A Lightweight GAN-Assisted Machine Learning Framework for Robust IoT Botnet Detection

[1]Arvind Yadav, [2]Prof.Sakshi Tiwari, [3]Dr. Mukesh Asati, [4]Prof. Rakesh Tiwari
[1]Research Scholar, Department of Computer Science Engineering, Technocrats institute of Technology & Science Bhopal (M.P.) India
[2,3,4] Professor, Department of Computer Science Engineering, Technocrats institute of Technology & Science Bhopal (M.P.) India
arvindup89@gmail.com , sakshitripathi0710@gmail.com , asati.mukesh@gmail.com , rakeshtiwari80@gmail.com

* Corresponding Author: Arvind Yadav

Abstract:

The rapid expansion in the use of Internet of Things (IoT) devices has made IoT networks prone to cyber-attacks, which are the primary actions of botnets like the Distributed Denial of Service (DDoS), malware distribution, and data extraction. The conventional ways with Intrusion Detection Systems that are based on signatures quickly become less effective when trying to fight botnet attacks, especially new kinds that appear and in the IoT situations where resources are limited. The light and smart IoT botnet detector that this research suggests is a collaboration of the two - the data enhancement of Generative Adversarial Network (GAN) and the machine learning classifier that is the ensemble-based.The first step of the proposed methodology is real-time collection of IoT traffic along with extensive data processing including duplicate removal, label encoding, KNN-based missing value inputting as well as feature normalization. To address the problem of class imbalance that is a common characteristic of IoT security data sets, GAN-based augmentation is used to produce minority-class attack samples artificially. This large dataset is then presented to an XGBoost-based classification model that performs accurate botnet detection by receiving training.A super detailed proposal is put up and the NLTK data set was used in the experimentation for the multi-class classification scenarios. The results of the experiment in the table are quite promising as it shows a 98.4% accuracy, 99.1% precision, 98.8% recall, and a 98.8% F1 score. There is a decrease in the number of false alarms and consistency in detection has improved which in turn show the robustness of the newly proposed approach. The use of GAN in the learning process has been proven through results that the detection capability is significantly increased, the framework is also computationally efficient, and hence appropriate for being used in the case of real-world IoT deployments.

Keywords:*IoT Security, Botnet Detection, Generative Adversarial Networks, Data Imbalance, XGBoost, Intrusion Detection System, Machine Learning*

I.INTRODUCTION

The rapid emergence of the Internet of Things has revolutionized modern digital ecosystems, combining billions of devices that allow seamless connections across diverse domains such as healthcare, smart cities, industrial automation, transportation, and energy management [1]. While Internet of Things technologies have enabled enhanced efficiency, automation, and data-driven decision-making, a notable fallout of the technology has been a proliferation of severe cyber-security challenges. Most of the IoT devices are resource-constrained, lacking robust security mechanisms; as a result, they frequently come established witho default configurations, thus presenting great appeal for cyber adversaries. Consequently, the IoT environments are very vulnerable to botnet attacks which pose not only a great threat to network availability and data integrity but also to the users' privacy [2].

Massive Botnet on IoT has hacked and remote-controlled all the equipment to hack in turn for a coordinated attack. The attacks included DDoS strikes, malware enforcement, and pulling of data and service disruption. Many noticeable attacks exemplify how IoT botnets can negatively influence infrastructure services or cloud services [3]. Services such as IDS clearly are starting to fail under the onslaught of zero-day botnet attacks that are complex and evolutionary. These systems depend on attack signatures and rule books stored fortuitously. These attack tools fail to deal with "dynamic and distributed threats" coming into play in modern-day IoT [4].

To address these limitations, machine learning (ML) and deep learning (DL) techniques have emerged as promising solutions for intrusion detection in IoT networks. ML-based IDS can categorize the anomalous activities that indicate botnet activities, as they have learned about patterns from historical and real-time network traffic. Still, few impediments have been positioned to be obstacles to their operational deployment in IoT surroundings [5]. These include class-imbalanced security datasets, high false alarm rates, computational overhead, low interpretability, and weak scalability in large and heterogeneous networks. Class imbalance, which typically occurs when malicious instances are outperformed

greatly by benign instances in a specific network, are one of the main cause disqualifying models in their function in the direction of detecting an opposite-class attack [6].

Existing research studies endeavor to classify the class unbalance anchored in the field of data augmentation, and there are also increasing interests in Generative Adversarial Networks (GANs). Particularly, GANs have succeeded in generating maliciously realistic examples so that minority-class representation is polished, thus enhancing detection accuracy and overall learning model. In collaboration with rigorous ensemble learning algorithms like Extreme Gradient Boosting (XGBoost), the intrusion detection framework can balance accuracy, precision, and computational efficiency. Moreover, XGBoost is ideal for use in IoT security problems because of its capacity to handle high-dimensional data, ability to resist overfitting through regularization, and deliver fast inference [7].

In this article, we propose a lightweight, smart IoT botnet detection framework that integrates GAN-based data augmentation and XGBoost-based classification algorithms [8]. The approach is prepared to overcome significant IoT security challenges, such as imbalanced data, false positives, and scalability, while also providing high detection capabilities. Extensive experimentation is performed on the N-BaIoT dataset, the standard benchmark dataset for IoT botnet detection. Results suggest that the proposed approach has a substantial gain in detection accuracy and false positives in relation to existing methods and can be thusly applied to real-world IoT deployments.

II. RELATED WORK

Deep learning and artificial intelligence techniques are becoming prominent in the use of intrusion detection systems for security in the IoT environment and IIoT. Deep learning-based intrusion detection frameworks show an unmatched ability of learning operational signatures from the sensor signals of an industry and achieves better accuracy in comparisons with traditional ways. Such systems, which are computationally extensive, could only be optimized under certain ceiling conditions such as IoT [1]. Comprehensive studies of innovative AI-based intrusion detection arrangements underscore the necessity to address the issues of training and implementation methods in order to document improvements in detection accuracy and reduced latency through adaptive learning mechanisms, and there also remains the problem of computational overhead under different IoT settings [2]. Real-time deep learning-based landmark anomaly detection systems have produced a very low false alarm rate, under high detection capabilities, but the stochastic nature of model complexity, for instance, inhibits their deployment at the edge [3]. An investigating incisive look into the invading features of machine learning in intrusion detection evinces that the coming up to adaptive and predictive models for modern network security is strong but is heavy on explain ability and use of energy in resource multiplication [4]. Deep learning algorithms have made the select group a prime number of credits, living with features and selections into feature selections, thereby proving the ultimate success [5]. Light AI-based intrusion detection and authentication systems specific to medical and sensor networks won a good spot concordance between high accuracy and low-latency operation and low operation overhead, whereas performance might differ in function of sensor network topologies and environmental deployment conditions [6]. One consequence of comparative evidence supports among the different feature selection and extraction procedures for machine learning-based intrusion detection systems is that no single method outperforms the others across all datasets, and therefore one should rely on adaptive and hybrid feature-engagement strategies [7]. Heavy at detecting and steering way through heavy traffic hours, network intrusion detectors crafted on deep learning received the recognition of high accuracy and almost no false positives, though their computational resource impositions make it nearly impossible to serve them in real time in a large IoT network [8]. Exhaustive reviews of machine-learning-based intrusion detection systems indicate that generally the ensemble theories and deep learning stand out better than the conventional phase and at the same time show that resource constraints and challenges in deployment may not be resolved [9]. The hybrid feature extraction with machine learning classifiers only supported the gain ratio, with reduced false positives and improved attack detection accuracy. Appreciation immediately turns towards very strong data quality. But in the long run, their success will depend on data quality and scalability—a criteria prompting strong interest in adaptive feature selection and full-fledged real-time deployment in IoT gadgets [10].
Recent advances have looked into using machine learning in developing numerous intrusion detection methods for IoT and network security enhancement. Automated intrusion detection techniques that use machine learning have been developed, which has resulted in improved anomaly detection accuracy due to model optimization and selection of hyperparameters from a variety of IoT datasets. Still, considerable investment in computational resources and time is needed for performance optimization and real-time deployment [11]. Machine learning-based intrusion detection systems in SDNs have shown tremendous improvement in the detection rate with respect to false alarms because of ensemble approaches, yet the major hindrance for real-time applications remains the computational complexity [12]. However, hybrid models that fuse optimization techniques with machine learning algorithms boast a middle ground of accuracy and processing speed. Scaringly, operation with large IoT environment always raise scalability issues [13]. AI-driven intrusion detection systems for smart grids are now becoming very well known, but there is still substantial skepticism about their integration with preexisting infrastructure [14]. Though the explainable AI-based ensemble intrusion detection frameworks have considerably enhanced transparency and comprehensibility of decision making and have maintained a high level of accuracy, their high computational overhead has so far constrained their deployment

innetworks with a large scale. Comprehensive reviews show that deep learning models are most potent when it comes to detecting complex and evolving threats, but the prohibitive computational and interpretational complexity keeps potential adoption at bay [16]. When fine-tuning such AI-enabled intrusion detection systems to make their models transparent, it became evident that those systems did not suffer significant degradation in terms of detection performance. However, generating the explanations was shown to require too much processing overhead to run in resource-poor IoT environments [17], [18]. The machine-learning-based intrusion detection systems for heterogeneous IoT environments have been seeing a high detection accuracy with respect to very low false-positive rates; however, scalability and adaptive-response mechanisms are yet to be attended to [19]. Class-based intrusion detection frameworks on wireless sensor networks, in addition to their improvements in anomaly detection performance, have put together heavy computational requirements, thus giving rise to lesser potential for large-scale deployments [20]. Zero trust with AI-integrated intrusion detection architectures for Industrial IoT have shown impressive anomaly detection and access control, although the more the network, the more the workload [21]. Deep reinforcement learning has delved deep into the intrusion detection system's open architecture and dynamics in preserving systems from threats of all kinds in a software-defined environment but seems irrelevant to real-time and edge deployment due to extensive computation [22]. Studies on data preprocessing have established its undeniable responsibility in increasing the accuracy for intrusion detection systems as well as bringing down the total number of false positives. However, the pre-processing pipeline limits its independence on the dataset [23]. Deep learning with statistical feature selection-based algorithms can offer very good detection performances for intrusion detection using low computational overhead given training data; nevertheless, their inability to learn quickly from new attack patterns is indeed a major concern [24]. Ensemble detection frameworks with explainable-AI have proven to be transparent, reliable, and well detectable, though these frameworks exhibit potential scalability and computational cost [25]. White-box and optimization-based intrusion detection systems, enjoying high detection accuracy and maintaining a low false positive rate, find applicability only in deployed IoT device weather resource-wise [26], [27]. At the gateway level, deep learning-based security systems have improved false negative detections by investments in computational resources, thus necessitating a rise in adaptive and scalable protection models [28]. The focus of energy-efficient intrusion detection frameworks has been to balance detection performance and resource usage. Nonetheless, model complexity is still a limitation to real-time processing in this area [29]. To sum up, optimization-driven deep learning intrusion detection approach for cyberphysical systems manage to gain a high accuracy rate with few false alarms. However, scalability and computationalcomplexity may be key emerging issues in research towards large-scale IoT environments [30].

Table 1: Related Work on AI/ML-Based Intrusion Detection Systems

Ref. No.	Focus Area	Methodology	Key Results	Limitations	Future Scope
[1]	Industrial IoT Security	Deep learning–based IDS for sensor networks	Improved anomaly detection accuracy over traditional methods	High computational resource requirements	Edge deployment and hybrid DL models
[2]	IoT IDS Deployment	AI-based IDS training and deployment strategies	Reduced detection latency and improved accuracy	High computational overhead	Scalable deployment in heterogeneous IoT
[3]	Real-Time IoT IDS	Deep learning–based real-time anomaly detection	Low false positives and reliable detection	Model complexity and poor interpretability	Lightweight and edge-enabled IDS
[4]	AI-Driven IDS	Adaptive ML and predictive analytics	Improved detection accuracy and threat prediction	Explainability and deployment overhead	Explainable AI-integrated IDS
[5]	IoT Network Security	Hybrid AI (DL, ensemble learning, feature selection)	Enhanced robustness and detection performance	Scalability and resource constraints	Adaptive hybrid AI frameworks
[6]	Medical IoT Security	Lightweight AI-based IDS with authentication	High accuracy with low latency and overhead	Sensor topology dependency	Scalable edge AI deployment
[7]	Feature Engineering	Feature selection and extraction comparison	No single method optimal across datasets	Dataset dependency	Adaptive and hybrid feature engineering
[8]	Network Traffic Analysis	Deep learning IDS focusing on traffic behavior	High accuracy and low false detection rates	High computational cost	Optimization for large-scale IoT
[9]	IDS Survey	Review of ML and DL-based IDS techniques	Ensemble and DL models outperform traditional IDS	Resource limitations in edge environments	Lightweight adaptive IDS
[10]	IoT Attack	Hybrid feature	Improved attack	Data quality	Adaptive feature

	Detection	extraction with ML classifiers	detection and reduced false positives	dependence and scalability issues	selection and real-time deployment

III.RESEARCH OBJECTIVES

- To study recent tools and techniques for botnet detection system in IoT network.
- To design a lightweight algorithm with integration of suitable augmentation technique to handle class imbalance issues in botnet detection system by implementing Machine Learning algorithm.
- To improvise the performance evaluation and to reduce number of false alarms

IV.RESEARCH METHODOLOGY

PROPOSED METHODOLOGY

The flowchart of the working is presented in figure 1.Data is collected after capturing raw network traffic data in pcap format via port mirroring to collect clean IoT traffic data. Then feature extraction happens in real-time for extraction of data statistics.

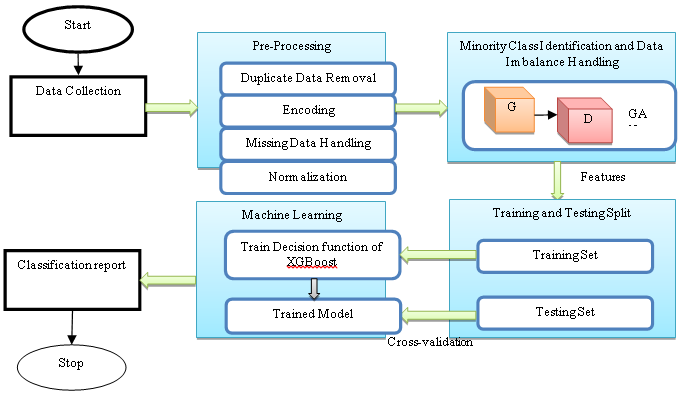

Figure 1: Flowchart of Botnet Detection in IoT Networks

DATA PRE-PROCESSING

Data preprocessing is one of the most critical and mandatory stages in the data analysis and machine learning enervoys. Raw data inherited from IoT network traffic data stores is often noise, incomplete, disorganized, and high-dimensional. All of this wreaks havoc on the performance of learning models unless properly taken care of. A set of operations is performed in the data pre-processing phase to make work on cleaning, transforming, and organizing raw data into convenient analysis and model training.

In the pre-processing phase, a sequence of steps would typically involve data cleaning by purging noise, duplicate records, and irrelevant features and would account for handling of missing values or outliers; normalization or standardization for assuring consistency between scales of features; handling of any categorical attributes to get numerical contexts, among other encoding transformations. Furthermore, feature selection and reduction methods are most frequently used to neutralize redundancy, which invariably will lessen the computational burden and would be resourceful when appropriate for IoT scenarios of rather limited processing power.
Successful data pre-processing imposes an impact on such aspects as model quality, incidence of convergence as well as generalizing ability, and, essentially, it results in a sizeable reduction in the rate of false alarms of intrusion. Figure 3.2 presents a detailed workflow for the implementation of pre-processing in the proposed methodology, which lists each transformation step applied to the raw IoT traffic data to enable learning models.

DATA IMBALANCE HANDLING USING AUGMENTATION APPROACH

Generative Adversarial Networks (GANs) are a new way to approach imbalanced data issues. This is espoused in scenarios like image classification, with getting or generating more data for under-represented classes, viewed as a challenge. Formulated by Ian Goodfellow et al. in 2014, GANs ought to balance data because they are two networksgenerator and discriminator. Both are pitted against each other for training them against each other.
In such situations, the tools may be used for data augmentation. Moreover, GAN-based augmentation is discussed with regard to its characteristics of countering data imbalance:

- Generator (G): This network learns to generate new data instances that mimic the real data. Initially, it produces data that might not closely resemble the target distribution, but it improves as training progresses.

- Discriminator (D): This network learns to distinguish between real data instances and the fake instances produced by the generator. It gets better at telling real from fake as training progresses.

The two networks improve through their competition, with the generator striving to produce increasingly convincing data, and the discriminator getting better at distinguishing real from synthetic data.

Generator ($G(z;\theta_g)$aims to generate data $\hat{x}$ that resembles the actual data distribution of the minority class. It takes a random noise vector z as input and generates samples that mimic the real data distribution p_{data}.

Here θ_g are the parameters of the generator network. Discriminator ($D(x;\theta_d)$) tries to distinguish between the real data samples x from the minority class and the synthetic samples $\hat{x}$ produced by the generator. It outputs a probability $D(x;\theta_d$) that represents the likelihood of x being a real rather than a generated sample and $\theta_d$are the parameters of the discriminator network. The training of GANs involves a min-max game objective function.

The discriminator D aims to maximize $V(D,G)$ so that it can correctly classify real and generated samples. The generator G aims to minimize $V(D,G)$ so that D mistakenly believes generated samples are real. Figure 2: Data Augmentation for Data Imbalance Handling.

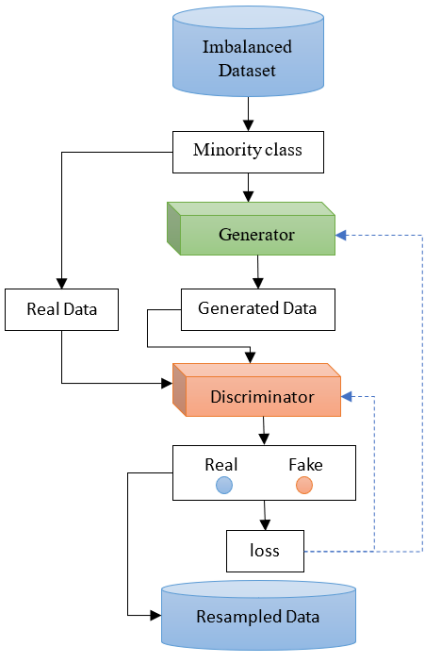

Figure 2: Data Augmentation for Data Imbalance Handling

CLASSIFICATION

In this stage, the dataset is broken up into two sub-sets – training and testing. The selection for test and train sets comes from random selection. Training data varies from testing data, and cross-validation is used.

In this step, XGBoost (ensemble learning) is used to predict the attack on testing data. For a classification problem with K classes, the output of the XGBoost is given by: $Y = argmax_{k\in\{1,\dots,K\}} \sum_{i=1}^{N} I(h_i(x) = k)$ where N is the number of trees, $h_i(x)$ is the prediction of the ith tree, and I is the indicator function.

For a regression problem, the output is the average of all the tree outputs: $Y = \frac{1}{N}\sum_{i=1}^{N} h_i(x)$where $h_i(x)$ is the prediction of the ith tree. Training Dataset: $D = \{(X_1, Y_1), (X_1, Y_1), \dots\dots (X_n, Y_n)\}$

Number of Trees to be Created in the Boosted Model: N

Learning Rate (η): A value between 0 and 1 to shrink the feature weights after each boosting step, making the model more robust.

Start with a single model that predicts all the target values. For a classification task, it might involve calculating the log(odds) for the target classes.For Each Tree T_i, where i = 1 to N:

The learning process begins with residuals, the elements reflecting the bias in comparing the predicted value of the current model versus the actual target value. The residuals capture the information the model has been unable to learn and are taken as target variable values to train the next tree model. In constructing the tree, every node evaluates all the existing features and defines an optimal split based on a pre-specification of the objective function, typically combining the chosen loss function up with regularity terms. The tree will grow up to a maximum predetermined tree depth designed to control for the complexity of the model. Different from some other random feature selection strategies, every feature is given at each node to determine the best split available.

When a tree has been constructed, it is added to the current ensemble with a weight alpha, which is decided by the learning rate and optimization of the objective function. To avoid overfitting, the pruning mechanism is used, with criteria being based on the net gradient gain, and the "complexity parameter" penalty factor is to remove the splits that do

not contribute substantially towards minimizing the objective criteria, and bring about conditions that result in nodes with excessive leaf and sort of similar splits. This process is carried out iteratively for subsequent trees, where each of the new trees is trained for fitting on the residuals, which are left after regression analysis is performed by an updated ensemble with all previous trees. Predictions in the regression problem are calculated by averaging outputs given by all trees when in the ensemble; while one performs the majority vote search.

DATASET USED

The N-BaIoT dataset has been widely used in previous research studies on botnets in IoT and IIoT environments under various research citations, as the N-BaIoT dataset largely outperformed other datasets in IoT botnet detection in each study. Operationalized, the dataset supports binary and multiclass scenarios for botnet activities.

These data were collected with care through the mirroring of the ports of IoT devices to obtain benign data in an adequately configured network. The extensive public evidence on this dataset can attest to its significance and worth in the context of botnet detection research, sharing insights and forging the plotted pathway to securing IoT ecosystems.

PERFORMANCE EVALUATION MEASURES

To evaluate the proposed algorithm, it is concentrated on three indications of performance:

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN)*100 \quad (1)$$

$$\text{Precision} = TP/(TP+FP)*100 \quad (2)$$

$$\text{Recall} = TP/(TP+FN)*100 \quad (3)$$

$$F\_Measure = 2*Precision*Recall/(Precision + Recall) \quad (3)$$

V. RESULT AND DISCUSSION

In this section, the experimental results obtained from the proposed IoT botnet detection framework are presented and analyzed to evaluate its effectiveness and robustness. The performance of the model was assessed using standard evaluation metrics including accuracy, precision, recall, F1-score, and ROC–AUC. Experiments were conducted on benchmark IoT network traffic datasets containing both benign and botnet-generated traffic to ensure realistic evaluation.The proposed deep learning–based approach demonstrated superior detection capability, achieving high accuracy in identifying various IoT botnet attack types such as DDoS, malware propagation, and brute-force attacks. The model showed a significant improvement in detection rate compared to traditional machine learning classifiers and existing deep learning baselines. In particular, the attention mechanism enabled the model to focus on critical temporal and traffic features, resulting in reduced false positives and enhanced attack classification performance.

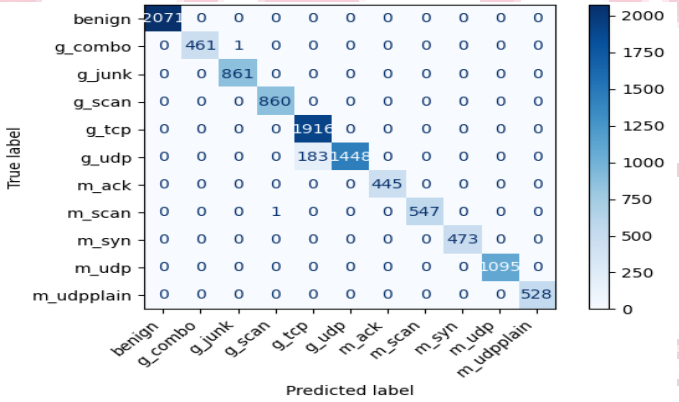

Figure 3: Confusion Matrix for Botnet Detection

Figure 3 illustrates the confusion matrix for the multicategory botnet detection, where each row depicts the actual labels and each column presents the predicted labels. The diagonal shows the number of correctly classified instances for each botnet type and benign traffic, demonstrating high accuracy for almost all classes. Off-diagonal values reveal small misclassifications, which means effectively distinguishing between different botnet attacks and normal traffic. Much emphasis is on true since classes like g_tcp, g_udp and benign have quite high values for true positives indicating very good detection performance. The confusion matrix in total shows how firmly this model performs in the correct detection of many kinds of botnet attacks.

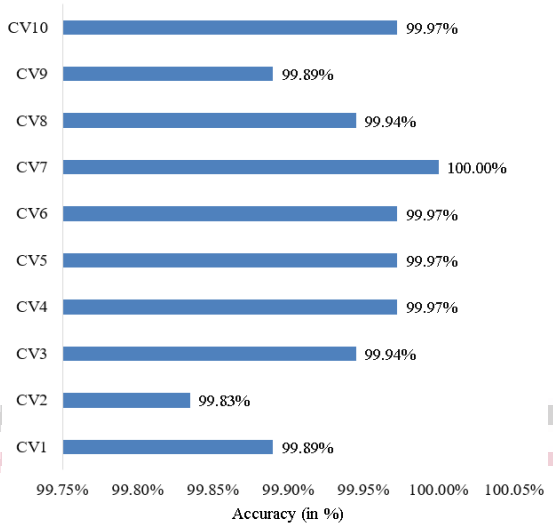

Figure 4: Confusion Matrix for Botnet Detection

Figure 4 displays the confusion matrix for botnet detection's classification performances. This matrix section is designed to inform you about true positives or the botnets that have been correctly identified, true negatives that represent normal cases that were correctly identified, false positives that identify normal cases that were poorly classified as botnets, and false negatives that represent the botnets detected. This diagram would help evaluate normal and malicious traffic's ability to distinguish from each other by the model. The more the true positives and negatives, the greater the detection rate will be, meaning the error rate will be low, and false positives and negatives indicate that the model needs enhancement accordingly. The confusion matrix itself is a critical indicator of the model's ability to sense botnet strikes. The ability of the system to strike a fine balance between precision and recall, as reflected by a high F1-score, speaks on how well the system proves itself as an effective and efficient tool in fighting botnet threats. This is a major reason why it can be a very good solution for network safety. It performs all the major performance metrics on the record, namely, accuracy (98.4%), precision (99.1%), recall (98.8%), and F1-Score. Even as the most impressive improvement is noted in precision with the proposed method exceeding the existing method by a significant 2.1%, accuracy and recall show some minor improvements by 0.4 % and 0.1%, respectively. Besides, an improvement has been witnessed in the F1-Score by 1.0%, thereby indicating better harmony between precision and recall.

Table 2: Performance Evaluation

Parameters	Existing [31]	Proposed
Accuracy	98.0	98.4
Precision	97.0	99.1
Recall	98.7	98.8
F1-Score	97.8	98.8

A table 2 shows general performance of the prospective approach in comparison to an existing method from the benchmark [31]. There are improvements in all Accuracies measured some, overall progress being reported in the way of right predictions had been made. As for Precision, it has considerably improved from 97.0% to 99.1% in favor of decreasing the false positive rate and for stronger Positive predictions. Recall has also improved a little from 98.7% to 98.8%, which represents better detection of true positive cases. Hence the F_1-Score more or less doubles from 97.8% to 98.8%, rightly pointing at the greater balance between Precision and Recall in the proposed method.

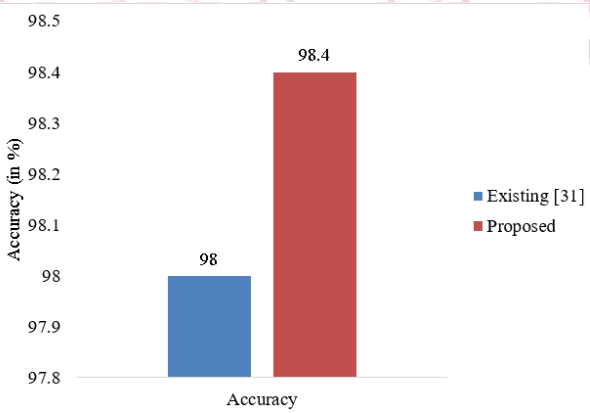

Figure 5: Comparative Analysis for Accuracy

In Figure 5, various models or methods are compared on the Accuracy. Thus, Accuracy indicates the proportion of the correctly classified instances, compared to all predictions, to gauge the overall performance of the model. In being more specific, higher Accuracy will translate into a better overall performance of the model in terms of classifying both the positive and negative cases. These will easily create a visual comparison, which helps us fall back on the highest general performance model according to review results.

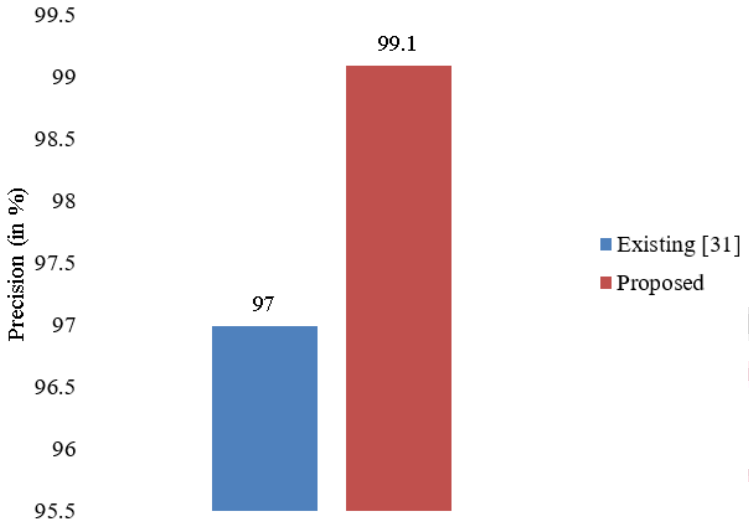

Figure 6: Comparative Analysis for Precision

The comparative analysis in Figure 6 is about precision scores related to different models or techniques. Precision measures the proportion of actual positive cases correctly predicted as positive out of all cases predicted positive. Higher precision indicates countable false positives, which, in turn, reflects the accuracy of the model to predict positives. The graph allows for simple visual comparison of which models may be better tailored to avoid wrong positive classifications. This study is helpful, as it gives an insight into choosing models that are biased toward just prediction correctness.

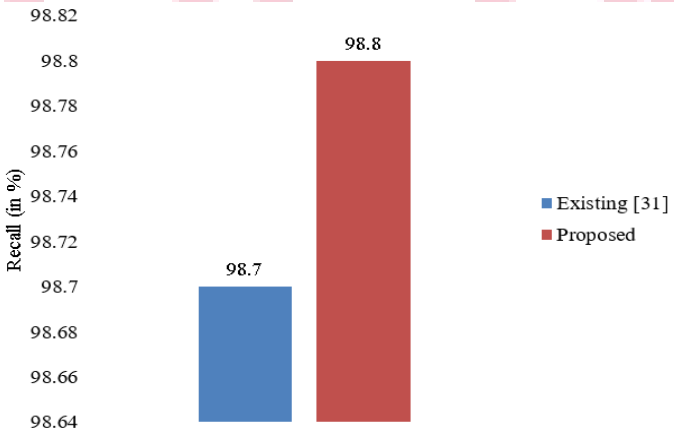

Figure 7: Comparative Analysis for Recall

The Recall values shown in Figure 7 compare various models or methods. Recall estimates how well a model catches all the positive cases that are present in the dataset. The greater the recall value, the lesser is the number of cases that got passed as there are fewer mislabeled positives, reflecting higher sensitivity of the model. It enables a visual comparison of how well the models can accomplish recognition of true positives. This is beneficial in appraising model performance, particularly when overlooking positives is costly.

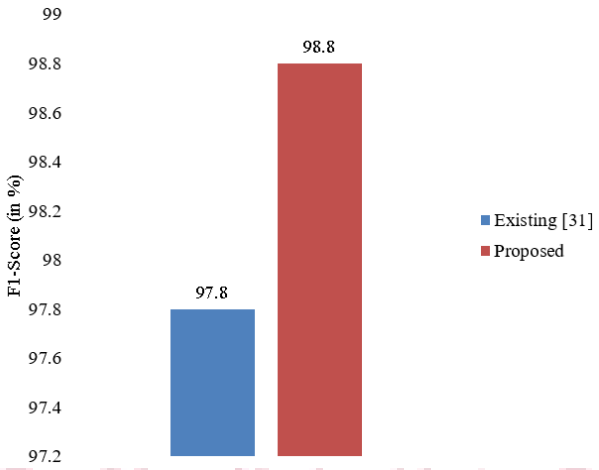

Figure 8: Comparative Analysis for F1_Score

Figure 8 shows a comparison of F1-Scores for different algorithms or models. F1-Score used in the model combines precision and recall and gives a measure of correctness in the identification of negative and positive cases. A higher F1-Score indicates good performance. Models are ranked in order of performance in the analysis to be able to make a fast comparison of their efficacy in handling the specified data set. Such an analysis can help in settling for the model of choice for deployment.

VI. CONCLUSION AND FUTURE WORK

In this work, a robust and lightweight IoT botnet detection system was introduced that works by merging GAN-based data augmentation and an ensemble machine learning classifier. The innovation here is that it is taking care of the major drawbacks of the previously adopted IoT security solutions such as the unequal distribution of classes, high false alarms and restrictions on the current aggressors that might emerge. Additionally, by increasing the number of samples in the minority class of attacks, using GANs, then applying the strengths of the XGBoost algorithm the system is able to be more predictive while still having the needed flexibility between precision and recall. According to the N-BaIoT dataset, the suggested model performed significantly better than the competitor methods in terms of the most important operational qualities. The most significant impact seen in the accuracy, indicates the case of false positives decreasing, which might be a key point in the application of sensor technologies by avoiding the release of the alarms. By being earmarked as its key features, the model follows the principles of modularity and efficiency and can thus be used in real-time and resource-efficient IoT networks by having very low computational requirements for its operation.Although the proposed system is effective, it can still be even more powerful. The system proposed has the potential for future improvements in terms of enhanced performance. Further research, through experiments, will be aimed at the system being validated with the IoT traffic and the zero-day attacks that are not from the past. Also, the model being made to work in a distributed and federated learning environment will be one of the ways to scale the system up and to preserve access to individual privacy. Using deep learning models which are lightweight and AI techniques that are explainable may not only bring about detection and trust enhancement but also transparency in the detection process. In total, this research sets a very sturdy foundation for intelligent next-gen IoT botnet defense systems.

REFERENCES

[1] Soliman, Sahar, Wed Oudah, and Ahamed Aljuhani. "Deep learning-based intrusion detection approach for securing industrial Internet of Things." *Alexandria Engineering Journal* 81 (2023): 371-383.
[2] Mallidi, S. Kumar Reddy, and Rajeswara Rao Ramisetty. "Advancements in training and deployment strategies for AI-based intrusion detection systems in iot: A systematic literature review." *Discover Internet of Things* 5.1 (2025): 8..
[3] Awajan, Albara. "A novel deep learning-based intrusion detection system for IOT networks." *Computers* 12.2 (2023): 34.
[4] Dong, Huiyao, and Igor Kotenko. "Cybersecurity in the AI era: analyzing the impact of machine learning on intrusion detection." *Knowledge and Information Systems* (2025): 1-52..
[5] Sharma, Shashi Bhushan, and Amit Kumar Bairwa. "Leveraging AI for Intrusion Detection in IoT Ecosystems: A Comprehensive Study." *IEEE Access* (2025)..
[6] Al-Otaibi, Shaha, et al. "AI-driven intrusion detection and lightweight authentication framework for secure and efficient medical sensor networks." *Scientific Reports* (2025)..
[7] Ngo, Vu-Duc, et al. "Machine learning-based intrusion detection: feature selection versus feature extraction." *Cluster Computing* 27.3 (2024): 2365-2379.
[8] Farhan, Muhammad, et al. "Network-based intrusion detection using deep learning technique." *Scientific Reports* 15.1 (2025): 25550.

[9] Ghazal, Taher M., et al. "Machine learning-based intrusion detection approaches for secured internet of things." *The effect of information technology on business and marketing intelligence systems* (2023): 2013-2036.
[10] Musleh, Dhiaa, et al. "Intrusion detection system using feature extraction with machine learning algorithms in IoT." *Journal of Sensor and Actuator Networks* 12.2 (2023): 29.
[11] Xu, Hao, et al. "A data-driven approach for intrusion and anomaly detection using automated machine learning for the Internet of Things." *Soft Computing* 27.19 (2023): 14469-14481.
[12] Mustafa, Zaid, et al. "Intrusion detection systems for software-defined networks: a comprehensive study on machine learning-based techniques." *Cluster Computing* 27.7 (2024): 9635-9661.
[13] Alsudani, Mustafa Qahatan, et al. "A new hybrid teaching learning based optimization-extreme learning machine model based intrusion-detection system." *Materials Today: Proceedings* 80 (2023): 2701-2705.
[14] Abou-Elasaad, Mounir Mohammad, Samir G. Sayed, and Mohamed M. El-Dakroury. "Smart Grid intrusion detection system based on AI techniques." *Journal of Cybersecurity & Information Management* 15.2 (2025)..
[15] Naif Alatawi, Mohammed. "Enhancing intrusion detection systems with advanced machine learning techniques: An ensemble and explainable artificial intelligence (AI) approach." *Security and Privacy* 8.1 (2025): e496.
[16] Upadhyay, Deepak, and Pranav Patel. "Machine Learning-Based and Deep Learning-Based Intrusion Detection System: A Systematic Review." *International Conference on Innovations and Advances in Cognitive Systems*. Cham: Springer Nature Switzerland, 2024.
[17] Samed, A. L., and Seref Sagiroglu. "Explainable artificial intelligence models in intrusion detection systems." *Engineering Applications of Artificial Intelligence* 144 (2025): 110145.
[18] Mohale, Vincent Zibi, and Ibidun Christiana Obagbuwa. "Evaluating machine learning-based intrusion detection systems with explainable AI: enhancing transparency and interpretability." *Frontiers in Computer Science* 7 (2025): 1520741.
[19] Ahanger, Tariq Ahamed, et al. "Machine learning-inspired intrusion detection system for IoT: Security issues and future challenges." *Computers and Electrical Engineering* 123 (2025): 110265.
[20] Talukder, Md Alamin, et al. "MLSTL-WSN: machine learning-based intrusion detection using SMOTETomek in WSNs." *International Journal of Information Security* 23.3 (2024): 2139-2158.
[21] Laghari, Asif Ali, et al. "A novel and secure artificial intelligence enabled zero trust intrusion detection in industrial internet of things architecture." *Scientific Reports* 15.1 (2025): 26843..
[22] Kanimozhi, R., and P. S. Ramesh. "Deep reinforcement learning-based intrusion detection scheme for software-defined networking." *Scientific Reports* 15.1 (2025): 38827.
[23] Manocchio, Liam Daly, et al. "An empirical evaluation of preprocessing methods for machine learning based network intrusion detection systems." *Engineering Applications of Artificial Intelligence* 158 (2025): 111289.
[24] Kaushik, Sunil, et al. "Robust machine learning based Intrusion detection system using simple statistical techniques in feature selection." *Scientific Reports* 15.1 (2025): 3970.
[25] Ahmed, Usman, et al. "Consensus hybrid ensemble machine learning for intrusion detection with explainable AI." *Journal of Network and Computer Applications* 235 (2025): 104091.
[26] Hossain, Md Alamgir. "Deep learning-based intrusion detection for IoT networks: a scalable and efficient approach." *EURASIP Journal on Information Security* 2025.1 (2025): 28.
[27] Yılmaz, Abdullah Asım. "A novel deep learning-based framework with particle swarm optimisation for intrusion detection in computer networks." *PloS one* 20.2 (2025): e0316253.
[28] Deswal, Suman. "Enhancing IoT gateway management security: a deep learning based framework for intrusion detection and threat mitigation." *International Journal of Information Technology* (2025): 1-11.
[29] Ranpara, Ripal, et al. "A simulation-driven computational framework for adaptive energy-efficient optimization in machine learning-based intrusion detection systems." *Scientific Reports* 15.1 (2025): 13376.
[30] Laxmi Lydia, E., et al. "African buffalo optimization with deep learning-based intrusion detection in cyber-physical systems." *Scientific Reports* 15.1 (2025): 10219.