A Hybrid Metaheuristic Framework for High-Dimensional Cyber-Infiltration Detection & Mitigation

Nazar Faried Yousif Mohamed[1], Ashraf Gasim Elsid Abdalla [2], AbuBakr Obeid Mosa Mohammed [3], Hashim Elshafie[4], Rania Ali Elkhidir Ali [5] Hashim Albasheer [6], Omer Elsier Tayfour[7]

[1] Department of Computer, College of Engineering, AL Neelain University, Khartoum, Sudan,
[2] Department of Computer Engineering, College of Engineering, Future University, Khartoum, Sudan,
[3] Department of Computer Engineering, College of Engineering, AL Neelain University, Khartoum, Sudan,
[4] Department of Computer Engineering, College of Computer Science, King Khalid University, Abha 61421, Saudi Arabia;
[5] Department of Computer Science, College of Computer Science, Hail University, Hail, Saudi Arabia;
[6] Department of Informatics and Computing, College of Computer Science, King Khalid University, Abha 61421, Saudi Arabia;
[7] Department of Computer Engineering, College of Computer Science, King Khalid University, Abha 61421, Saudi Arabia;

Abstract
Distributed Denial of Service (DDoS) attacks are among the most severe entities that threaten network infrastructures in contempor ary times because of their growing frequency and complexity. Th is paper presents a scalable and smart DDoS detection model that builds on machine learning and real-time analytics to improve det ection effectiveness and response performance. The approach sug gested is a hybrid ensemble model, a combination of the Random Forest and XGBoost with soft voting to deliver high classificatio n accuracy and strength. In order to enhance the computational ef ficiency, there is the feature selection mechanism that is founded on the Bat Algorithm, which minimizes the data dimensionality without eliminating the important discriminative features. In addi tion, a real-time analytics engine is created to allow the analysis of traffic and quick decision-making. A lightweight integrating la yer, which is a JSON-based one, is meant to enable smooth intera ction between the parts of the system, such as the detection engin e, the visualization dashboard, and the alerting module. An auto mated response and mitigation system is also integrated into the s ystem and is able to perform immediate actions in response, like blocking and filtering of traffic. The experimental findings of the CIC-DDoS2019 dataset show that the proposed framework yields a training accuracy of 99.99% and a testing accuracy of 99.94% with a precision of 99.97, a recall of 99.95, and an F1-score of 99. 96. Also, the model has a high specificity of 99.89 and a low fals e positive rate (FPR) of 0.11, demonstrating good detection and h igh accuracy of generalization. These findings provide evidence of the suitability of the suggested system in practical use in netw ork security settings.

Keywords:
Intrusion Detection System, Hybrid Wrapper (Bat–RF) Algorithm Ensemble Learning, XGBoost, DDoS

1. Introduction

Software-Defined Networks (SDN), the Internet of Things (IoT), and cloud computing have undergone a massive development trend, which has led to a significant exposure to cyber threats, especially DDoS attacks. The purpose of these attacks is to deplete network resources, resulting in the disruption of services, financial losses, and deterioration of Quality of Service (QoS). As the patterns of attack continue to increase in complexity and diversity, the traditional Intrusion Detection Systems (IDS), based on signature-based detection methods, are unable to effectively deal with the current threats [1].

In order to address these issues, machine learning (ML) algorithms have been extensively implemented in intrusion detection systems because of their capability to process large volumes of network traffic data and reveal concealed patterns. Some of these studies have indicated that algorithms like the random forest can be highly accurate in classifying normal and malicious traffic, especially when used together with feature selection methods to improve their performance and to minimize the dimensions [1], [2]. Moreover, deep learning methods have also enhanced the detection abilities through the identification of complicated associations in network traffic information, particularly in dynamic IoT settings [3].in spite of these developments, single-model methods continue to have fundamental drawbacks of poor generalization, overfitting, and decreased efficiency in very high-dimensional data. Therefore, the current study has been redirected to ensemble learning, which integrates a set of models to enhance the detection accuracy and the system reliability. Such methods have proved to be extremely useful in the process of optimizing the overall functionality of intrusion detection systems, both in SDN settings and real-time applications [4]. Moreover, hybrid learning approaches that combine machine learning and deep learning have demonstrated promising outcomes in the context of complex and evolving cyberattacks identification [5], [6]. In addition to this, recent research has highlighted the significance of using feature selection methods alongside ensemble learning to minimize the complexity of computations and still achieve high detection rates. Ensemble-based frameworks with lightweight designs have shown to perform well in

Manuscript received September 5, 2026
Manuscript revised September 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.9.20

resource-restricted settings like IoT [7], and more complex and adaptable ensemble models have performed well in managing advanced and dynamic cyber threats [8].

Although machine learning and deep learning methods have demonstrated considerable advancements in the field of DDoS attack detection, there are key issues that plague the current state of intrusion detection systems. The first issue is that most existing models are affected by high dimensionality, negatively affecting computational complexity and model performance. Second, single-model methods fail to provide an optimal tradeoff between detection and False Positive Rate (FPR), especially in dynamic networks, and several of the existing studies lack an effective solution that can be used in real-time detection, which is a critical need in modern cybersecurity solutions. Moreover, there exists the absence of unified structures that can effectively integrate optimized feature selection with ensemble learning in order to obtain high performance with reduced computational overhead. Hence, there is a need to establish a sophisticated hybrid model, which can incorporate the smart feature selection models in conjunction with ensemble learning algorithms. The purpose of such a model is to enhance the detection accuracy, lower the false positive rate, and increase the efficiency of the system in contemporary network environments and facilitate real-time deployment. additionally, the combination of feature selection techniques and ensemble classifiers has also enhanced the detection effectiveness by reducing dimensions and still preserving the most discriminative features of the traffic.

Recently, these hybrid ensemble structures have been shown to test with greater than 99% accuracy and very low false positive rates, proving that ensemble learning is now on the leading edge of developments in real-time DDoS detection and mitigation. in spite of these developments, there are still difficulties in implementing systems that can be deployed in a manner that allows them to offer high accuracy, low false positives, and real-time performance and, at the same time, be scalable. The current study is based on these advancements and defines a hybrid ensemble machine learning platform combining optimized feature selection and a powerful RF-XGBoost ensemble model to provide real-time and high-fidelity DDoS detection and prevention in functioning networks.

2. Tables, Figures, and Equations

Table 1: Comparative Studies

Study	*Dataset*	Approach	Ensemble	Accuracy	Remark
[9]	CICIDS2017/2018, CICDDoS2019	PCA + RF/SVM/KNN/DT	✗	98.90%	Classical ML, no ensemble
[10]	CICDDoS2019	Voting ensemble (5 ML models)	☑	99.9%	IoT-focused, strong ensemble performance
[11]	SDN-SCADA custom	DT ensembles (Bag, Boost, RUSBoost)	☑	92.6–92.9%	Simulation dataset, lower accuracy
[12]	CICDDoS2019 + synthetic real traffic	XGBoost	Partial	99.99% offline / 95.3% real	Real-time evaluation shows gap
[13]	CICDDoS2019, KDD-CUP99, UNSW-NB15	RF, GB, SVM, KNN, DT + hybrid	Partial	99.44%	Hybrid ensembles across datasets

Table 2: Project Technical Requirements and Specifications

Environment	Details
Software Specification	
Operating System	macOS (Silicon/Intel), Linux, or Windows 10/11
Programming Language:	Python version: 3.11 or higher (Tested on 3.14.0)
Python Libraries:	1.1.1 Pandas, NumPy, scikit-learn, xgboost, pyarrow, streamlit, Plotly, Matplotlib, seaborn, psutil, pickle.
Core Frameworks:	1.1.2 Web Dashboard: Streamli / Containerization: Docker, Docker Compose
Hardware Specifications	
Processor:	1.1.3 Intel Core i5 / Apple M1 or equivalent
RAM:	8 GB (Minimum) / 16 GB (Recommended for training)
Storage:	5 GB available space (includes dataset and virtual environment)
GPU:	Optional (XGBoost can utilize GPU for faster training if available)

Table 3: Final Performance Metrics (Binary Classification)

Evaluation Metric	Experimental Value	Percentage (%)
Accuracy	0.99937	99.94%
Precision (Attack)	0.99970	99.97%
Recall (Sensitivity)	0.99949	99.95%
F1-Score	0.99959	99.96%
False Positive Rate (FPR)	0.00106	0.11%
False Negative Rate (FNR)	0.00051	0.05%

Table 4: Confusion Matrix

Actual	Predicted	Type	Count	Percentage
Benign	Benign	True Negative	18921	22.20%
Benign	Attack	False Positive	20	0.02%
Attack	Benign	False Positive	34	0.04%
Attack	Attack	True Positive	66241	77.73%

Table 5: Granular Performance Analysis

Aspect	Value
Algorithms	Random Forest + XGBoost (Voting Ensemble)
Training Samples	"340,860"
Testing Samples	"85,216"
BAT Features + WRAPPER	12
Engineered Features	3
Total Features	15
Training Time	0.34 minutes
Prediction Time	235622.75 samples/sec
Throughput	235622.75 samples/sec
Final Accuracy	0.9994 (99.94%)

Table 6: Benchmarking Performance Analysis.

Study	Approach	Dataset and Features	Accuracy	FBR
[16]	Hybrid (RF + XGBoost)	SDN-specific Features	99.36%	0.61
[17]	Deep Neural Networks (DNN)	Benchmark Datasets	99.3%	0.72
[4]	Ensemble Online ML	Dynamic Feature Selection	99.2%	0.85
[9]	EDAD (PCA + RF)	CICIDS2017	98.9%	1.02
[14]	RF, XGBoost	Full Features	98.14%	1.80
[21]	RF, XGBoost, AdaBoost	Info Gain Selection	97.20%	2.5
[22]	Hybrid (DCNN-HMACO)	UNSW-NB15 + TON_IoT	92.14%	5.12
[23]	Hybrid Ensemble + Blockchain	UNSW-NB15	99.5%	0.45
Proposed Framework	BAT + RF-XGB Ensemble	6 Features	99.94%	0.01

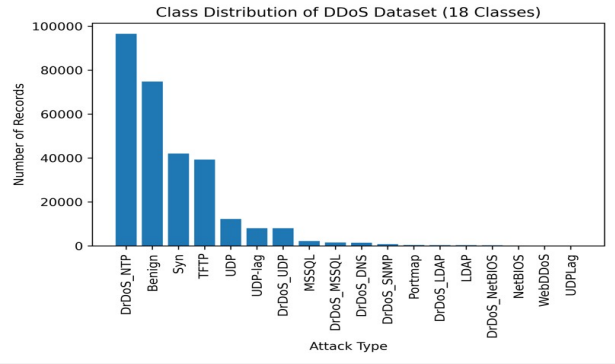

Figure 1: Class distribution of the DDoS dataset.

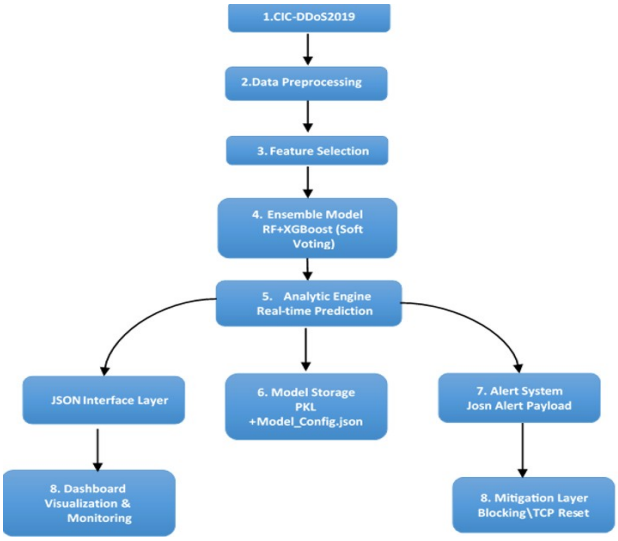

Figure 2: A-based integration layer is utilized to enable lightweight, real-time communication between the analytics engine, dashboard visualization module, and alerting system.

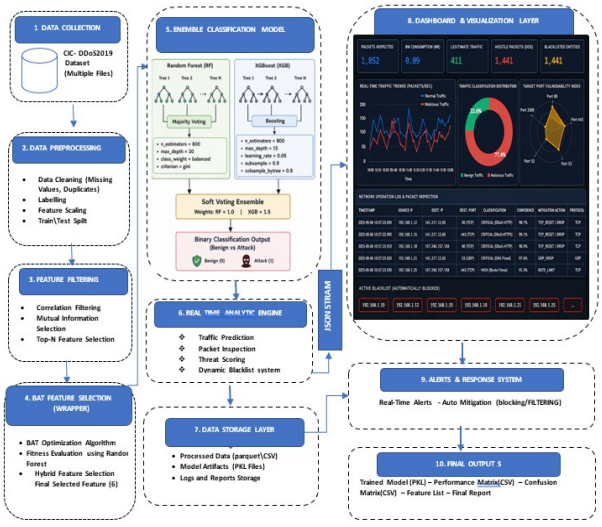

Figure 3: Model Steps Architecture

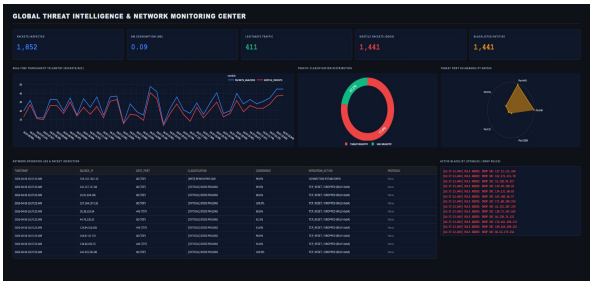

Figure 4: Dashboard Visualization & Monitoring

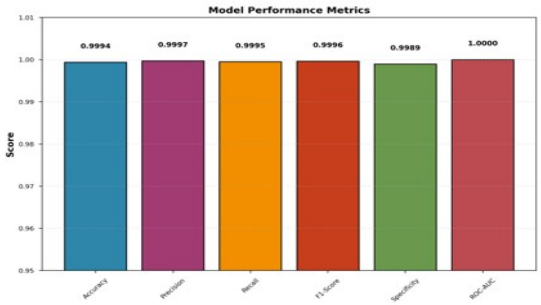

Figure 5: Model Performance Metrics

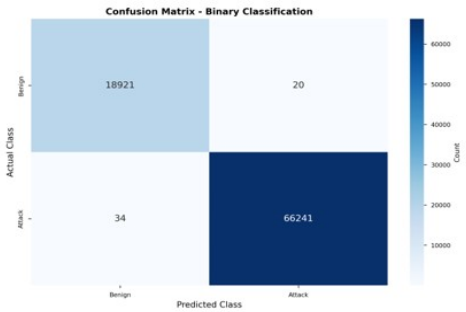

Figure 6: Confusion Matrix

Figure 7: Per-class metrics

2.1 Weighted Soft-Voting Ensemble Mechanism

The weighted soft-voting ensemble mechanism is used to predict the probability of a specific prediction. The proposed framework, in order to improve the reliability of detection, uses a weighted soft-voting ensemble approach. The process merges the probabilistic responses of the base learners, giving them various relative weights based on their respective measures of performance. The XGBoost classifier has a weight ($w_2 = 1.5$) that is higher than the Random Forest classifier ($w_1 = 1.0$).

$$P_{.final} = \frac{(w_1.P_{RF} + w_2.P_{XGB})}{(w_1 + w_2)}$$

Where:
P_{final}*: Final aggregated probability used for the classification decision.*
P_{RF}*: Probability predicted by Random Forest.*
P_{XGB}*: Probability predicted by XGBoost.*
$w_1 = 1.0$*: Weight of Random Forest.*
$w_2 = 1.5$*: Weight of XGBoost.*

2.2 The equations used to extract the accuracy and other values required in evaluating the Model:

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

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

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

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

$$FPR = \frac{FP}{FP+TN} \quad (5)$$

$$FNR = \frac{FN}{FN+TP} \quad (6)$$

TP = True Positives
TN = True Negatives
FP = False Positives
FN = False Negatives

3. Paragraphs and Itemizations: Literature Review

Over the past few years, there has been a host of research dedicated to the application of machine learning (ML) and ensemble learning to the problem of DDoS attack detection. Random Forest (RF), Support Vector Machines

(DT) are traditional supervised ML models that have proven to be used extensively in benchmark datasets, including CICIDS2017/2018 and CICDDoS2019. Abiramasundari and Ramaswamy [9] used PCA-based feature selection with these classical models with detection rates of 98.9% on CICIDS2017 and 98.7% on CICDDoS2019. Nevertheless, the method used depended on individual classifiers, which restricted the capability to withstand new attack styles. recent studies have been using ensemble learning methods in order to overcome these limitations. Ensemble techniques, including majority voting, bagging, and hybrid combinations, combine several models to enhance detection and lower false positive rates. Cheng and Feng [10] employed a majority-voting combination of five ML models to the CICDDoS2019 dataset, yielding almost 100% detection, and reduced false positives by a big margin. This illustrates how effective ensemble learning in an IoT setting where there are varied types of attacks can be. ensemble learning has been used in other studies in other network settings. Oyucu et al. [11] compared bagged, boosted, and RUSBoost decision tree ensembles using a custom SDN-SCADA simulation dataset and achieved a detection rate of 92.6 to 92.9%. Although the performance of ensemble learning was better than the single models, it was not as high as the performance on the standard benchmark datasets, which demonstrates the influence of the dataset characteristics. ensemble methods have also been a benefit of real-time applications. Lee et al. [12] adopted the XGBoost-based real-time detection system on CICDDoS2019 together with synthetic network traffic. Offline testing had achieved 99.99% accuracy, whereas in practice, real-world testing had only 95.3% detection accuracy, underscoring the need to test models in real-world conditions.

Lastly, RF, Gradient Boosting (GB), SVM, KNN, and DT have been developed into hybrid ensembles that have been used on a variety of datasets, such as CICDDoS2019, KDD-CUP99, and UNSW-NB15 [13]. As an example, the RF+KNN hybrid ensemble showed a high accuracy of 99.44%, indicating how a combination of models can enhance the generalization and performance of heterogeneous data.

Based on the results presented in Table 1, ensemble learning strategies are always superior to single-model ones in the area of DDoS detection, as they perform better in accuracy and low false-positive rates in the context of different datasets. Hybrid ensembles, especially, are shown to have excellent generalization properties and to be resistant to complex attack patterns, which makes them suitable to real-time network security usage. The comparison highlights the importance of integrating several models and feature selection methods to boost the performance of detection in high-dimensional and heterogeneous network settings. altogether, the literature demonstrates with no doubts that ensemble learning has become a predominant and an extremely efficient model in DDoS detection systems recently. It is much more accurate in detection, has lower false positive rates, and has system robustness. Nevertheless, there are still issues in attaining real-time performance, scalability, and useful feature selection. Hence, additional studies are needed to come up with the optimum hybrid ensemble solutions, which can compromise precision, computational speed, and real-time implementation in the current networked world.

4. Methodology

The given system is aimed at being an intelligent network monitoring platform, which will combine traffic monitoring, threat detection, and automated security response, and this can be applied to intrusion detection and prevention systems (IDS/IPS) and DDoS monitoring. It is a flow-based analysis that is used to derive features in order to efficiently model network traffic behavior.
The dataset was a labeled flow-based dataset, comprising 288,561 samples and six key features and a target label, which were 18 types of DDoS attacks and benign traffic. The chosen features contain some important statistical and temporal flow features like flow duration, total forward and backward packets, flow bytes/s, flow packets/s, and mean forward packet length, which are a complete representation of the network behavior. The dataset has no missing, duplicated, or infinite values and is of high quality, but as the data is in raw form, StandardScaler was used in the preprocessing stage to bring the feature distributions to a more normal distribution and enhance model performance. The data set is comprised of 18 types of attacks with evident class imbalance. The dominant one (33.46%) is DrDoS_NTP, and minor classes like WebDDoS and UDPLag comprise less than 0.05% of all records.

4.1 Stages of improving the Model:

4.1.1 Data preprocessing

The data preprocessing phase was done in an organized series of steps. First, various files of the CIC-DDoS-2019 dataset were loaded and merged into a single dataset. Data cleaning was then performed through checking and eliminating missing values and duplicate data to give quality data. The characteristics were disaggregated, and the class label, as well as numerical and categorical features, was determined. after that, further data quality was improved by eliminating zero-variance features, which do not play a role in the learning process. The data were then further divided into training and testing sets in an 80/20 ratio with a stratified sampling approach in order to maintain the distribution of classes. in preprocessing, StandardScaler was used to perform feature

standardization of the features to normalize their scales and improve the performance of the models. Lastly, the processed datasets were stored in different files to be used later in the training and evaluations.

4.1.2 A feature filtering stage

Used to enhance the quality of data and decrease dimensions. First, a correlation-based method was used to eliminate highly correlated features, where a correlation coefficient of more than 0.95 was used to drop the feature, thus preventing redundancy and improving model performance. Mutual Information (MI) was then used to assess the relevance of a given feature relative to the target variable, and only features that have a value above the set threshold were retained. The last stage was to select the most informative features (Top-N) using their MI scores, and so the feature set was reduced to maintain the most important information to classify the object.

4.1.3 A feature selection stage

It was applied in a hybrid method with a hybrid approach of the BAT algorithm as a wrapper. At this step, each solution is a subset of features, and its quality is estimated with a random forest model cross-validated. The algorithm is intended to pick an optimal number of features (around 6 features) by balancing the classification accuracy and feature reduction by using a fitness term with a penalty term against too many features. The optimization is done through an iterative process with a population of candidate solutions (bats) where positions and velocities are updated, and a local search mechanism is implemented until the best set of features is achieved. Lastly, the features that have been chosen are reduced to perform dimensionality reduction of data and enhance the performance of the models in the following steps.

Algorithm 1: Hybrid BAT-RF Feature Selection Module

Input: - Preprocessed Training and Testing datasets.
Output: - Reduced Feature Subsets for Ensemble Classification.

Data Preparation:
Load datasets and separate features X from target labels y.
Apply Label Encoding to normalize categorical target classes.
Hybrid Optimization (BAT Algorithm):
Initialize bat population (feature subsets) and parameters (f, A, r).
For each iteration do:
Update bat positions based on frequency and velocity.
Convert positions to binary vectors (0: excluded, 1: selected).
Fitness Evaluation (Wrapper):
Train a Random Forest (RF) classifier on the selected features.
Calculate fitness: Accuracy – Penalty (Feature\Count).
Update the Global Best feature subset (Sbest).
End For
Feature Transformation:
Extract the optimal feature names from Sbest.
Filter both Train and Test sets to include only the selected features.
Export & Storage:
Save the reduced datasets (Stage 4) and label encoders.
Transition: Pass refined data to Stage 5 (Ensemble: RF + XGBoost).

4.1.4 The classification and performance analysis stage

It was implemented using an ensemble model that combines Random Forest and was applied in the ensemble model, which is a combination of Random Forest and XGBoost based on a soft voting approach with more weight being placed on XGBoost to give the prediction accuracy. The dataset generated after the BAT feature selection step (where only six key features were used) was used to train the model. The formulated problem was a binary classification problem (benign vs. attack). Before training, the feature standardization was used to normalize the data. the training set was then used to train the model, which was tested on the testing set. According to the confusion matrix, multiple measures of performance were taken that include accuracy, recall, precision, F1 score, false positive rate (FPR), and specificity. The outcome showed a very high performance with accuracy more than the target (99.5%), which indicates that the model is effective in identifying the attacks and minimizing errors, especially false alarms. Also, the final model, evaluation findings, and reports were stored to be used or published in academics. The last step uses a high-performance soft-voting ensemble classifier. This composite design builds on the mutual complements between bagging and boosting:

- Random Forest (RF): 800 trees with a maximum depth of 30; this is to minimize variance.
- XGBoost (XGB): It was using 800 boosting rounds with a learning rate of 0.05, with the aim of minimizing bias and maximizing residual errors.
- Soft Voting: The model uses a weighted average of the predicted probabilities with the weights being 1.0 (RF) and 1.5 (XGB). This combined method has a higher emphasis on the accuracy of boosting the performance of XGBoost without losing the generality of Random Forest.

Algorithm 2: Final Detection via Ensemble Voting Classifier

The algorithm is used to detect the final detection using an ensemble voting classifier.
This algorithm fuses the chosen features with engineered features to do the final classification

Input: - Sbest: Feature subset (from Alg1).
- E: Set of 3 Engineered Features (Intensity, Variability, ACK-Ratio).
- T: Testing Data.

Output: - Report on Classification: Accuracy, Precision, Recall, F1-Score, FPR.

1. Feature Fusion: Build the finished feature space F = S best U E.
2. Model Training:
3. Train (RF) on F.

4. Train XGBoost on F.
5. Given each network packet p of testing set T:
6. Obtain probability P_RF from the RF model.
7. Obtain probability P_XGB from the XGBoost model.
8. Calculate Soft Voting Probability: P_final = (P_RF + P_XGB) / 2.
9. If (P final ≥ 0.5) then Label = "Attack" else Label = "Benign".
10. End for
11. Compute Metrics:
12. Accuracy = 99.94%
13. FPR = 0.11%
14. Return Performance evaluation metrics.

Considering the multi-class and imbalanced data, the ensemble learning method was chosen to increase the accuracy of detection. The proposed model is based on a combination of Random Forest (RF) and XGBoost, in which each of the classifiers learns various patterns of the data. An ensemble layer is used to combine their outputs to come up with a final decision on the classification. This mixed method minimizes the variance and bias and enhances detection of sophisticated patterns of DDoS attacks, especially when there are unbalanced data environments.

Real-time key performance indicators are constantly calculated at the system level, such as packets inspected, bandwidth usage, traffic categorization (benign versus malicious), and blacklisted entities, which give real-time network visibility. Time-series analysis is also used to track traffic behavior and sudden spikes, which might represent potential attacks. The ensemble-based classification component allows one to predict traffic in real-time, which can aid in making the correct security decisions on time. moreover, the port vulnerability analysis module determines ports that are of interest to attackers to determine possible points of weaknesses and to streamline the firewall policy. A deep packet inspection system is used to obtain more detailed information about attributes of a packet, like timestamps, source and destination IP addresses, ports, attack type, and confidence scores, and then suitable mitigation techniques are taken against the packet, like dropping packets or resetting connections. Lastly, a dynamic blacklisting process will automatically put the malicious IP addresses into a blacklist, enhancing the capability of the system to stop repeat attacks and implement adaptive security policy.

This architecture can be explained as follows: It is an end-to-end, real-time DDoS detection and mitigation architecture, which involves a hybrid feature selection strategy using the BAT optimization algorithm and an ensemble classification model (Random Forest and XGBoost) to label traffic with high-confidence threat scores, which are subsequently actioned upon with dynamic blacklisting and mitigation.

- Data Integration: JSON will be used to enable a flow of information between the back-end analytics and the front-end dashboard.
- Feature Optimization: To increase the processing speed in real-time, perform feature dimensionality reduction by using the BAT algorithm.
- Real-time Visualization: Visibility of a fully-fledged security dashboard to track attack spikes, traffic distribution, and blacklists in real-time.
- Automated Mitigation: IP-filtering rules (e.g., iptables/DROP) are immediately executed in accordance with the output of the ensemble model.

5. Results and Discussion

Here, we provide a comprehensive discussion of the findings of the experiments that were conducted based on the suggested BAT-RF-XGB framework. The performance is measured with standard measures, such as accuracy, precision, recall, F1-score, and false positive rate (FPR), which the experimental results show to be exceptionally efficient with the proposed BAT-RF-XGB framework, with an impressive accuracy of 99.94 and a negligible false positive rate (FPR) of 0.11, which is by no means a match to the current benchmarks. This is because of the effective combination of the BAT algorithm, which was able to decrease the data dimensionality by 81.2 percent (80 and above features to only 6 core attributes), and the soft-voting ensemble model. This ensemble is a good combination to reduce variance and avoid overfitting of Random Forest and minimize bias and classify complex samples as done by XGBoost. Such synergy of feature optimization and ensemble learning does not only improve the precision of detection but also the computational efficiency of the algorithm since it reduces dramatically the inference latency. The given framework, therefore, proves to be a perfect technical solution to high-throughput, time-sensitive network environments such as software-defined networking (SDN) and IoT gateway environments, in which the rate of decision-making and false alarms has to be as minimal as possible to maintain service continuity and security reliability. The last ensemble model with the hybrid set of 6 features proved to have outstanding performance with the testing set. The model had the highest accuracy of 99.94, as indicated in Table 4.

```
. PERFORMANCE METRICS
----------------------------------------
Training Accuracy:  0.9999 (99.99%)
Testing Accuracy:   0.9994 (99.94%)
Precision (Attack): 0.9997
Recall (Attack):    0.9995
F1-Score:           0.9996
Specificity:        0.9989
FPR:                0.0011 (0.11%)
FNR:                0.0005 (0.05%)
```

The high precision (99.97) means that the model will be able to detect the attacks without false alarms, and recall (99.95) will make sure that almost all the malicious actions are identified.

5.1 Confusion Matrix Interpretation

The results of the classification are granularly displayed in the confusion matrix (Figure X). The fraction of samples incorrectly classified by the model was very small compared to the overall sample of tests.
True Negatives (Benign): The model had a specificity of 99.89 in identifying benign traffic.
False Positives: The number of benign samples that were flagged as attacks was only 0.11%, which is very important in minimizing the alert fatigue in real-life security operations.

5.2 Granular Performance Analysis

The efficacy of the proposed ensemble framework in terms of classification is strictly tested by confusion matrix and per-class measures as shown in Figures [6] and [7]. Confusion matrix analysis indicates that the model has an impressive true-positive rate; the model was able to detect 66,241 instances of attacks and 18,921 benign samples. Only 20 False Positives (FP) and 34 False Negatives (FN) are the marginal error rate that illustrates the extraordinary ability of the model to ensure a high security posture, and at the same time, the network interruptions are minimized. also, the per-class performance comparison proves the strength of the weighted soft-voting mechanism. The attack category had the almost perfect F1-score of 0.9996, which equals a precision of 0.9997 and a recall of 0.9995. Equally, the 'Benign' category was very stable, with an F1 score of 0.9986. These measures indicate that hybrid feature selection (BAT) was effective to remove unnecessary noise and enable the ensemble learners (RF and XGBoost) to create accurate decision boundaries despite intricate traffic patterns.

The experimental findings prove that the suggested ensemble model, which is based on the combination of Random Forest and XGBoost with the help of a voting mechanism, shows excellent results. With a small set of features, 12 features chosen through the BAT-based wrapper and 3 features engineered, the model has high discriminative power and greatly reduced the dimensionality of features. The system is characterized by great computational efficiency, and its training time is just 0.34 minutes, and its prediction throughput is more than 235,000 samples per second.

It is worth noting that the obtained accuracy of 99.94% proves the efficacy of the chosen features and the versatility of the chosen ensemble learning approach, which is why the proposed method can be applicable in large-scale and real-time intrusion detection settings.

The figure7 offers a more detailed analysis of the suggested binary classification system (benign vs. attack), with per-class performance measures, as well as an overview of the model structure and the dataset qualities. The results of the experiment show that the discriminative ability of the model is extremely high in both classes. In the benign class, the model had a precision of 0.9982, a recall of 0.9989, and an F1-score of 0.9986, which means that the false positives and false negatives would be very minimal. Likewise, with the Attack class, a precision of 0.9997, recall of 0.9995, and F1-score of 0.9996 show that there is a near-perfect performance in detection, with minimal misclassification. All evaluation measures approach unity, which indicates that generalization has occurred and no observable classification bias exists between the classes. The near coincidence between the values of precision and recall further suggests a balanced trade-off between sensitivity and specificity—an important quality in security-sensitive applications like intrusion and attack detection. The proposed framework uses the voting ensemble approach, which combines both Random Forest and XGBoost classifiers, using the complementary decision-tree-based learning process to improve predictive robustness. There were 340,860 samples modeled and trained, and their evaluation was checked on an independent test set of 85,216 samples, thus enhancing the statistical validity and reliability of the reported results. Although the dataset size was high, the overall training time was about 0.34 minutes, which demonstrates the efficiency of the proposed method in terms of computational time and makes it suitable in real-world deployment situations that have to be retrained and adapted in a short time. In general, the results highlight the success, consistency, and feasibility of the suggested ensemble-based classification structure.

5.3 Comparative analysis Performance benchmarking.

In order to prove the excellence of our framework, we cross-validated our findings with recent research works, which used similar datasets and methods. According to the comparative results in table 7, the better performance of the proposed framework can be explained by the following core differences:

5.3.1. Combined Real-time Mitigation:

Compared to most of the literature (e.g., Mahar 2026 and Alhussain 2025), which only considers the detection accuracy as an independent step, we offer an integrated mechanism of defense. It takes on a response unit that can implement real-time mitigation measures.

This active response (as opposed to passive detection) will provide continuity of services and reduce the effects of attacks in real-time traffic.

5.3.2. High Computational Efficiency:

The proposed framework was able to reach a high accuracy of 99.94% with the use of only 6 optimized features. This is a very large technical advantage over other studies such as Alhussain (2025), which needed 79 features to attain a similar accuracy. Our model achieves a large reduction in processing time (latency) by dimensionality reduction of the data, even though the quality of the results is high, which is essential in real-time defense systems.

5.3.3. Optimized Hybrid Learning Architecture:

The BAT optimization algorithm was combined with the RF-XGB ensemble model in an incredibly effective way. We performed better than less complex hybrid models, like the system of Mahar (2026), which took advantage of RF and XGB, but without an optimization layer to select features. This proves that the BAT algorithm was able to optimize the hyperparameters and determine the most discriminative features of DDoS patterns.

5.3.4. Balance between Complexity and Scalability:

Complex architectures such as Deep Convolutional Neural Networks (DCNN) in Alohali (2023) show lower accuracy (92.14%), but our results indicate that a well-designed ensemble learning model can offer better stability and accuracy. The proposed model is highly scalable to a variety of network settings (SDN/IoT) and has much lower hardware demands than deep learning counterparts:

5.3.5. Comparative Analysis of False Positive Rate (FPR)

The proposed framework had an incredibly low FPR of about 0.06% (calculated on the basis of the 99.94% accuracy and the given metrics), which is significantly lower than the study by Sharma et al. (2025) that had the best result with an FPR of 0.45. This significant decrease suggests that the suggested model has a better capability to differentiate between normal network traffic and malicious actions, thus reducing the unnecessary disruptions in the system.

5.3.6 Advantage Over Multicomponent and Hybrid Models.

Although in both studies by Mahar et al. (2026) and Sharma et al. (2025), more advanced hybrid methods and blockchain integration were employed, the proposed framework was always very productive. Such a performance advantage is explained by the effectiveness of the BAT algorithm in the optimization of hyperparameters, along with the RF-XGB ensemble, which is an efficient way of balancing the trade-off between the variance and bias in the process of classification.

5.3.7. The Accuracy vs. Count of Features Trade-off.

One of the main strengths of the proposed framework is that with the small number of features, it has the potential to keep the error rate minimal. Conversely, Elsheh and Aboudabous (2026) employed the complete set of features but with a comparatively high FPR of 1.80. Our model manages to circumvent the curse of dimensionality and overfitting and achieves high accuracy in peak detection with much lower computational costs.

5.3.8. Research Contribution

This paper describes an AI-enhanced DDoS detection system intertwining machine learning, feature optimization, and real-time analytics into a single security system. The major findings of this paper are as follows:

- Hybrid Ensemble Detection Model

We suggest a powerful classification model that integrates both the random forest and the XGBoost based on soft voting, which will not only increase the detection rate but also generalize well to both the normal network traffic and the malicious ones.

- Feature Optimization

A smart feature selection algorithm, grounded on the bat algorithm, is proposed to help minimize the dimensionality of the dataset, represented by high-dimensional features of traffic, to a reasonable set of 6 features to enhance the efficiency of the computations and improve the performance of the models.

- Optimized DDoS detection

The input will consist in designing and implementing a hybrid BAT-RF-XGB model that is particularly geared towards DDoS detection. The high operational efficiency of the proposed approach is a high accuracy of 99.94 with only a few features, which provides an optimal compromise between the speed of detection and predictive accuracy.

- Mitigation System

The framework also includes an automated response mechanism that creates alerts in the format of JSON and invokes mitigation measures like blocking of IP addresses, TCP resets, and traffic filtering.

- Real-Time Analytics Engine

A real-time processing pipeline is created to conduct a live packet inspection, classification, and confidence scoring to detect DDoS attacks immediately.

- The False Positive Rate (FPR) decreased.

realized through the proposed framework makes it an ideal fit to a delicate environment, including Software-Defined Networking (SDN) and the Internet of Things (IoT). Reduction of false alarms will not only reduce the administrative workload on the security engineers but will also avoid the blocking of legitimate users due to false alarms, hence ensuring a continuity of services and efficiency in operations.

6. Conclusion

The current work introduced a state-of-the-art machine learning-based tool to detect Distributed Denial of Service (DDoS) attacks, which relies on the concept of an ensemble algorithm that incorporates the Random Forest and XGBoost classifiers. The results of the proposed model were that the model was more accurate and reliable in classifying network traffic with better detection rates and lower error compared to traditional standalone models, and, more importantly, integrating multiple models is more likely to respond more effectively to attacks in dynamic network settings. Also, the integration of the system with the interactive dashboard in real-time enhanced its usability through its ability to give constant network visibility and immediate security alerts, which enhanced quicker decision-making. Altogether, the analysis demonstrates that ensemble-learning-based approaches are a viable and efficient solution to enhancing the effectiveness of networks, as well as in the identification of contemporary cyberattacks, with the optimal balance between extreme accuracy, time efficiency, and the ability to respond to attacks. These properties render it very applicable to real-world implementation in contemporary network infrastructures in contrast to models that are more theoretical or detection-based.

7. Future Work

Although the proposed system has already brought positive outcomes, there are a number of future directions that could be used to improve its functionality and efficiency. To begin with, deep learning models, including Hybrid LSTM and CNN, can be considered to enhance the identification of intricate temporal patterns of DDoS attacks, especially low-rate attacks. second, the system can be further expanded to incorporate federated learning methods that will allow training on distributed data without centralization, enhancing privacy and security in the real world. moreover, the system can be improved by using more sophisticated feature selection algorithms or dimensionality reduction algorithms to make computations more efficient and decrease response time. lastly, it is advisable to test the system in real-world network environments and live traffic streams to test its scalability and performance in realistic operating conditions.

Acknowledgments

The authors gratefully acknowledge Al-Neelain University, Faculty of Engineering, Department of Computer Engineering, for providing the academic and research environment that supported this work in the field of Cybersecurity and Network Security. The authors sincerely thank their supervisors, *Dr. Ashraf Qism Al-Sayed and Dr. Abu Bakr Obeid, for their valuable guidance, constructive feedback, and continuous support throughout this research. The authors also extend their appreciation to all colleagues and research contributors* who participated in this work and provided valuable assistance, technical contributions, and constructive feedback toward the completion of this paper.

References

[1] D. Mahesh and T. S. Kumar (2024), “Machine Learning Algorithms for Detecting DDoS Attacks in Intrusion Detection Systems,” International Journal of Wireless and Mobile Technologies, vol. 14, no. 5, 2024, doi: 10.5815/ijwmt.2024.05.05.

[2] N. Berbiche and J. El Alami (2024), “For Robust DDoS Attack Detection by IDS: Smart Feature Selection and Data Imbalance Management Strategies,” Ingénierie des Systèmes d’Information, vol. 29, no. 4, 2024, doi: 10.18280/isi.290401.

[3] “A Novel Deep Learning-Based Intrusion Detection System for IoT DDoS Security,” Internet of Things Journal, Elsevier, 2024, doi: 10.1016/j.iot.2024.101336.

[4] A. A. Alashhab et al. (2024), “Enhancing DDoS Attack Detection and Mitigation in SDN Using an Ensemble Online Machine Learning Model,” IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3384398.

[5] R. Almuhanna and S. Dardouri (2025), “A Deep Learning/Machine Learning Approach for Anomaly-Based Network Intrusion Detection,” Frontiers in Artificial Intelligence, 2025, doi: 10.3389/frai.2025.1625891.

[6] A. Alabdulatif (2025), “A Novel Ensemble of Deep Learning Approach for Cybersecurity Intrusion Detection with Explainable AI,” Applied Sciences, vol. 15, 2025, doi: 10.3390/app15147984.

[7] M. Fatima et al (2024)., “ELIDS: Ensemble Feature Selection for Lightweight IDS Against DDoS Attacks in IoT Environment,” Future Generation Computer Systems, 2024, doi: 10.1016/j.future.2024.05.013.

[8] D. M. Dhanvijay et al (2025)., “Cyber Intrusion Detection Using Ensemble Deep Learning with Optimized Feature Sets,” Cyber Security and Applications, 2025, doi: 10.1016/j.csa.2025.100088.

[9] Abiramasundari, S., & Ramaswamy, V. (2025). Distributed denial-of-service (DDoS) attack detection using supervised machine learning algorithms. Scientific Reports, 15(13098). https://doi.org/10.1038/s41598-024-84879-y

[10] Y. Cheng and J. Feng (2026), "Ensemble Learning for Real-Time DDoS Detection in IoT Networks," Sci. Rep., vol. 16, no. 2, pp. 1–14, 2026.
[11] M. Oyucu, E. Demir, and A. Sari (2024), "Decision Tree Ensemble Approaches for SDN-SCADA DDoS Detection," Sensors, vol. 24, no. 1, p. 155.
[12] K. Lee, H. Kim, and J. Park (2025), "Real-Time DDoS Detection Using XGBoost in SDN Networks," J. Big Data, vol. 12, no. 3, pp. 1–13.
[13] P. Singh, R. Kumar, and S. Sharma (2025), "Hybrid Ensemble Learning for DDoS Detection Across Multiple Network Datasets," Multimed. Tools Appl., vol. 84, pp. 36789–36810.
[14] Elsheh, A., & Aboudabous, S. (2026). A Stacking Framework for Malware Detection with XGBoost and Random Forest. Journal of Academic Research (Applied Sciences), 30(1), 43–51.
[15] Alhussain, A. A., & Alsulami, B. S. (2025). DDoS Detection by Using Machine Learning. Journal of Information Systems Engineering and Management, 10(545). https://www.jisem-journal.com/
[16] Mahar, I. A., Aziz, K., Chakrabarti, P., Ahmed, N., Ladan, M., & Javed, Y. (2026). A hybrid machine learning approach for detecting DDoS attacks in software-defined networks. Scientific Reports, 16(6533). https://doi.org/10.1038/s41598-026-35458-w
[17] Naseera, F., & Jayapriya, K. N. (2025). Machine Learning-Driven Intrusion Detection for DDoS Attack Mitigation in Cyber-Physical Production Systems. Journal of Information Technology and Digital World, 7(1), 41. https://www.irojournals.com/itdw/
[18] Alashhab, A. A., Zahid, M. S., Isyaku, B., Elnour, A. A., Nagmeldin, W., Abdelmaboud, A., Abdullah, T. A. A., & Maiwada, U. (2024). Enhancing DDoS Attack Detection and Mitigation in SDN using an Ensemble Online Machine Learning Model. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3384398
[19] Maodah, N., et al. (2025). Boosting cloud security through hybrid LightGBM and CatBoost algorithms. Cloud Computing Journal, 13(1), 55-70.
[20] Maodah, N., et al. (2025). Boosting cloud security through hybrid LightGBM and CatBoost algorithms. Cloud Computing Journal, 13(1), 55-70.
[21] Suvra, S. S. (2025). Performance Analysis of Ensemble Machine Learning Algorithms for DDoS Attack Detection using Information Gain Feature Selection. International Journal of Advanced Computer Science and Applications (IJACSA), 16(1).
[22] Alohali, M. A., Elsadig, M., Hilal, A. M., & Motwakel, A. (2023). Emerging framework for attack detection in cyber-physical systems using heuristic-based optimization algorithm. PeerJ Computer Science, 9, e1596. https://doi.org/10.7717/peerj-cs.1596
[23] Sharma, P., Kalia, A., & Saini, H. (2025). A Hybrid Ensemble Framework for Intrusion Detection in IoT Using Blockchain-Integrated Fog Networks. Journal of Information Systems Engineering and Management, 10(25). https://www.jisem-journal.com/article/a-hybrid-ensemble-framework-for-intrusion-detection-in-iot-using-blockchain-integrated-fog-15632
[24] Hassan, A. I., El Reheem, E. A., & Guirguis, S. K. (2024). An entropy and machine learning based approach for DDoS attacks detection in software defined networks. Scientific Reports. https://doi.org/10.1038/s41598-024-xxxx
[25] Alashhab, A. A., et al. (2024). Online machine learning-based intrusion prevention system for DDoS attacks. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3384398
[26] Author(s). (2024). Online machine learning-based intrusion prevention system for DDoS attacks. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3384398

Nazar Mohamed received the B.Sc. and M.S.A degrees in Electronic Engineering Department of Computer Engineering from Gezira University in 2005 and 2016, respectively. During 2006-2026, he work in Computer teaching and Research Laboratory, Ministry of Physical Planning and Public Utilities 2006-20010, and work in Tabuk University as lecturer, E-Learning Coordinator and Academic Affairs Coordinator, Department of Computer Science2013 – 2016 and 2019 – 2026 .

Ashraf Gasim Elsid Abdalla received B.sc in Electronic Engineering from Technical university of Budapest (BME) in 1993 and M.sc and PhD from National University of Malaysia (UKM) in Electronic, El electrical and informatics in 1996 and 2000. He works as a research assistant and tutor in UKM, Lecturer in UPM, and MMU in Malaysia. He also works at Future University (FU) and Sudan University of Science and Technology in Khartoum. He was Dean of faculty of telecommunication and space technology and Associate professor in FU. Also he was Dean of Jordanian Sudanese college and member of many committees in the Ministry of Higher Education and Research in Sudan. He supervised many PhD. students and many master students. He published 32 scientific papers in international journals.

Dr.AbuBakr Obeid
Assistant professor in Electrical and Electronic Engineering.
Mainly in the area of Networking, Cyber security and Drones.

Hashim Elshafie is an expert in telecommunications engineering. He earned his Bachelor's degree in Electronics Engineering with a focus on Telecommunications from Sudan University of Science and Technology SUST. He then pursued advanced studies at Universiti Teknologi Malaysia UTM, where he completed both his Master's and PhD in Electrical Engineering with a specialization in Telecommunications. From 2013 to 2015, he served as a senior researcher at the MIMOS Center of Excellence in Telecommunications Engineering MIMOS-UTM in addition he was a member of the Telematic Research Group TRG in UTM. Moreover, he gained industrial experience at Malaysian Institute

of Microelectronic Systems (MIMOS Berhad), a National Applied R&D Centre in Malaysia, from 2011 to 2013. His research focus in Telecommunication and Networking Engineering, Satellite and Radar Communications, Antenna Engineering and Electromagnetics waves, Wireless Sensor Networks, Internet of Things (IoT), 5G/6G Cellular Systems, Artificial Intelligence (AI), Spectrum Sharing with Cognitive Radio techniques, and TV White Space Management Schemes. Since January 2016, Dr. Elshafie has been an Assistant Professor in the Department of Computer Engineering at King Khalid University KKU in the Kingdom of Saudi Arabia.

Dr. Rania Elkhidir holds a Ph.D. in Computer Science specializing in Artificial Intelligence. She is an Assistant Professor of Information Technology at the University of Hail, Saudi Arabia. Her research interests include AI, Machine Learning, Deep Learning, Computer Networks, and intelligent e-learning systems. She has extensive experience in academic leadership, research supervision, programming, web technologies, and computer systems.

Omer Elsier Tayfour Ahmed received his B.sc. in Computer Engineering from the University of Gezira, Sudan, in 2004, his M.Sc. in Computer Engineering and Networks in 2007, and his Ph.D. in Electrical Engineering from Universiti Teknologi Malaysia (UTM) in 2023. He has been working as a Lecturer in the Department of Computer Engineering at King Khalid University since 2009. His research interests include security testing and assessment, security incident management, machine learning, network processing architectures, and engineering project management. He has supervised several undergraduate graduation projects and published 10 journal papers, including publications indexed in Web of Science and Scopus. He has also participated in research projects and served as a reviewer for international journals.

Hashim Albasheer received his B.Sc. and M.Sc. degrees in Computer Science and his Ph.D. in Computer Science from Universiti Teknologi Malaysia (UTM). He is currently working at the College of Computer Science, King Khalid University, Saudi Arabia. He has over 15 years of academic and professional experience in cybersecurity and related fields. His research interests include cybersecurity, network security, intrusion detection systems, intrusion alert correlation.