] M. Vishwakarma and N. Kesswani, "DIDS: A Deep Neural Network based real-time Intrusion detection system for IoT," *Decision Analytics Journal*, vol. 5, p. 100142, Dec. 2022, doi: 10.1016/j.dajour.2022.100142.

[35] M. Karthik and M. Krishnan, "Securing an Internet of Things from Distributed Denial of Service and Mirai Botnet Attacks Using a Novel Hybrid Detection and Mitigation Mechanism," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 1, pp. 113–123, Feb. 2021, doi: 10.22266/ijies2021.0228.12.

[36] D. Mohamed and O. Ismael, "Enhancement of an IoT hybrid intrusion detection system based on fog-to-cloud computing," *Journal of Cloud Computing*, vol. 12, no. 1, Mar. 2023, doi: 10.1186/s13677-023-00420-y.

[37] A. Awajan, "A Novel Deep Learning-Based Intrusion Detection System for IoT Networks," *Computers*, vol. 12, no. 2, p. 34, Feb. 2023, doi: 10.3390/computers12020034.

[38] F. S. Alrayes et al., "Modeling of Botnet Detection Using Barnacles Mating Optimizer with Machine Learning Model for Internet of Things Environment," *Electronics*, vol. 11, no. 20, p. 3411, Oct. 2022, doi: 10.3390/electronics11203411.

[39] J. Kim, H. Won, M. Shim, S. Hong, and E. Choi, "Feature Analysis of IoT Botnet Attacks based on RNN and LSTM," *International Journal of Engineering Trends and Technology*, vol. 68, no. 4, pp. 43–47, Apr. 2020, doi: 10.14445/22315381/ijett-v68i4p208s.

[40] B. Sharma, L. Sharma, C. Lal, and S. Roy, "Anomaly based network intrusion detection for IoT attacks using deep learning technique," *Computers and Electrical Engineering*, vol. 107, p. 108626, Apr. 2023, doi: 10.1016/j.compeleceng.2023.108626.

[41] D. Rani, N. S. Gill, P. Gulia, F. Arena, and G. Pau, "Design of an Intrusion Detection Model for IoT-Enabled Smart Home," *IEEE Access*, vol. 11, pp. 52509-52526, 2023, doi: 10.1109/access.2023.3276863.

[42] G. Almahadin, M. O. Hiari, A. H. Hussein, N. M. M. Turab, A. Alkhresheh, and M. A. Al-Tarawneh, "Performance Evaluation of an Intelligent and Optimized Machine Learning Framework for Attack Detection," *International Journal of Communication Networks and Information Security*, vol. 14, no. 3, pp. 358-371, 2022.

[43] D. H. Mustafa and I. M. Husien, "Adaptive DBSCAN with Gray Wolf Optimizer for Botnet Detection," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 4, pp. 409–421, Aug. 2023, doi: 10.22266/ijies2023.0831.33.

[44] S. Jain, P. M. Pawar, and R. Muthalagu, "Hybrid Intelligent Intrusion Detection System for Internet of Things," *SSRN Electronic Journal*, vol. 8, p. 100030 2022, doi: 10.2139/ssrn.4097433.

[45] G. Çtin, "An Effective Classifier Model for Imbalanced Network Attack Data," *Computers, Materials & Continua*, vol. 73, no. 3, pp. 4519–4539, 2022, doi: 10.32604/cmc.2022.031734.

[46] V. S. A. Raju and B. Suma, "Network Intrusion Detection for IoT-Botnet Attacks Using ML Algorithms," *2023 7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, pp. 1-6, Nov. 2023, doi: 10.1109/csitss60515.2023.10334188.

[47] M. Ali, M. Shahroz, M. F. Mushtaq, S. Alfarhood, M. Safran, and I. Ashraf, "Hybrid Machine Learning Model for Efficient Botnet Attack Detection in IoT Environment," *IEEE Access*, vol. 12, pp. 40682–40699, 2024, doi: 10.1109/access.2024.3376400.

[48] H. Alkahtani and T. H. H. Aldhyani, "Botnet Attack Detection by Using CNN-LSTM Model for Internet of Things Applications," *Security and Communication Networks*, vol. 2021, pp. 1–23, Sept. 2021, doi: 10.1155/2021/3806459.

[49] I. Ullah and Q. H. Mahmoud, "A Framework for Anomaly Detection in IoT Networks Using Conditional Generative Adversarial Networks," *IEEE Access*, vol. 9, pp. 165907–165931, 2021, doi: 10.1109/access.2021.3132127.

[50] H.-C. Chu and Y.-J. Lin, "Improving the IoT Attack Classification Mechanism with Data Augmentation for Generative Adversarial Networks," *Applied Sciences*, vol. 13, no. 23, p. 12592, Nov. 2023, doi: 10.3390/app132312592.

[51] T. Zhukabayeva, A. Buja, and M. Pacolli, "Evaluating Security Mechanisms for Wireless Sensor Networks in IoT and IIoT," *Journal of Robotics and Control (JRC)*, vol. 5, no. 4, pp. 931-943, 2024, doi: 10.18196/jrc.v5i4.21683

[52] Zhukabaeva *et al*., "Towards robust security in WSNs: a comprehensive analytical review and future research directions," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 1, p. 318, Oct. 2024, doi: 10.11591/ijeecs.v36.i1.pp318-337.

[53] S. Afrifa, V. Varadarajan, P. Appiahene, T. Zhang, and E. A. Domfeh, "Ensemble Machine Learning Techniques for Accurate and Efficient Detection of Botnet Attacks in Connected Computers," *Eng.*, vol. 4, no. 1, pp. 650-664, Feb. 2023, doi: 10.3390/eng4010039.

[54] G. Kambourakis, C. Kolias, and A. Stavrou, "The Mirai botnet and the IoT Zombie Armies," *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*, pp. 267-272, Oct. 2017, doi: 10.1109/milcom.2017.8170867.

[55] R. Ahmad and I. Alsmadi, "Machine learning approaches to IoT security: A systematic literature review," *Internet of Things*, vol. 14, p. 100365, Jun. 2021, doi: 10.1016/j.iot.2021.100365.

[56] S. M. Tahsien, H. Karimipour, and P. Spachos, "Machine learning based solutions for security of Internet of Things (IoT): A survey," *Journal of Network and Computer Applications*, vol. 161, p. 102630, Jul. 2020, doi: 10.1016/j.jnca.2020.102630.

[57] R. Ahmad, R. Wazirali, and T. Abu-Ain, "Machine Learning for Wireless Sensor Networks Security: An Overview of Challenges and Issues," *Sensors*, vol. 22, no. 13, p. 4730, Jun. 2022, doi: 10.3390/s22134730.

[58] A. Mehta, J. K. Sandhu, and L. Sapra, "Machine Learning in Wireless Sensor Networks: A Retrospective," *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp. 328-331, Nov. 2020, doi: 10.1109/pdgc50313.2020.9315767.

[59] S. H. Haji and S. Y. Ameen, "Attack and Anomaly Detection in IoT Networks using Machine Learning Techniques: A Review," *Asian*

Journal of Research in Computer Science, pp. 30–46, Jun. 2021, doi: 10.9734/ajrcos/2021/v9i230218.

[60] M. Mamdouh, M. A. I. Elrukhsi, and A. Khattab, "Securing the Internet of Things and Wireless Sensor Networks via Machine Learning: A Survey," *2018 International Conference on Computer and Applications (ICCA)*, pp. 215-218, Aug. 2018, doi: 10.1109/comapp.2018.8460440.

[61] S. Ismail, D. W. Dawoud, and H. Reza, "Securing Wireless Sensor Networks Using Machine Learning and Blockchain: A Review," *Future Internet*, vol. 15, no. 6, p. 200, May 2023, doi: 10.3390/fi15060200

[62] M. Alqahtani, H. Mathkour, and M. M. Ben Ismail, "IoT Botnet Attack Detection Based on Optimized Extreme Gradient Boosting and Feature Selection," *Sensors*, vol. 20, no. 21, p. 6336, Nov. 2020, doi: 10.3390/s20216336.

[63] R. Hallman, J. Bryan, G. Palavicini, J. Divita, and J. Romero-Mariona, "IoDDoS—The Internet of Distributed Denial of Sevice Attacks - A Case Study of the Mirai Malware and IoT-Based Botnets," *Proceedings of the 2nd International Conference on Internet of Things, Big Data and Security*, pp. 47-58, 2017, doi: 10.5220/0006246600470058.

[64] M. Gelgi, Y. Guan, S. Arunachala, M. Samba Siva Rao, and N. Dragoni, "A systematic literature review on DDOS attacks on IoT botnets and evaluation of detection methods," *Sensors*, vol. 24, no. 11, p. 3571, Jun. 2024, doi: 10.3390/s24113571.

[65] P. Victor, A. H. Lashkari, R. Lu, T. Sasi, P. Xiong, and S. Iqbal, "IoT malware: Attribute-based taxonomy, detection mechanisms, and challenges," *Peer-to-Peer Networking and Applications*, vol. 16, no. 3, pp. 1380-1431, May 2023, doi: 10.1007/s12083-023-01478-w.

[66] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and Other Botnets," *Computer*, vol. 50, no. 7, pp. 80–84, 2017, doi: 10.1109/mc.2017.201.

[67] Monthly number of IoT attacks global 2022| Statista: Available: https://www.statista.com/statistics/1322216/worldwide-internet-of-things-attacks/

[68] F. Louati, F. B. Ktata, and I. Amous, "Enhancing Intrusion Detection Systems with Reinforcement Learning: A Comprehensive Survey of RL-based Approaches and Techniques," *SN Computer Science*, vol. 5, no. 6, Jun. 2024, doi: 10.1007/s42979-024-03001-1.41.

[69] T. Chen and C. Guestrin, "XGBoost," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, doi: 10.1145/2939672.2939785.

[70] T. Kim, L. F. Vecchietti, K. Choi, S. Lee, and D. Har, "Machine Learning for Advanced Wireless Sensor Networks: A Review," *IEEE Sensors Journal*, vol. 21, no. 11, pp. 12379–12397, Jun. 2021, doi: 10.1109/jsen.2020.3035846.

[71] S. Bhattacharya *et al.*, "A Novel PCA-Firefly Based XGBoost Classification Model for Intrusion Detection in Networks Using GPU," *Electronics*, vol. 9, no. 2, p. 219, Jan. 2020, doi: 10.3390/electronics9020219.

[72] M. Al-kasassbeh, M. A. Abbadi, and A. M. Al-Bustanji, "LightGBM Algorithm for Malware Detection," *Intelligent Computing*, pp. 391–403, 2020, doi: 10.1007/978-3-030-52243-8_28.

[73] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, "A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting," *Energy*, vol. 214, p. 118874, Jan. 2021, doi: 10.1016/j.energy.2020.118874.

[74] R. Bukhowah, A. Aljughaiman, and M. M. H. Rahman, "Detection of DoS Attacks for IoT in Information-Centric Networks Using Machine Learning: Opportunities, Challenges, and Future Research Directions," *Electronics*, vol. 13, no. 6, p. 1031, Mar. 2024, doi: 10.3390/electronics13061031.

[75] A. Parmar, R. Kataria, and V. Patel, "A review of Random Forest: An ensemble classifier," I*nternational Conference on Intelligent Communication Technologies and the Internet of Things (ICICI)*, pp. 758-763, 2018, Dec. 2018, doi: 10.1007/978-3-030-03146-6_86.

[76] V. G. Costa и C. E. Pedreira, "Recent advances in decision trees: an updated survey," *Artificial Intelligence Review*, vol. 56, no. 5, pp. 4765-4800, Oct. 2022, doi: 10.1007/s10462-022-10275-5.

[77] J. Su, S. He, and Y. Wu, "Features selection and prediction for IoT attacks," *High-Confidence Computing*, vol. 2, no. 2, p. 100047, Jun. 2022, doi: 10.1016/j.hcc.2021.100047.

[78] Kaggle, "N-BaIoT Dataset To Detect IoT Botnet Attacks," Kaggle, 2023, https://www.kaggle.com/datasets/mkashifn/nbaiot-dataset (accessed on 2 December 2023).

[79] Y. Meidan *et al.*, "N-BaIoT—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 12–22, Jul. 2018, doi: 10.1109/mprv.2018.03367731.

[80] H. Tanveer, M. A. Adam, M. A. Khan, M. A. Ali, and A. Shakoor, "Analyzing the performance and effectiveness of machine learning algorithms such as Deep Learning, Decision Trees, or Support Vector Machines on different datasets and applications," *Asian Bulletin of Big Data Management*, vol. 3, no. 2, pp. 126-136, Jan. 2024, doi: 10.62019/abbdm.v3i2.83.