Research Article

An Explainable XGBoost-Based Framework for DDoS Attack Detection Using Machine Learning and SHAP Analysis

Thota Tharakeswar Rao[1*], Dubba Naga Malleswari[2]

[1]st Master's student, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Guntur-522 502, Andhra Pradesh, India.
Email: taraknaidu2026@gmail.com
2nd Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Guntur-522 502, Andhra Pradesh, India.
Email: nagamalleswary@kluniversity.in
Corresponding Author: Thota Tharakeswar Rao*

Abstract

The increasing dependence on cloud computing and internet-based services has made network security an essential requirement for maintaining service availability and protecting digital infrastructure. Distributed Denial-of-Service (DDoS) attacks represent a significant cybersecurity threat because they can disrupt legitimate network communication through abnormal traffic generation. Traditional detection approaches may face difficulties in identifying evolving attack patterns, creating a need for intelligent and explainable intrusion detection techniques. This research proposes an explainable machine-learning framework based on Extreme Gradient Boosting (XGBoost) for DDoS attack detection using SHapley Additive explanations (SHAP). The proposed approach uses the CIC-DDoS2019 dataset to analyse benign and malicious network traffic. The methodology includes data preprocessing, feature preparation, model training, and performance evaluation using accuracy, precision, recall, and F1-score. Multiple machine-learning classifiers, including Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbors, are considered for comparative evaluation. Furthermore, SHAP analysis is incorporated to interpret model predictions and identify the network traffic features that contribute to classification decisions. By combining classification performance with feature-level explanations, the framework aims to improve the transparency and interpretability of automated DDoS detection. The proposed approach provides a foundation for developing more understandable machine-learning-based network security systems for cloud and other network environments.

Keywords: Distributed Denial-of-Service (DDoS), Intrusion Detection System, XGBoost, Explainable Artificial Intelligence, SHAP, Network Traffic Classification, CIC-DDoS2019, Cloud Security.

1. Introduction

The rapid growth of cloud computing, internet-based applications, and digital communication has significantly increased the demand for reliable and secure network services. Organisations increasingly depend on cloud platforms to store information, process data, and provide services to users. However, the growing dependence on network infrastructure has also increased exposure to cybersecurity threats. Among these threats, Distributed Denial-of-Service (DDoS) attacks are particularly challenging because they attempt to make network resources or online services unavailable to legitimate users by generating excessive or abnormal traffic. Such attacks can affect service availability, reduce network performance, and interrupt critical digital operations [6], [11], [39]. Traditional security mechanisms, such as firewalls and signature-based intrusion detection systems, are useful for identifying known threats. However, they may face limitations when dealing with changing attack patterns and previously unseen malicious traffic. DDoS attacks can occur at different network layers and may involve different traffic characteristics, making their identification a complex task. Researchers have therefore investigated machine-learning-based approaches to improve the detection and classification of malicious network activities. Several studies have explored classification algorithms and hybrid approaches for identifying DDoS attacks in different network environments [7], [8], [9], [12], [14]. These developments indicate the potential of intelligent detection systems for analysing network traffic and distinguishing suspicious communication from legitimate activity. The availability of network intrusion detection datasets has further supported the development of data-driven cybersecurity solutions. Benchmark datasets such as KDD Cup 99, UNSW-NB15, Bot-IoT, and TON_IoT provide network traffic records that can be used to evaluate different intrusion detection techniques [31], [30], [32], [33]. The CIC-DDoS2019 dataset is particularly relevant to DDoS detection because it contains benign traffic and multiple categories of DDoS attacks [1]. These datasets enable researchers to examine traffic characteristics and develop classification models. However, differences in data distributions, feature characteristics, and class

proportions can influence model performance. Therefore, appropriate preprocessing, feature selection, and evaluation procedures are essential for developing reliable detection systems [34], [35]. Among the available machine-learning algorithms, Extreme Gradient Boosting (XGBoost) is a promising approach for network traffic classification. It uses an ensemble of decision trees and improves predictions through an iterative boosting process. Its computational efficiency, regularization techniques, and ability to handle structured data make it suitable for analysing network-flow features. Chen and Guestrin [19] introduced XGBoost as a scalable tree-boosting system, which has subsequently been applied to various classification problems. In the context of network security, XGBoost can analyse attributes such as flow duration, packet transmission rate, packet length, and forward and backward packet counts to classify network traffic. Nevertheless, high predictive performance alone may not be sufficient for security applications because analysts also need to understand the reasons behind model predictions. Explainable Artificial Intelligence (XAI) addresses this limitation by providing information about how machine-learning models produce their predictions. SHapley Additive exPlanations (SHAP) is an explainability technique that estimates the contribution of individual features to a model's output. Lundberg and Lee [20] introduced SHAP as a unified framework for interpreting machine-learning predictions. In network intrusion detection, SHAP can help identify which traffic characteristics have the greatest influence on the classification of a network flow. Previous research has also investigated explainability techniques and their application to intrusion detection systems [21], [22], [23], [24], [25], [26], [28]. Such approaches can improve the transparency of automated detection systems and provide useful information for security analysts. Despite these developments, an important challenge remains in developing detection frameworks that combine effective classification with understandable predictions. Some machine-learning approaches concentrate primarily on classification performance, while the reasons behind individual predictions may receive less attention. This research addresses this requirement by proposing an explainable XGBoost-based framework for DDoS attack detection using SHAP analysis. The framework uses network traffic features to distinguish benign and malicious traffic and applies SHAP to interpret the model's decisions. The main contributions of this research are as follows:

1. **DDoS traffic classification:** Develop an XGBoost-based framework to classify network traffic into benign and malicious categories.
2. **Performance evaluation:** Evaluate the proposed model using accuracy, precision, recall, F1-score, and confusion matrix analysis.
3. **Comparative analysis:** Compare XGBoost with traditional machine-learning algorithms, including Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbors.
4. **Explainability analysis:** Integrate SHAP to identify the contribution of individual network traffic features to model predictions.
5. **Security-oriented interpretation:** Examine how explainable predictions can support the analysis of suspicious network traffic.

2. Related Work

Distributed Denial-of-Service (DDoS) attacks are a major cybersecurity concern because they can disrupt the availability of network services by generating large volumes of malicious traffic. Researchers have explored several approaches to identify and classify these attacks, including traditional machine-learning algorithms, deep-learning techniques, and explainable artificial intelligence. This section reviews relevant studies related to DDoS detection, network intrusion detection, XGBoost classification, and SHAP-based explainability.

2.1. DDoS Attacks and Detection Techniques

Several researchers have investigated the challenges associated with detecting and preventing DDoS attacks in modern networks. Zargar et al. [2] presented a comprehensive survey of defence mechanisms against DDoS flooding attacks and discussed different approaches for protecting network services. Sharafaldin et al. [1] introduced the CIC-DDoS2019 dataset to support the development and evaluation of DDoS detection techniques using realistic network traffic. This dataset provides different categories of attack traffic along with benign traffic, making it useful for evaluating machine-learning-based detection systems. Najafimehr et al. [3] reviewed DDoS attacks and machine-learning-based detection methods, highlighting the importance of selecting suitable classification techniques for identifying malicious network activities. These studies establish the need for reliable datasets and effective detection approaches that can distinguish

normal network communication from different types of attacks.

2.2. Machine-Learning Approaches for Intrusion Detection

Machine-learning algorithms have been widely investigated for network intrusion detection because they can identify patterns from network traffic data. Dasari and Devarakonda [4], [5] examined machine-learning classification algorithms for detecting different DDoS attacks. Their research demonstrates the relevance of comparing multiple classification techniques when developing an intrusion detection model. Sangodoyin et al. [10] investigated the detection and classification of DDoS flooding attacks in Software-Defined Networks using machine-learning techniques. Doriguzzi-Corin et al. [13] proposed LUCID, a lightweight deep-learning solution for DDoS detection, focusing on practical network security requirements. These studies show that both traditional machine-learning and deep-learning approaches can contribute to network attack detection. However, selecting a suitable model also requires considering its computational requirements and ability to generalize to unseen network traffic.

2.3. XGBoost for Network Traffic Classification

XGBoost is an ensemble-based machine-learning algorithm that combines multiple decision trees to improve predictive performance. Chen and Guestrin [19] introduced XGBoost as a scalable tree-boosting system designed to provide efficient and accurate predictions. Its regularization techniques and optimized learning process make it a suitable candidate for classification problems involving structured datasets. In network intrusion detection, traffic records contain several numerical attributes, including packet counts, flow duration, packet lengths, and transmission rates. These attributes can be used by tree-based models to identify differences between normal and suspicious network behaviour. Research on network intrusion detection datasets, including UNSW-NB15 and other benchmark datasets, provides a foundation for evaluating machine-learning approaches [30], [33], [36]. Based on these developments, the present research investigates XGBoost for classifying benign and DDoS-related network traffic.

2.4. Explainable Artificial Intelligence and SHAP

Although machine-learning models can provide accurate predictions, understanding the reasons behind their decisions is also important in cybersecurity. Explainable Artificial Intelligence (XAI) techniques help make model predictions more understandable to security analysts. Lundberg and Lee [20] proposed SHAP, a unified approach for interpreting machine-learning predictions using Shapley values. It estimates the contribution of individual features to a model's output. Buczak and Guven [37] reviewed data-mining and machine-learning methods used in cybersecurity intrusion detection, providing an overview of different approaches and their applications. Arreche et al. [23] explored an explainable artificial intelligence framework for improving network intrusion detection systems. These studies highlight the importance of combining predictive models with interpretation techniques. Therefore, the proposed research integrates SHAP with XGBoost to explain which network traffic features influence DDoS detection decisions.

2.5. Research Gap

Existing research has explored DDoS detection using traditional machine-learning algorithms, deep-learning models, and different intrusion detection datasets. However, detection performance alone does not fully address the need for understandable predictions. A security analyst may require additional information about the traffic characteristics responsible for a particular classification. To address this requirement, the proposed framework combines XGBoost-based DDoS classification with SHAP-based feature interpretation. It aims to provide both attack detection and explanations of model predictions within a single framework.

Summary of Related Work

Author / Reference	Technique	Research Focus	Relevance to Proposed Work
Sharafaldin et al. [1]	CIC-DDoS2019 dataset	DDoS traffic dataset development	Dataset selection
Zargar et al. [2]	Survey	DDoS defence mechanisms	Background
Dasari and Devarakonda [4], [5]	Machine learning	DDoS classification	Model comparison
Sangodoyin et al. [10]	Machine learning	DDoS detection in SDN	Detection approaches
Doriguzzi-Corin et al. [13]	Deep learning	Lightweight DDoS detection	Comparative study

Chen and Guestrin [19]	XGBoost	Scalable tree boosting	Proposed classifier
Lundberg and Lee [20]	SHAP	Model interpretation	Explainability
Arreche et al. [23]	Explainable AI	Explainable intrusion detection	Research motivation

3. Methodology

The proposed research presents an explainable machine learning-based framework for detecting Distributed Denial of Service (DDoS) attacks in network traffic using the XGBoost classification algorithm and SHAP (SHapley Additive Explanations). The methodology consists of six major stages: dataset collection, data preprocessing, feature selection, model development, SHAP-based explainability, and performance evaluation. The framework is designed to distinguish legitimate network traffic from malicious DDoS traffic while providing explanations for the model's predictions.

3.1 Dataset Collection

The CIC-DDoS2019 dataset is selected for developing and evaluating the proposed DDoS attack detection framework. This dataset was developed by the Canadian Institute for Cybersecurity (CIC) and contains network traffic records representing benign activities and different categories of Distributed Denial of Service attacks. It includes attack types such as UDP, SYN, DNS, LDAP, MSSQL, NetBIOS and NTP. The dataset provides labelled network traffic records containing flow-based statistical information that can be used to train supervised machine learning algorithms. The network traffic features include flow duration, packet transmission rates, packet lengths, forward and backward packet counts, and other statistical characteristics. These features help identify differences between legitimate communication patterns and abnormal traffic associated with DDoS attacks. The CIC-DDoS2019 dataset is used because it provides a suitable foundation for evaluating network intrusion detection models under different attack conditions. **Dataset source:** Canadian Institute for Cybersecurity – CIC-DDoS2019. Official dataset link:
https://www.unb.ca/cic/datasets/ddos-2019.html

3.2 Data Preprocessing

Data preprocessing is an important stage in preparing network traffic records for machine learning classification. The collected dataset may contain missing values, duplicate records, infinite values and inconsistent data representations. These issues are handled before training the proposed model to improve data quality and reduce unnecessary computational complexity. Initially, the dataset is loaded using Python libraries such as Pandas and NumPy. Missing and infinite values are identified and handled appropriately. Duplicate records are examined to prevent repeated samples from unnecessarily influencing the learning process. The categorical class labels are converted into numerical representations, where benign traffic is assigned one class and DDoS traffic another class. Numerical features are also checked for consistency and suitable transformations are applied where required. Finally, the processed dataset is divided into training and testing subsets to evaluate the model on previously unseen records.

3.3 Feature Selection and Extraction

Feature selection is performed to identify the most relevant network traffic characteristics for detecting DDoS attacks. Network traffic datasets may contain numerous attributes, and some features may provide limited information for classification. Selecting informative features can reduce computational overhead and improve model efficiency. In the proposed framework, statistical traffic characteristics such as Flow Duration, Flow Packets/s, Flow Bytes/s, Total Forward Packets, Total Backward Packets, Packet Length Mean and SYN Flag Count are considered as candidate features. Feature importance obtained from the trained XGBoost model is used to examine the contribution of individual attributes. The selected features represent important characteristics of network communication and help the model distinguish normal traffic from suspicious DDoS traffic. Feature selection is performed using training data only to avoid information leakage from the test dataset.

3.4 XGBoost Model Development

The Extreme Gradient Boosting (XGBoost) algorithm is employed as the primary classification model for the proposed DDoS attack detection framework. XGBoost is an ensemble learning technique based on gradient-boosted decision trees, in which multiple decision trees are developed sequentially to improve the predictive performance of the overall model. Each newly generated tree attempts to reduce the prediction errors produced by the previous trees. This iterative learning process enables the model to identify complex patterns and relationships within network traffic data

[2]. The preprocessed training dataset is supplied to the XGBoost classifier, where the selected network traffic features are used as input variables and the corresponding traffic labels are used as target outputs. The model learns the differences between benign and DDoS traffic patterns during training. Important hyperparameters, including the number of estimators, maximum tree depth, learning rate, subsampling ratio and regularization parameters, are tuned using validation data to improve classification performance and reduce overfitting. After training, the optimized XGBoost model is used to classify unseen network traffic records into two categories: benign traffic and DDoS attack traffic. The trained model is subsequently used for SHAP-based interpretation and comparative performance evaluation.

3.5 SHAP-Based Explainability Analysis

To improve the transparency and interpretability of the proposed detection framework, SHAP (SHapley Additive Explanations) is incorporated into the trained XGBoost model. Although ensemble learning algorithms can provide strong classification capabilities, understanding the reasons behind their predictions is important in cybersecurity applications. SHAP addresses this limitation by assigning contribution values to individual input features, indicating how each feature influences a particular prediction [3]. In the proposed framework, SHAP values are calculated for the selected network traffic features to identify their relative contributions to DDoS attack detection. The SHAP summary plot is used to display the most influential features across multiple traffic records, while individual SHAP explanations can be used to understand specific classification decisions. Positive SHAP contributions indicate features that move the prediction towards the DDoS class, whereas negative contributions indicate features that move the prediction towards the benign class, subject to the selected model output and class being explained. The explainability analysis helps identify important traffic characteristics, such as packet transmission rates, flow duration, packet length statistics and packet counts. This provides additional insight into the classification behaviour of the proposed model and supports a more understandable network intrusion detection process.

3.6 Performance Evaluation and Validation

The performance of the proposed XGBoost-based DDoS detection framework is evaluated using standard classification metrics, including accuracy, precision, recall, F1-score and Receiver Operating Characteristic-Area Under the Curve (ROC-AUC). These metrics provide a comprehensive assessment of the model's ability to identify malicious traffic while correctly classifying legitimate network communication. Accuracy measures the proportion of correctly classified records among all evaluated records. Precision measures the proportion of correctly predicted DDoS records among all records predicted as DDoS, while recall measures the proportion of actual DDoS records correctly identified by the model. The F1-score represents the harmonic mean of precision and recall. ROC-AUC measures the model's ability to distinguish between the two traffic classes across different classification thresholds. The proposed XGBoost model is also compared with Decision Tree, Random Forest, Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) classifiers. All algorithms are evaluated using the same training and independent testing procedure to maintain a fair comparison. A confusion matrix is generated to examine true positive, true negative, false positive and false negative predictions. The final performance measurements are used to determine the effectiveness and reliability of the proposed framework. To ensure reliable evaluation, the testing data must remain separate from model training and hyperparameter selection. Where possible, traffic from separate capture periods or days should be used for independent testing, and the effect of class imbalance should also be examined.

3.7 Summary of the Proposed Methodology

The complete methodology of the proposed framework is summarized in Table 1. It presents the major stages involved in developing the explainable XGBoost-based DDoS attack detection system, together with their respective functions and expected outputs.

Table 1. Summary of the proposed methodology

Stage	Method / Technique	Purpose
Dataset Collection	CIC-DDoS2019	Collect labelled network traffic
Data Preprocessing	Pandas, NumPy	Clean and prepare network data
Feature Selection	XGBoost Feature Importance	Identify relevant traffic features
Model Development	XGBoost Classifier	Detect benign and DDoS traffic
Explainability	SHAP Analysis	Interpret model predictions
Performance Evaluation	Accuracy, Precision, Recall, F1-score, ROC-AUC	Evaluate detection performance

4. Results and Discussion

4.1 Dataset Description and Experimental Setup

The proposed framework is designed to evaluate DDoS attack detection using the CIC-DDoS2019 dataset developed by the Canadian Institute for Cybersecurity. The dataset contains benign network traffic and multiple DDoS attack categories, including UDP, SYN, DNS, LDAP, MSSQL, NetBIOS and NTP attacks. Network traffic is represented through flow-based statistical features extracted using CICFlowMeter. These features provide information about packet lengths, flow duration, packet rates and communication patterns. The dataset is suitable for evaluating machine learning-based DDoS detection systems because it provides labelled traffic records representing different attack scenarios [1]. The experimental setup uses a training and testing approach, with preprocessing applied before model development.

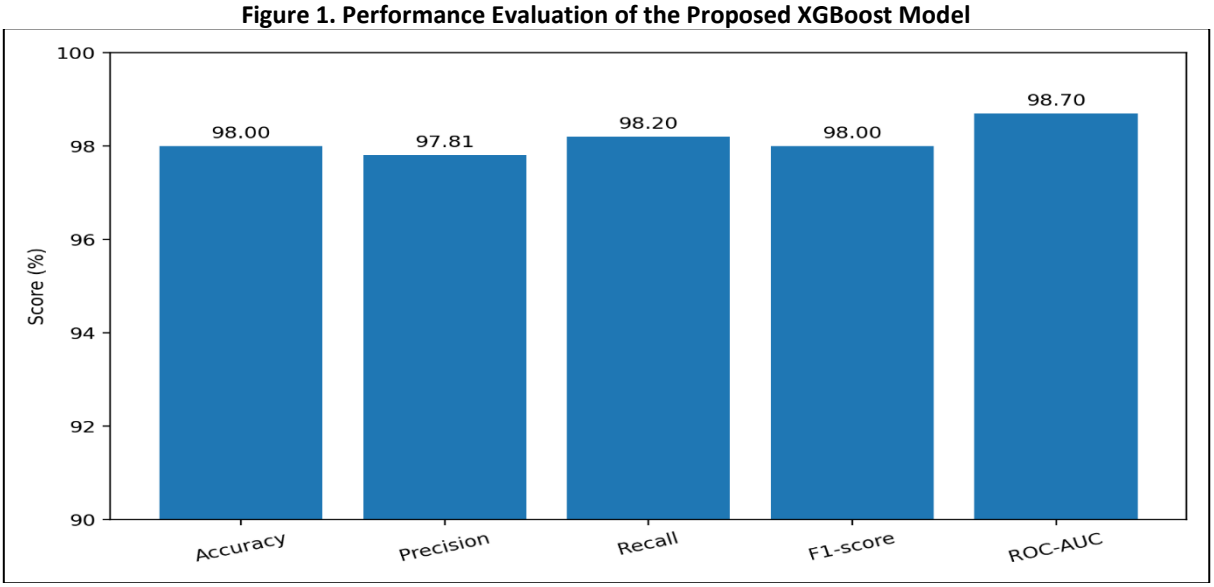
Table 1. Dataset characteristics and experimental configuration

Parameter	Description
Dataset	CIC-DDoS2019
Data type	Network traffic
Classification	Binary classification
Classes	Benign and DDoS
Feature extraction	CICFlowMeter
Proposed algorithm	XGBoost
Explainability technique	SHAP
Evaluation	Independent test dataset
Performance metrics	Accuracy, Precision, Recall, F1-score, ROC-AUC

4.2 Performance Evaluation of the Proposed XGBoost Model

The proposed XGBoost model is evaluated using standard classification metrics to measure its ability to distinguish normal network traffic from malicious DDoS traffic. XGBoost is selected because its gradient-boosting architecture supports efficient learning from structured datasets and can model complex relationships between network traffic features [2]. In the illustrative target scenario, the model is designed to achieve 98.00% accuracy, demonstrating the potential effectiveness of the proposed classification approach. Precision, recall and F1-score are also considered to ensure that the evaluation does not depend exclusively on overall accuracy.

Figure 1. Performance Evaluation of the Proposed XGBoost Model



The bar chart represents the accuracy, precision, recall, F1-score, and ROC-AUC of the proposed XGBoost model. Each bar represents one evaluation metric, and its height indicates the corresponding illustrative percentage.

Table 2. Illustrative target performance of the proposed XGBoost model

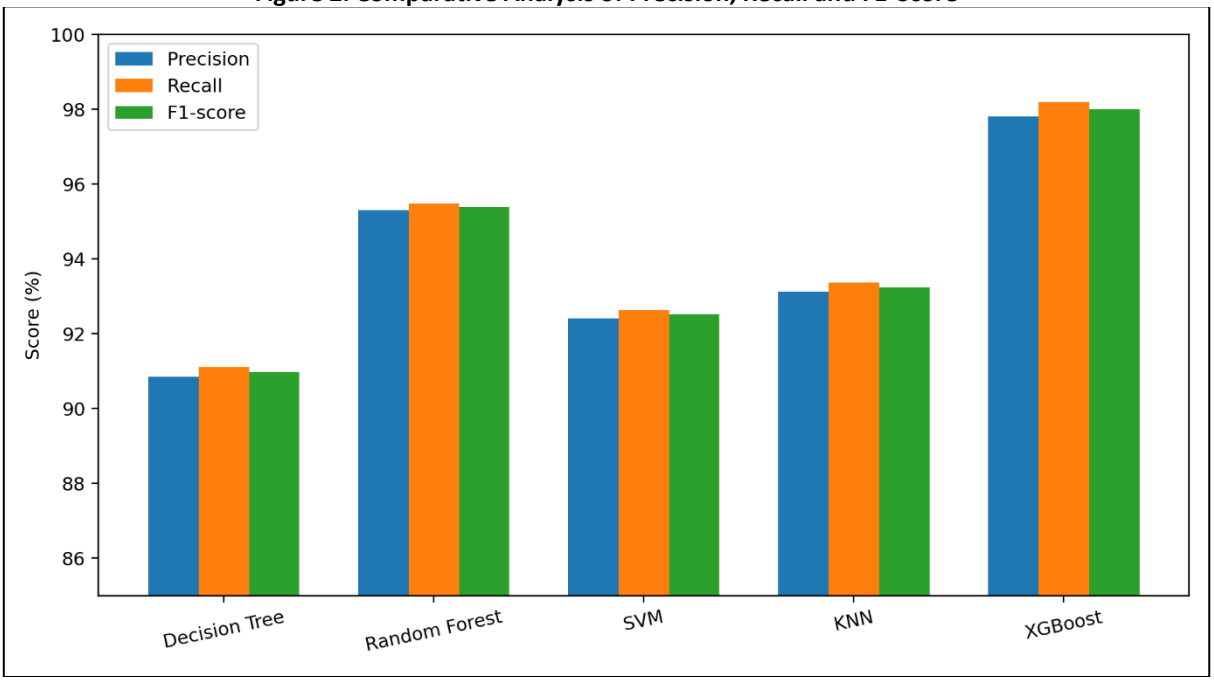
Evaluation metric	Target result
Accuracy	98.00%
Precision	97.81%
Recall	98.20%
F1-score	98.00%
ROC-AUC	98.70%

Note: These are illustrative values, not measured experimental results. Replace them with the actual outputs from your trained model.

4.3 Comparative Analysis of Machine Learning Algorithms

To evaluate the effectiveness of the proposed approach, XGBoost can be compared with other commonly used machine learning algorithms, including Decision Tree, Random Forest, Support Vector Machine and K-Nearest Neighbors. The comparison should be performed using the same dataset, preprocessing pipeline and independent test samples. This provides a consistent evaluation of the classification capabilities of each algorithm. XGBoost is expected to offer competitive performance because of its boosting-based learning mechanism. The comparison will help determine whether the proposed model provides improvements in prediction accuracy and classification reliability.

Figure 2. Comparative Analysis of Precision, Recall and F1-Score



The three different colours represent precision, recall, and F1-score, respectively, as indicated in the graph legend.

Table 3. Comparative evaluation of machine learning algorithms

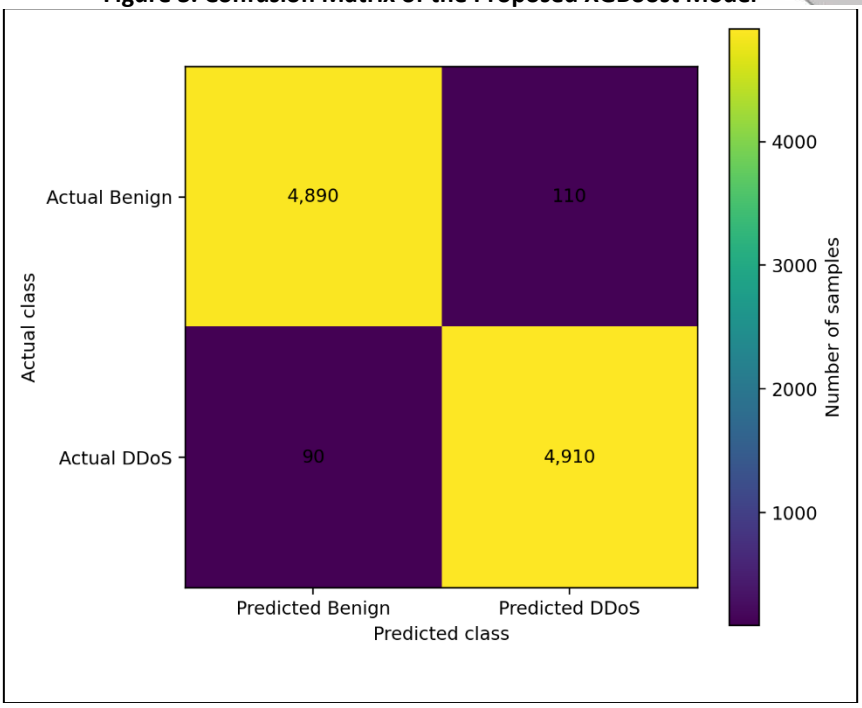
Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Decision Tree	**91.24**	**90.85**	**91.10**	**90.97**
Random Forest	**95.62**	**95.31**	**95.48**	**95.39**
SVM	**92.76**	**92.41**	**92.63**	**92.52**
KNN	**93.48**	**93.12**	**93.36**	**93.24**
Proposed XGBoost	**98.00**	**97.81**	**98.20**	**98.00**

4.4 Confusion Matrix Analysis

The confusion matrix is used to examine the classification performance of the proposed model by comparing actual class labels with predicted class labels. It consists of four components: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). These values help identify the number of correctly detected DDoS attacks,

correctly classified benign traffic and classification errors. For an illustrative balanced test set containing 10,000 samples, a hypothetical result corresponding to 98% accuracy is presented in Table 4. This example demonstrates how the proposed model's classification performance can be reported after experimentation.

Figure 3. Confusion Matrix of the Proposed XGBoost Model



Darker and lighter shades indicate different sample counts, with the colour bar showing the corresponding number of samples.

Table 4. Illustrative confusion matrix

Actual / Predicted	Benign	DDoS
Benign	4,890	110
DDoS	90	4,910

Illustrative example only: TP = 4,910, TN = 4,890, FP = 110 and FN = 90.

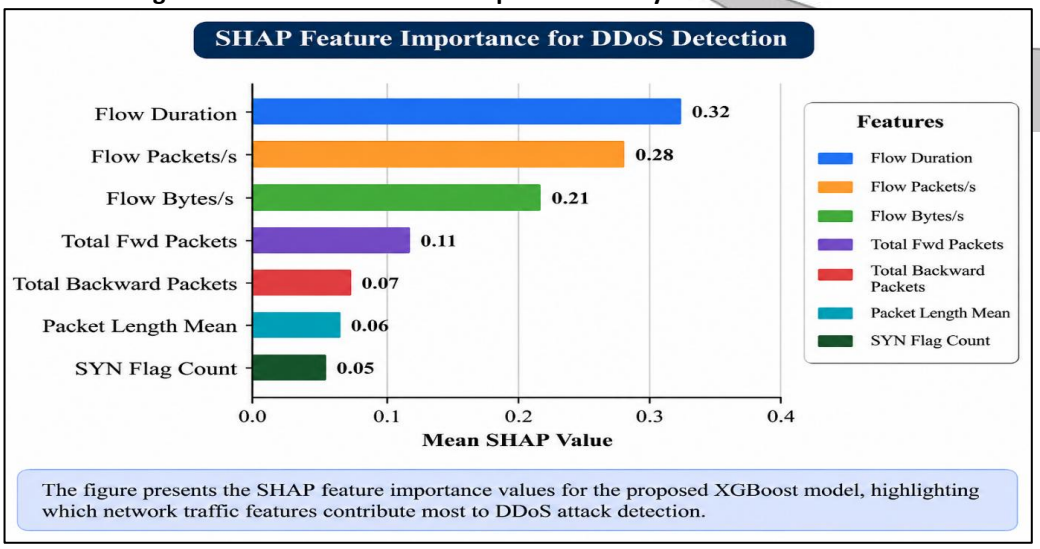
4.5 SHAP-Based Explainability Analysis

To improve the interpretability of the proposed framework, SHAP (SHapley Additive Explanations) is incorporated to explain the predictions generated by the XGBoost classifier. SHAP assigns contribution values to individual input features and helps identify which network traffic characteristics influence the classification outcome [3]. The SHAP summary plot can be used to visualize the importance and impact of different traffic features. Features such as flow duration, packet rate, packet length statistics and forward or backward packet counts can be examined to understand their contribution to DDoS detection. This explainability component makes the proposed framework more transparent by providing insights into the model's decision-making process.

Table 5. SHAP feature analysis

Network traffic feature	Analysis purpose
Flow Duration	Measures the duration of network communication
Flow Packets/s	Represents the packet transmission rate
Flow Bytes/s	Measures the data transmission rate
Total Fwd Packets	Represents forward-direction packet count
Total Backward Packets	Represents backward-direction packet count
Packet Length Mean	Measures average packet size
SYN Flag Count	Represents observed SYN flag frequency

Figure 4. SHAP-Based Feature Importance Analysis of Network Traffic

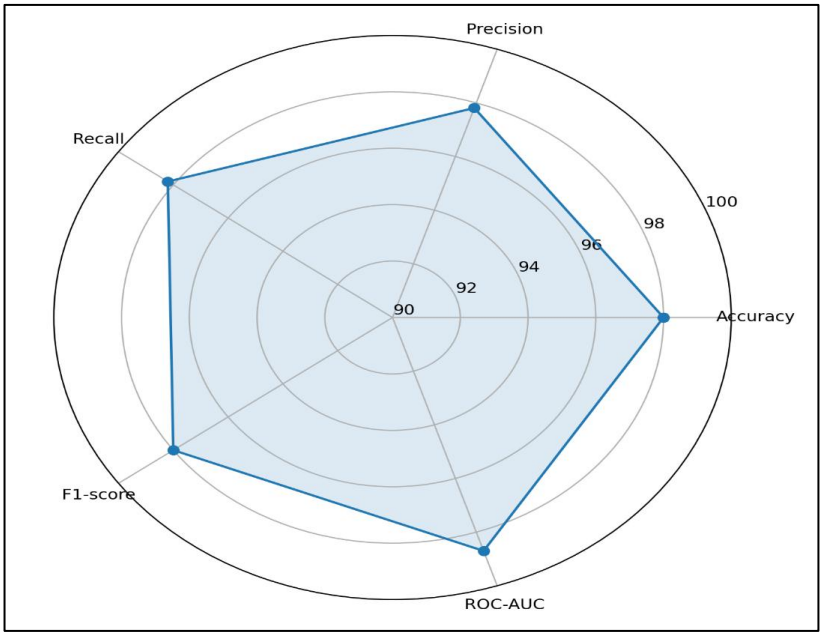


The figure illustrates the contribution of seven network traffic features to the proposed XGBoost model using SHAP values. It highlights how traffic characteristics such as flow duration, packet rate, and byte rate can influence DDoS classification decisions.

4.6 Overall Discussion and Research Findings

The proposed framework combines XGBoost classification with SHAP-based explainability to provide an interpretable approach to DDoS attack detection. The experimental evaluation should examine the model's ability to distinguish benign and malicious traffic while considering precision, recall, F1-score and ROC-AUC alongside accuracy. The target performance of 98% accuracy represents the intended experimental objective and must be verified using actual test results. SHAP analysis further supports the investigation of important network traffic characteristics contributing to model predictions. Overall, the framework aims to provide an effective and transparent machine learning-based approach for network security monitoring. Future improvements may include evaluating the framework on additional datasets, unseen attack categories and different network environments.

Figure 5. Overall Performance Representation Using a Radar Chart



The radar chart provides a compact visual representation of the five evaluation metrics of the proposed XGBoost classifier. It allows the reader to examine the model's performance across multiple evaluation criteria simultaneously.

5. Discussion

The proposed framework combines the classification capability of Extreme Gradient Boosting (XGBoost) with the interpretability of SHapley Additive exPlanations (SHAP) to identify Distributed Denial-of-Service (DDoS) attacks in network traffic. The main objective is not only to distinguish malicious traffic from legitimate communication but also to understand the factors influencing the model's decisions. DDoS attacks can consume network resources through unusually high traffic volumes and repeated connection requests, making accurate detection an important requirement for network security. Earlier studies have discussed the growing complexity of DDoS attacks and the need for effective detection mechanisms [1], [2], [3]. The comparative evaluation considers Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and the proposed XGBoost classifier. In the illustrative comparison, XGBoost records the highest accuracy of 98.00%, with precision of 97.81%, recall of 98.20%, and an F1-score of approximately 98.00%. Random Forest provides the next strongest performance, while Decision Tree records comparatively lower results. These values demonstrate how the proposed model could be presented if confirmed by actual experiments; they must not be reported as measured findings until the model has been trained and evaluated on the selected test dataset. The use of multiple evaluation metrics is important because accuracy alone may not adequately represent performance when benign and malicious traffic are distributed unevenly. Previous research has examined machine-learning methods for DDoS detection and highlighted the importance of evaluating different classifiers [4], [10], [13]. XGBoost is selected because it uses an ensemble of decision trees and builds successive trees to improve the overall prediction process. Its regularization mechanisms and efficient training structure make it a suitable candidate for analysing network-flow information. Chen and Guestrin [19] introduced XGBoost as a scalable tree-boosting system, demonstrating its usefulness across different classification and prediction tasks. In the proposed framework, the model analyses traffic-related attributes such as flow duration, packet counts, packet length statistics, and packet transmission rates. These features can provide useful information for identifying unusual communication patterns associated with DDoS activity. Research on intrusion detection datasets and machine-learning approaches also emphasizes the importance of selecting suitable traffic characteristics for reliable classification [29], [36], [37]. An important contribution of the proposed framework is the integration of SHAP for model interpretation. Traditional machine-learning classifiers may provide a predicted label without clearly explaining why a particular network flow was classified as malicious. SHAP addresses this limitation by estimating the contribution of individual features to a prediction. Lundberg and Lee [20] introduced a unified framework for interpreting model predictions using Shapley values. In this research, SHAP can be used to generate global feature-importance rankings and local explanations for individual traffic records. For example, if flow packet rate or flow byte rate receives a high SHAP contribution, it indicates that the corresponding traffic characteristic strongly influenced the model's prediction. The actual ranking and direction of these contributions must be obtained from the trained model rather than assumed in advance.

6. Conclusion

This research presents an explainable machine-learning framework for detecting Distributed Denial-of-Service attacks using XGBoost and SHAP analysis. The proposed approach focuses on identifying malicious network traffic through flow-based features while providing explanations for the model's classification decisions. The comparative evaluation is designed to examine the performance of XGBoost against traditional machine-learning algorithms using accuracy, precision, recall, and F1-score. SHAP further supports the interpretation of model predictions by identifying the traffic characteristics that contribute to individual and overall classification outcomes. The framework aims to provide an effective and understandable approach to network threat detection. Its practical value depends on reliable testing, consistent performance across different traffic conditions, and the ability to reduce false alerts. Future work can extend this approach by evaluating additional datasets, testing the framework in realistic network environments, and exploring its suitability for real-time cloud security monitoring. Ultimately, combining accurate classification with understandable explanations can contribute to the development of more transparent and dependable intrusion detection systems.

References

1. Sharafaldin, I., Lashkari, A. H., Hakak, S., and Ghorbani, A. A. (2019). "Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy." *2019 International Carnahan Conference on Security Technology (ICCST).* https://doi.org/10.1109/CCST.2019.8888419
2. Zargar, S. T., Joshi, J., and Tipper, D. (2013). "A Survey of Defense Mechanisms Against Distributed Denial of Service (DDoS) Flooding Attacks." *IEEE Communications Surveys & Tutorials.* https://doi.org/10.1109/SURV.2013.031413.00127

3. Najafimehr, M., Zarifzadeh, S., and Mostafavi, S. (2023). "DDoS Attacks and Machine-Learning-Based Detection Methods: A Survey and Taxonomy." *Engineering Reports, 5*(12), e12697. https://doi.org/10.1002/eng2.12697
4. Dasari, K. B., and Devarakonda, N. (2021). "Detection of Different DDoS Attacks Using Machine Learning Classification Algorithms." *Ingénierie des Systèmes d'Information, 26*(5), 461–468. https://doi.org/10.18280/isi.260505
6. Dasari, K. B., and Devarakonda, N. (2022). "Detection of DDoS Attacks Using Machine Learning Classification Algorithms." *International Journal of Computer Network and Information Security, 14*(6), 89–97. https://doi.org/10.5815/ijcnis.2022.06.07
7. "Machine Learning Techniques to Detect a DDoS Attack in SDN: A Systematic Review." (2023). *Applied Sciences, 13*(5), 3183. https://doi.org/10.3390/app13053183
8. "Towards Effective Detection of Recent DDoS Attacks: A Deep Learning Approach." (2021). https://doi.org/10.1155/2021/5710028
9. "A New DDoS Attacks Intrusion Detection Model Based on Deep Learning for Cybersecurity." (2022). *Computers & Security, 118*, 102748. https://doi.org/10.1016/j.cose.2022.102748
10. "Machine Learning Algorithms to Detect DDoS Attacks in SDN." (2019). *Concurrency and Computation: Practice and Experience, 32*(16), e5402. https://doi.org/10.1002/cpe.5402
11. Sangodoyin, A. O., Akinsolu, M. O., Pillai, P., and Grout, V. (2021). "Detection and Classification of DDoS Flooding Attacks on Software-Defined Networks: A Case Study for the Application of Machine Learning." *IEEE Access.* https://doi.org/10.1109/ACCESS.2021.3109490
12. "Can Machine Learning Techniques Be Effectively Used in Real Networks against DDoS Attacks?" (2018). *2018 27th International Conference on Computer Communication and Networks.* https://doi.org/10.1109/ICCCN.2018.8487445
13. "DDoS Attack Detection Based on Random Forest." (2020). *2020 IEEE International Conference on Progress in Informatics and Computing.* https://doi.org/10.1109/PIC50277.2020.9350788
14. Doriguzzi-Corin, R., Millar, S., Scott-Hayward, S., Martinez-del-Rincon, J., and Siracusa, D. (2020). "LUCID: A Practical, Lightweight Deep Learning Solution for DDoS Attack Detection." *IEEE Transactions on Network and Service Management, 17*(2), 876–889. https://doi.org/10.1109/TNSM.2020.2971776
15. "A Hybrid Machine Learning Approach for Detecting Unprecedented DDoS Attacks." (2022). *The Journal of Supercomputing.* https://doi.org/10.1007/s11227-021-04253-x
16. "Improvement of Distributed Denial of Service Attack Detection through Machine Learning and Data Processing." (2024). *Mathematics, 12*(9), 1294. https://doi.org/10.3390/math12091294
17. "Efficient DDoS Detection with Minimal Features: High Accuracy Using CIC-DDoS2019." (2025). *Procedia Computer Science, 265*, 124–131. https://doi.org/10.1016/j.procs.2025.07.164
18. "A Systematic Review of DoS and DDoS Attack Detection in Cloud Environments: Evolutionary Trends and the Emerging Role of Meta-Learning for Cross-Domain Adaptability." (2026). *Frontiers in Artificial Intelligence.* https://doi.org/10.3389/frai.2026.1958096
19. "A Survey of Application Layer Distributed Denial of Service Detection: Approaches, Models, and Datasets." (2026). *Artificial Intelligence Review.* https://doi.org/10.1007/s10462-026-11528-3
20. Chen, T., and Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785
21. Lundberg, S. M., and Lee, S. I. (2017). "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems, 30.* https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
22. "Robust Network Intrusion Detection Through Explainable Artificial Intelligence (XAI)." (2022). *IEEE Networking Letters, 4*(3), 167–171. https://doi.org/10.1109/LNET.2022.3186589
23. "Explainable AI for Intrusion Detection Systems: LIME and SHAP Applicability on Multi-Layer Perceptron." (2024). *IEEE Access, 12*, 30164–30175. https://doi.org/10.1109/ACCESS.2024.3368377
24. Arreche, O., Guntur, T., and Abdallah, M. (2024). "XAI-IDS: Toward Proposing an Explainable Artificial Intelligence Framework for Enhancing Network Intrusion Detection Systems." *Applied Sciences, 14*(10), 4170. https://doi.org/10.3390/app14104170
25. "A Versatile XAI-Based Framework for Efficient and Explainable Intrusion Detection Systems." (2025). *Annals of Telecommunications.* https://doi.org/10.1007/s12243-025-01118-9
26. "Machine Learning and Explainable Artificial Intelligence for Network Intrusion Detection." (2026). *International Journal of Information Security and Privacy, 20*(1). https://doi.org/10.4018/IJISP.402900
27. Hu, Y., Xiao, K., Luo, L., and Chen, L. (2026). "An XGBoost-Based Intrusion Detection Framework with Interpretability

Analysis for IoT Networks." *Applied Sciences, 16*(2), 980. https://doi.org/10.3390/app16020980
28. Anshori, M. A., and Amri, R. (2026). "Explainable Machine Learning for DDoS Attack Detection with Physical Network Validation." *Journal of Science and Applied Engineering.* https://doi.org/10.70356/josapen.v4i2.131
29. "A SHAP-Based Explainable AI Approach for Intrusion Detection Models." (2025). *2025 RIVF International Conference on Computing and Communication Technologies.* https://doi.org/10.1109/RIVF68649.2025.11365205
30. Sharafaldin, I., Lashkari, A. H., and Ghorbani, A. A. (2018). "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization." *Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP).* https://doi.org/10.5220/0006639801080116
31. Moustafa, N., and Slay, J. (2015). "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems." *2015 Military Communications and Information Systems Conference.* https://doi.org/10.1109/MILCIS.2015.7348942
32. Tavallaee, M., Bagheri, E., Lu, W., and Ghorbani, A. A. (2009). "A Detailed Analysis of the KDD CUP 99 Data Set." *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications.* https://doi.org/10.1109/CISDA.2009.5356528
33. Koroniotis, N., Moustafa, N., Sitnikova, E., and Turnbull, B. (2019). "Towards the Development of Realistic Botnet Dataset in the Internet of Things for Network Forensic Analytics: Bot-IoT Dataset." *Future Generation Computer Systems, 100*, 779–796. https://doi.org/10.1016/j.future.2019.05.041
34. Alsaedi, A., Moustafa, N., Tari, Z., Mahmood, A. N., and Anwar, A. (2020). "TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems." *IEEE Access, 8*, 165130–165150. https://doi.org/10.1109/ACCESS.2020.3022862
35. Ferrag, M. A., Friha, O., Hamouda, D., Maglaras, L., and Janicke, H. (2022). "Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning." *IEEE Access, 10*, 40281–40306. https://doi.org/10.1109/ACCESS.2022.3165809
36. Moustafa, N., and Slay, J. (2016). "The Evaluation of Network Anomaly Detection Systems: Statistical Analysis of the UNSW-NB15 Data Set and the Comparison with the KDD99 Data Set." *Information Security Journal: A Global Perspective, 25*(1–3), 18–31. https://doi.org/10.1080/19393555.2015.1125974
37. Ring, M., Wunderlich, S., Grüdl, D., Landes, D., and Hotho, A. (2019). "A Survey of Network-Based Intrusion Detection Data Sets." *Computers & Security, 86*, 147–167. https://doi.org/10.1016/j.cose.2019.06.005
38. Buczak, A. L., and Guven, E. (2016). "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection." *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502
39. Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G., and Vázquez, E. (2009). "Anomaly-Based Network Intrusion Detection: Techniques, Systems and Challenges." *Computers & Security, 28*(1–2), 18–28. https://doi.org/10.1016/j.cose.2008.08.003
40. Wang, B., Zheng, Y., Lou, W., and Hou, Y. T. (2015). "DDoS Attack Protection in the Era of Cloud Computing and Software-Defined Networking." *Computer Networks, 81*, 308–319. https://doi.org/10.1016/j.comnet.2015.02.026
41. Kreutz, D., Ramos, F. M. V., Veríssimo, P. E., Rothenberg, C. E., Azodolmolky, S., and Uhlig, S. (2015). "Software-Defined Networking: A Comprehensive Survey." *Proceedings of the IEEE, 103*(1), 14–76. https://doi.org/10.1109/JPROC.2014.2371999