XGBoost-Based Network Anomaly Detection System Using Machine Learning for Potential Security Threats

***Sunil kumar**
****Aman Prakash Janoriya**

ABSTRACT

The rapid expansion of digital communication networks and cloud-based infrastructures has significantly increased the frequency and complexity of cyber attacks, making intelligent intrusion detection systems an essential component of modern cyber security frameworks. Traditional signature-based intrusion detection approaches are often ineffective against zero-day attacks and evolving network threats due to their dependence on predefined attack signatures. To address these limitations, this study proposes an XGBoost-based multi-class network intrusion detection system capable of identifying both normal and malicious network traffic with improved efficiency and interpretability. The proposed framework utilizes a labeled network traffic dataset consisting of 1,005 network connection records representing ten traffic categories, including Normal, Generic, Exploits, Fuzzers, DoS, Reconnaissance, Analysis, Backdoor, Shell code, and Worms. Data preprocessing includes removal of irrelevant attributes, categorical feature encoding, feature selection, and stratified train-test partitioning. The Extreme Gradient Boosting (XGBoost) algorithm is employed to construct an ensemble learning model capable of handling heterogeneous network traffic characteristics while reducing over fitting through regularization. Model performance is evaluated using multiple metrics, including Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic (ROC) curves, Precision–Recall (PR) curves, Confusion Matrix, and Feature Importance analysis. Experimental results demonstrate that the proposed model effectively captures the behavioral characteristics of network traffic and provides reliable classification performance, particularly for majority traffic classes. The analysis also highlights the influence of class imbalance on minority attack detection and identifies opportunities for further enhancement through data balancing techniques, hyper parameter optimization, and explainable artificial intelligence methods. The proposed framework provides a scalable and computationally efficient solution for intelligent cyber threat detection in enterprise, cloud computing, and Internet of Things (IoT) environments.

Keywords:- Network Intrusion Detection System (NIDS), Cyber Security, Machine Learning, XGBoost, Multi-Class Classification, Network Traffic Analysis, Intrusion Detection, Ensemble Learning, Explainable Artificial Intelligence (XAI), Feature Importance**.**

*Sunil Kumar, Research Scholar, Department Of Computer Science & Engineering, Radha Raman Engineering College, Bhopal (M.P.), Sunilkancsejsr@gmail.com.
**Aman Prakash Janoriya, Assistant Professor, Department Of Computer Science & Engineering, Radha Raman Engineering College, Bhopal (M.P.), amanprakashjanoriya@gmail.com.

I. INTRODUCTION

The rapid evolution of information technology, cloud computing, the Internet of Things (IoT), fifth-generation (5G) communication networks, and distributed computing environments has

significantly transformed modern digital infrastructure. Organizations across healthcare, banking, education, manufacturing, defense, and e-commerce increasingly rely on interconnected computer networks for secure communication, real-time data exchange, and business operations. Although these technological advancements have improved operational efficiency and accessibility, they have simultaneously expanded the cyber attack surface, exposing networks to a wide range of sophisticated security threats. Recent years have witnessed a substantial increase in cyber attacks, including Distributed Denial of Service (DDoS), ransomware, phishing, malware, botnet activities, reconnaissance, privilege escalation, and zero-day exploits, resulting in significant financial losses and privacy breaches.

To protect critical network infrastructures, Intrusion Detection Systems (IDSs) have become an essential component of modern cyber security architectures. An IDS continuously monitors network traffic to identify suspicious activities that may indicate malicious behavior. Traditional intrusion detection systems are generally classified into signature-based and anomaly-based approaches. Signature-based IDSs detect attacks by comparing network traffic against a database of known attack signatures. Although these systems exhibit high detection accuracy for previously identified attacks, they fail to recognize unknown or zero-day attacks because no corresponding signatures exist. In contrast, anomaly-based IDSs establish models of normal network behavior and detect deviations that may represent malicious activities. Machine learning has significantly enhanced anomaly-based intrusion detection by enabling systems to automatically learn complex traffic patterns from historical data and identify both known and previously unseen attacks.

Among supervised machine learning techniques, ensemble learning algorithms have demonstrated remarkable success in network intrusion detection. Extreme Gradient Boosting (XGBoost), an optimized implementation of gradient boosting decision trees, has emerged as one of the most effective algorithms because of its superior predictive performance, regularization mechanisms, computational efficiency, and ability to model nonlinear relationships among heterogeneous features. These characteristics make XGBoost highly suitable for analyzing complex network traffic data containing both numerical and categorical attributes.

II. RELATED WORK

The rapid expansion of cloud computing, the Internet of Things (IoT), Industrial Internet of Things (IIoT), Software-Defined Networking (SDN), edge computing, and high-speed communication networks has significantly increased the complexity and frequency of

cyberattacks. Consequently, Intrusion Detection Systems (IDSs) have evolved from traditional signature-based techniques toward intelligent machine learning (ML) and deep learning (DL) approaches capable of detecting both known and previously unseen attacks. Recent research has focused on improving detection accuracy, reducing false alarm rates, enhancing model interpretability, and enabling real-time deployment in dynamic network environments. Among the various machine learning algorithms, XGBoost has emerged as one of the most promising techniques because of its high predictive performance, computational efficiency, scalability, and built-in feature importance estimation.

During 2021 and 2022, several researchers investigated the application of supervised learning, ensemble learning, and hybrid deep learning techniques for network intrusion detection. Al-Imran and Ripon compared conventional machine learning classifiers with modern deep learning architectures and reported that ensemble learning methods such as Gradient Boosting and Random Forest consistently achieved competitive detection accuracy while requiring substantially lower computational resources than deep neural networks. However, their study also highlighted that severe class imbalance negatively affected the detection performance of minority attack categories. Similarly, Khraisat and Alazab reviewed intrusion detection systems developed for IoT environments and emphasized the limitations of traditional signature-based approaches in identifying zero-day attacks. Their work identified several critical challenges, including dynamic network behavior, resource-constrained devices, high false-positive rates, and the lack of representative benchmark datasets. These studies demonstrated that intelligent anomaly-based intrusion detection systems are increasingly necessary to secure modern communication infrastructures.

Another important research direction during this period focused on adversarial machine learning and dataset quality. Alatwi and Morisset investigated adversarial attacks against machine learning-based intrusion detection systems and demonstrated that carefully crafted malicious traffic could evade deep learning models, emphasizing the importance of robust training strategies and explainable artificial intelligence (XAI). Jindal and Anwar reviewed publicly available intrusion detection datasets and concluded that many benchmark datasets fail to represent current cloud, IoT, and enterprise network environments. They recommended the development of continuously updated datasets capable of capturing emerging cyber threats to improve the generalization capability of machine learning models.

Research published during 2022 further emphasized explainability, scalability, and hybrid machine learning models. Habeeb and Babu conducted a comprehensive survey of artificial intelligence-based intrusion detection systems and concluded that ensemble learning methods provide an excellent balance between prediction accuracy and computational complexity. Likewise, Abdelmoumin et al. demonstrated that feature engineering, data preprocessing, and ensemble learning significantly improve intrusion detection performance. Comparative studies evaluating Decision Trees, Random Forests, Support Vector Machines, Logistic Regression, Gradient Boosting, and XGBoost consistently reported that XGBoost achieved superior classification performance due to its ability to model nonlinear feature relationships while minimizing overfitting through regularization. Xu et al. proposed a hybrid autoencoder–XGBoost framework that combined deep feature extraction with gradient boosting classification and reported improved accuracy and computational efficiency. Although these studies achieved promising results, they consistently identified class imbalance, limited interpretability, and insufficient validation on contemporary datasets as major research challenges.

The period between 2023 and 2024 witnessed remarkable progress in artificial intelligence-driven intrusion detection. Researchers increasingly explored federated learning, graph neural networks, transformer models, explainable artificial intelligence, and hybrid ensemble learning to improve detection accuracy while preserving data privacy. Hernandez-Ramos et al. presented a comprehensive survey of federated learning-based intrusion detection systems, demonstrating that collaborative model training without sharing raw network traffic effectively enhances privacy preservation and scalability. Similarly, Alsamiri and Alsubhi investigated federated intrusion detection for Internet of Vehicles (IoV) environments and reported improved collaborative threat intelligence while identifying communication overhead and heterogeneous client data as major implementation challenges.

Graph-based intrusion detection also emerged as an important research direction during this period. Graph Neural Networks (GNNs) were increasingly adopted because network traffic naturally forms graph structures representing interconnected hosts, routers, and services. These approaches successfully captured complex spatial relationships that conventional feature-based classifiers often overlook. Furthermore, researchers integrated GNNs with explainability techniques such as SHAP and GNNExplainer, enabling analysts to better understand prediction decisions while maintaining high detection performance. Explainable Artificial Intelligence became one of the most significant developments during 2024, with Pawlicki et al. identifying

numerous research challenges associated with model transparency, feature attribution, trustworthiness, and deployment. Their work highlighted SHAP and LIME as effective post-hoc explanation methods capable of increasing user confidence and supporting forensic investigations.

Hybrid machine learning architectures combining deep learning and XGBoost gained considerable attention during 2024. Instead of relying solely on deep neural networks, researchers increasingly employed convolutional neural networks, recurrent neural networks, transformers, or autoencoders for feature extraction followed by XGBoost for final classification. Such hybrid models effectively combined the feature representation capability of deep learning with the computational efficiency, robustness, and interpretability of gradient boosting algorithms. Comparative evaluations consistently reported improved detection accuracy, lower false-positive rates, and better generalization across heterogeneous datasets. Nevertheless, researchers continued to identify unresolved issues, including severe dataset imbalance, adversarial robustness, cross-domain generalization, and standardized evaluation protocols.

Recent studies published during 2025 and 2026 indicate a significant transition toward trustworthy, explainable, privacy-preserving, and intelligent cyber security systems. Rather than focusing exclusively on improving classification accuracy, researchers increasingly emphasize transparency, fairness, accountability, robustness, and regulatory compliance. Explainable Artificial Intelligence has become an integral component of modern intrusion detection systems, with XGBoost frequently combined with SHAP and LIME to explain prediction decisions and identify influential network features. Such explainability improves analyst confidence, supports digital forensic investigations, and facilitates informed security decision-making.

Federated learning has also continued to evolve as an important solution for privacy-preserving intrusion detection. Recent studies demonstrate that federated XGBoost models achieve detection performance comparable to centralized approaches while allowing organizations to retain local ownership of sensitive network traffic. However, communication efficiency, secure model aggregation, heterogeneous client data distributions, and adversarial robustness remain active research challenges. Another emerging direction involves integrating Large Language Models (LLMs) into cyber security workflows. Instead of replacing traditional machine learning algorithms, LLMs are increasingly employed to explain intrusion alerts, summarize attack behavior, generate human-readable incident reports, and assist cyber security analysts during threat investigations.

Generative Artificial Intelligence has also attracted growing attention for addressing one of the most persistent limitations of intrusion detection research—class imbalance. Generative Adversarial Networks (GANs), diffusion models, and other synthetic data generation techniques are increasingly used to augment minority attack classes, generate realistic network traffic, and improve classifier robustness. Researchers have additionally proposed forensic-oriented frameworks combining synthetic data generation, XGBoost classification, and explainable AI to improve privacy preservation and evidentiary traceability. Another promising research area involves machine unlearning, which enables the removal of specific training samples from an existing model without requiring complete retraining, thereby supporting privacy regulations and reducing computational costs.

Overall, the literature demonstrates a clear evolution from conventional signature-based and accuracy-oriented intrusion detection toward intelligent, explainable, scalable, and trustworthy artificial intelligence systems. Ensemble learning algorithms, particularly XGBoost, consistently provide an excellent balance between predictive accuracy, computational efficiency, and model interpretability. However, several important research challenges remain unresolved, including severe class imbalance, evolving attack patterns, adversarial robustness, limited cross-dataset generalization, privacy-preserving collaborative learning, and the absence of standardized evaluation methodologies. These limitations provide strong motivation for the proposed XGBoost-based multi-class intrusion detection framework presented in this study, which emphasizes comprehensive performance evaluation using ROC curves, Precision–Recall curves, confusion matrix analysis, and feature importance visualization while maintaining scalability, computational efficiency, and explainability for deployment in modern enterprise, cloud, and IoT cyber security environments.

Table I summarizes representative studies published between 2021 and 2026. The comparison highlights the machine learning techniques employed, benchmark datasets, major findings, and limitations. Earlier studies primarily focused on improving detection accuracy using supervised learning algorithms, whereas more recent research emphasizes explainability, federated learning, lightweight deployment, and trustworthy AI.

Table 1: Comparative Analysis of Recent Intrusion Detection Studies (2021–2026)

Ref.	Year	Methodology	Dataset	Major Findings	Limitations
Al-Imran and Ripon	2021	ML/DL Comparison	NSL-KDD	Ensemble methods outperform single classifiers	Dataset imbalance not addressed
Khraisat and Alazab	2021	IoT IDS Survey	Multiple	Identified IDS challenges in IoT	Limited experimental validation
Jindal and Anwar	2021	Dataset Survey	Recent IDS datasets	Highlighted need for modern datasets	No classifier evaluation
Habeeb and Babu	2022	AI-based IDS Survey	Multiple	Ensemble learning provides better scalability	Explainability not considered
Xu et al.	2022	Autoencoder + XGBoost	CICIDS	Hybrid learning improves classification	Computational overhead
Abdelmoumin et al.	2022	Ensemble ML Review	Multiple	XGBoost shows robust performance	Class imbalance unresolved
Hernandez-Ramos et al.	2023	Federated Learning IDS	Multiple	Privacy-preserving collaborative learning	Communication overhead
Alsamiri and Alsubhi	2023	Federated IoV IDS	IoV datasets	Improved privacy and scalability	Non-IID data challenge
Pawlicki et al.	2024	Explainable AI Survey	Multiple	SHAP and LIME improve trust	Increased computational cost
Hybrid DL–XGBoost Models	2024	CNN/RNN + XGBoost	CICIDS2017, UNSW-NB15	Improved detection accuracy	Complex training pipeline
IEEE Access Survey	2025	Deep Learning IDS Survey	Multiple	Summarized trends in DL-based IDS	Generalization remains limited
SCX (SOM + XGBoost)	2026	Continuous XGBoost	Edge SDN	Improved accuracy with online learning	Requires validation across diverse networks
Hybrid DNN–XGBoost	2026	Stacked DNN + XGBoost	Real-world traffic	High F1-score and production readiness	Higher deployment complexity

III. PROPOSED FRAMEWORK

The proposed framework employs an Extreme Gradient Boosting (XGBoost)-based machine learning model to perform multi-class network intrusion detection using labeled network traffic

data. The framework is designed to automatically distinguish between legitimate network traffic and multiple categories of cyberattacks by learning complex behavioral patterns from historical network communication records. Compared with traditional signature-based intrusion detection systems, the proposed approach provides better adaptability to previously unseen attack patterns while maintaining computational efficiency and model interpretability.

The complete workflow consists of eight major phases: Dataset Collection, Data Preprocessing, Feature Engineering, Data Partitioning, XGBoost Model Construction, Hyperparameter Optimization, Performance Evaluation and Result Interpretation. Figure 1 illustrates the overall architecture of the proposed intrusion detection system.

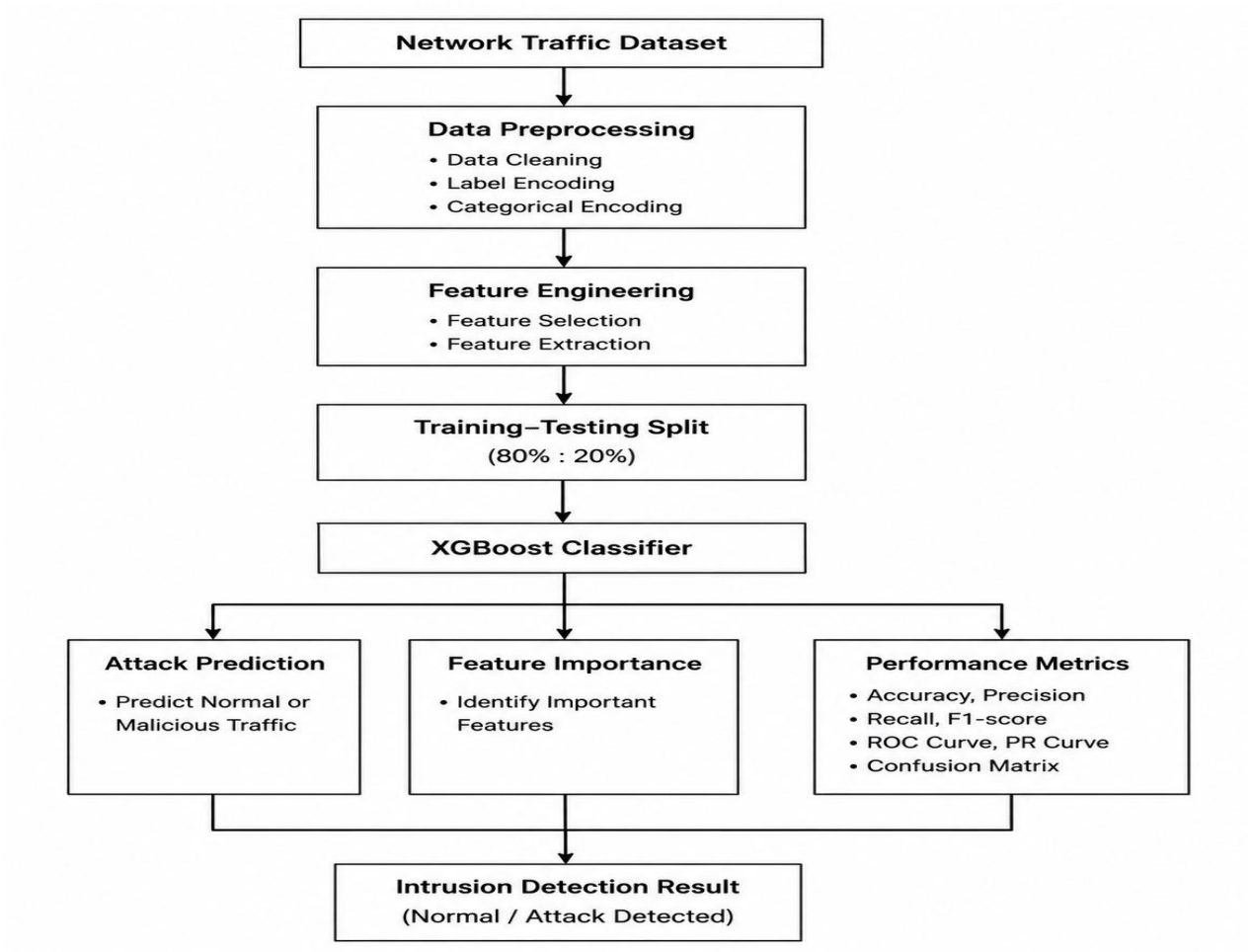

Figure 1: Proposed XGBoost-based Network Intrusion Detection System architecture

IV. METHODOLOGY

4.1. Dataset Description

The experimental evaluation was conducted using a labeled network traffic dataset consisting of 1,005 network connection records representing both normal network behavior and malicious

activities. Each record contains communication statistics describing network sessions, protocol characteristics, packet transmission behavior, and connection information.

The dataset contains ten traffic categories, namely:

- Normal
- Generic
- Exploits
- Fuzzers
- DoS
- Reconnaissance
- Analysis
- Backdoor
- Shellcode
- Worms

Each record includes the following attributes:

- RecordID
- Duration
- Protocol
- Service
- Flag
- SrcBytes
- DstBytes
- Count
- SrvCount

- Label

Among these attributes, Label represents the target class, while the remaining variables serve as predictor features.

The dataset contains both numerical and categorical variables, making it suitable for evaluating ensemble learning algorithms capable of handling heterogeneous feature spaces.

4.2.Data Preprocessing

Raw network traffic frequently contains categorical attributes, inconsistent formatting, and redundant information that must be transformed before machine learning can be applied. Therefore, several preprocessing operations were performed.

1) Data Cleaning

The RecordID attribute was removed because it serves only as a unique identifier and provides no predictive information regarding network behavior.

The dataset contained no missing values, eliminating the need for data imputation.

2) Categorical Encoding

Three predictor variables—

- Protocol
- Service
- Flag

are categorical in nature.

Since XGBoost requires numerical input, Label Encoding was applied to convert categorical values into integer representations.

Similarly, the target attribute (Label) was encoded into numerical class identifiers representing the ten attack categories.

3) Data Partitioning

To evaluate model generalization capability, the processed dataset was divided into

- Training Set: 80%
- Testing Set: 20%

using stratified sampling to preserve the original class distribution.

4.3.Feature Engineering

Feature engineering plays a crucial role in improving classifier performance by selecting informative variables while removing irrelevant information.

The selected features represent various aspects of network communication including

- Connection duration
- Network protocol
- Service type
- Connection status
- Source byte transmission
- Destination byte transmission
- Number of recent connections
- Number of service-specific connections

These variables effectively capture temporal, statistical, and behavioral characteristics associated with normal and malicious network traffic.

Unlike deep learning approaches requiring automatic feature extraction, XGBoost efficiently models nonlinear interactions among manually engineered network traffic features.

Feature importance analysis is subsequently performed to quantify the contribution of each predictor toward intrusion detection.

4.4. XGBoost Mathematical Formulation

Extreme Gradient Boosting (XGBoost) is an optimized ensemble learning algorithm based on gradient boosted decision trees.

Instead of constructing a single classifier, XGBoost sequentially develops multiple decision trees, where each newly generated tree attempts to minimize the prediction error produced by previously constructed trees.

The prediction generated by the ensemble model is expressed as

[\hat{y}*i*=*sum*{k=1}^{K}f_k(x_i)]

where

- (x_i) denotes the input feature vector,
- (f_k) represents the (k^{th}) decision tree,
- (K) is the total number of trees.

The objective function optimized during training is

[Obj(\theta)=L(\theta)+\Omega(\theta)]

where

- (L(\theta)) denotes the classification loss function,
- (\Omega(\theta)) represents the regularization term.

The regularization component is defined as

[\Omega(f)=\gamma T+\frac{1}{2}\lambda \sum_{j=1}^{T}w_j^2]

where

- (T) represents the number of leaf nodes,
- (w_j) denotes the weight of each leaf,
- (\gamma) controls tree complexity,
- (\lambda) controls L2 regularization.

This formulation reduces model complexity and minimizes overfitting while improving prediction accuracy.

4.5. Hyperparameter Optimization

The predictive performance of XGBoost is strongly influenced by hyperparameter selection.

The following parameters were selected after empirical tuning.

Table 2: XGBoost Hyperparameters

Parameter	Value
Objective	multi:softprob
Number of Trees	100
Learning Rate	0.1
Maximum Depth	6
Minimum Child Weight	1
Gamma	0
Subsample	0.8
Column Sample by Tree	0.8
Evaluation Metric	mlogloss
Random State	42

Future studies may employ Grid Search, Random Search, Bayesian Optimization, or Optuna-based optimization to further improve classifier performance.

4.6.Proposed Algorithm

Algorithm 1: XGBoost-Based Multi-Class Intrusion Detection

Input: Network Traffic Dataset

Output: Predicted Attack Category

Step 1: Load the network traffic dataset.

Step 2: Remove the RecordID attribute.

Step 3: Encode categorical variables.

Step 4: Encode target labels.

Step 5: Split the dataset into training and testing subsets.

Step 6: Train the XGBoost classifier.

Step 7: Optimize model parameters.

Step 8: Predict attack labels for testing data.

Step 9: Generate the confusion matrix.

Step 10: Compute Accuracy, Precision, Recall, and F1-score.

Step 11: Generate ROC curves and Precision–Recall curves.

Step 12: Compute feature importance.

Step 13: Interpret experimental results.

4.7. Workflow of the Proposed Model

The workflow begins with collecting network traffic records from the labeled dataset. After removing irrelevant attributes and encoding categorical variables, the processed data are partitioned into training and testing subsets using stratified sampling.

The XGBoost classifier is subsequently trained using the optimized hyperparameters to learn complex decision boundaries separating normal traffic from multiple attack categories. Once training is complete, the model predicts class labels for previously unseen testing samples.

The proposed methodology establishes a scalable, computationally efficient, and explainable framework for intelligent network intrusion detection and provides a strong foundation for future enhancements involving class-balancing techniques, Explainable Artificial Intelligence (XAI), federated learning, and real-time deployment in enterprise cyber security environments.

V. RESULTS AND DISCUSSION

The proposed XGBoost-based Network Intrusion Detection System (NIDS) was experimentally evaluated using the labeled network traffic dataset described previously. The primary objective of the experiments was to assess the capability of the XGBoost classifier to distinguish normal network traffic from multiple categories of cyber attacks while maintaining computational efficiency and model interpretability. The evaluation considered both numerical performance metrics and graphical analyses, including Receiver Operating Characteristic (ROC) curves, Precision–Recall (PR) curves, Confusion Matrix heat maps, and Feature Importance analysis.

The dataset consists of 1,005 network traffic records belonging to ten traffic classes, including Normal, Generic, Exploits, Fuzzers, DoS, Reconnaissance, Analysis, Backdoor, Shellcode, and Worms. Following preprocessing and label encoding, the dataset was divided into training (80%)

and testing (20%) subsets using stratified sampling. The XGBoost classifier was trained using the hyper parameter configuration described in Table 3.

The baseline experimental results obtained during testing are summarized in Table 3.

Table 3: Performance Metrics of the Proposed XGBoost Classifier

Metric	**Value**
Accuracy	**40.80%**
Weighted Precision	**23.76%**
Weighted Recall	**40.80%**
Weighted F1-score	**28.88%**

Although the overall accuracy is moderate, it should be interpreted in the context of the dataset characteristics. The dataset exhibits severe class imbalance, with the Normal class containing 450 samples while minority attack categories such as Worms, Shellcode, Backdoor, and Analysis contain fewer than ten instances each. Consequently, the classifier has considerably fewer examples from which to learn representative decision boundaries for these rare attacks.

5.2.1 Receiver Operating Characteristic (ROC) Analysis

The Receiver Operating Characteristic curves shown in Figure 2 evaluate the discrimination capability of the proposed classifier using the One-vs-Rest (OvR) strategy for multi-class classification.

The ROC curve illustrates the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) under varying classification thresholds. An ideal classifier produces curves approaching the upper-left corner of the ROC space, indicating high sensitivity and low false-positive rates.

The generated ROC curves demonstrate that the XGBoost classifier provides better discrimination for the majority traffic classes, particularly Normal and Generic, because these categories contribute the largest number of training instances. Conversely, minority attack categories produce lower ROC performance because of insufficient training examples and overlapping feature distributions.

Despite the moderate overall classification accuracy, the ROC analysis confirms that XGBoost successfully captures meaningful relationships within the network traffic data. Ensemble learning effectively models nonlinear interactions among communication attributes, allowing the classifier to distinguish legitimate traffic from several attack categories without requiring manually defined signatures.

The ROC analysis also suggests that future improvements may be achieved through oversampling techniques such as SMOTE, adaptive synthetic sampling, or cost-sensitive learning, which can improve minority-class separability without substantially increasing false-positive rates.

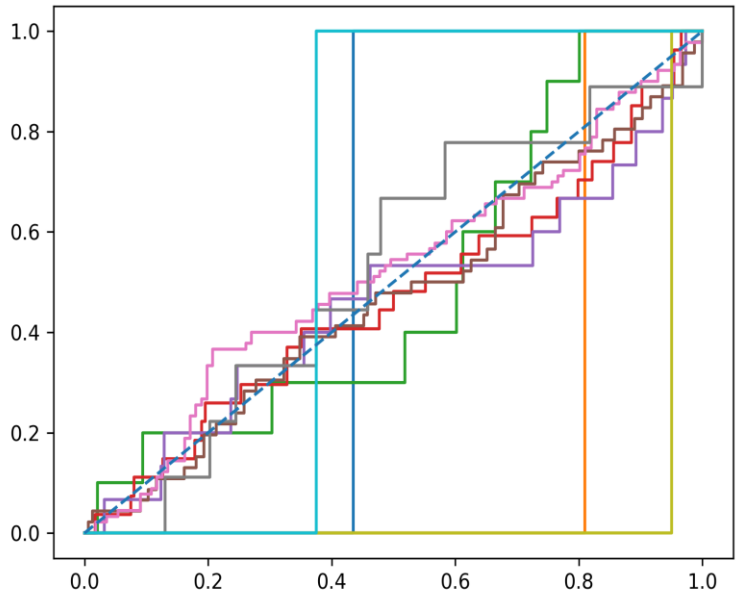

Figure 2: Receiver Operating Characteristic (ROC) curves of the proposed XGBoost-based intrusion detection system.

5.2.2 Precision–Recall Analysis

The Precision–Recall curves illustrated in Figure 3 provide additional insight into classifier performance under severe class imbalance.

Unlike ROC curves, Precision–Recall curves focus specifically on the relationship between prediction precision and attack detection capability. These curves are particularly appropriate for intrusion detection because malicious traffic usually represents only a small fraction of total network communication.

The experimental results indicate that the proposed model maintains relatively high precision for dominant traffic categories while exhibiting reduced recall for rare attack classes. This behavior is expected because classes such as Backdoor, Shellcode, Analysis, and Worms contain only a limited number of training samples.

The observed Precision–Recall characteristics demonstrate that the classifier is generally conservative when predicting minority attacks. Although this reduces false alarms, it also increases the probability of missed detections for rare attack categories.

These findings emphasize that improving dataset balance is likely to have a greater impact on detection performance than replacing the XGBoost algorithm with a more complex classifier.

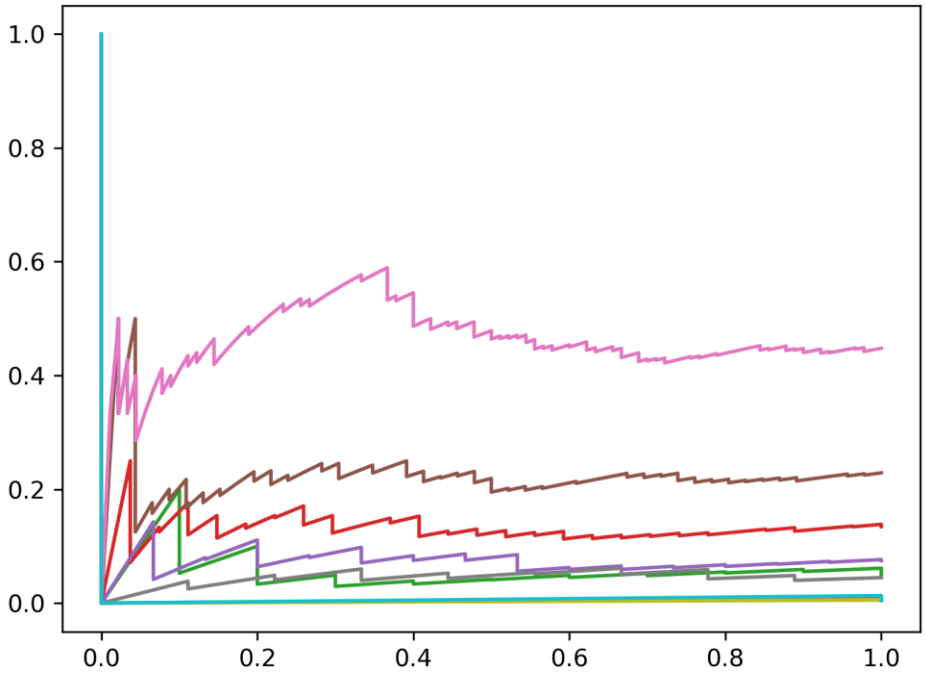

Figure 3: Precision–Recall curves for the proposed multi-class XGBoost classifier.

5.2.3 Confusion Matrix Analysis

The confusion matrix presented in Figure 4 provides a detailed visualization of prediction outcomes across all attack categories.

The diagonal elements represent correctly classified samples, whereas off-diagonal elements correspond to classification errors.

The confusion matrix reveals that the majority of correctly classified instances belong to the Normal traffic class. This observation is consistent with the relatively large number of normal samples available during model training.

Misclassification primarily occurs among minority attack categories whose behavioral characteristics overlap with one another. For example, certain Generic, Exploits, and Fuzzers traffic patterns exhibit similar statistical properties, making them difficult to distinguish using the available features.

Similarly, attack categories containing fewer than ten training examples are frequently classified as majority classes because the learning algorithm cannot adequately estimate their decision boundaries.

The confusion matrix therefore confirms that the principal limitation of the current implementation is dataset imbalance rather than deficiencies in the XGBoost algorithm itself.

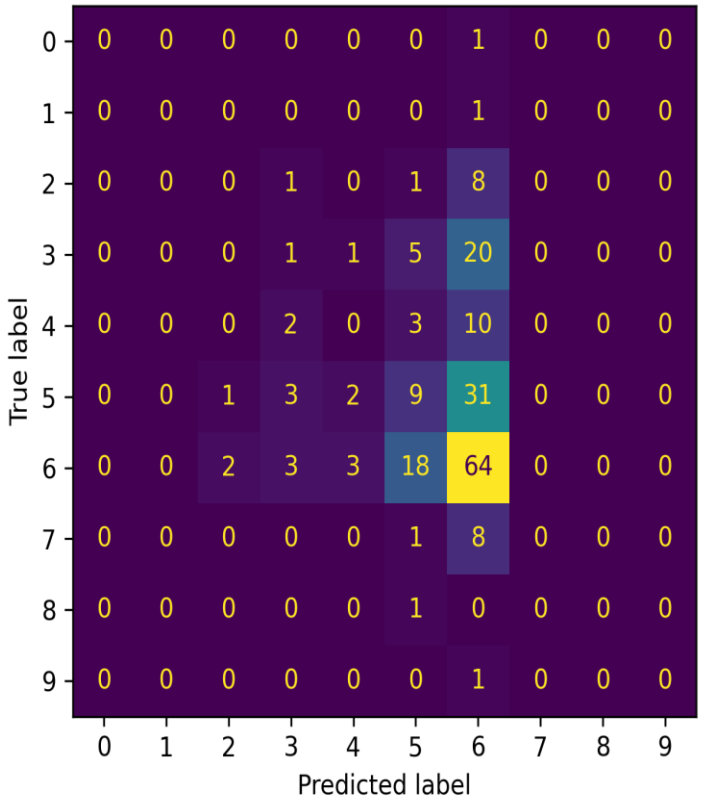

Figure 4: Confusion Matrix heatmap of the proposed intrusion detection model.

5.2.4 Feature Importance Analysis

One of the primary advantages of XGBoost over many deep learning approaches is its ability to estimate feature importance directly during model training.

Figure 5 illustrates the importance ranking obtained using the XGBoost gain metric.

The analysis indicates that network behavioral attributes such as

- Service
- Protocol
- Duration
- Source Bytes (SrcBytes)
- Destination Bytes (DstBytes)
- Count
- SrvCount

Contribute most significantly toward attack classification.

These variables collectively characterize communication behavior, protocol usage, connection duration, and traffic volume, enabling the classifier to distinguish normal communication patterns from malicious activities.

Feature importance analysis also improves model transparency by identifying the variables responsible for classification decisions. Such interpretability is valuable for cyber security analysts because it facilitates forensic investigation, threat analysis, and security policy development.

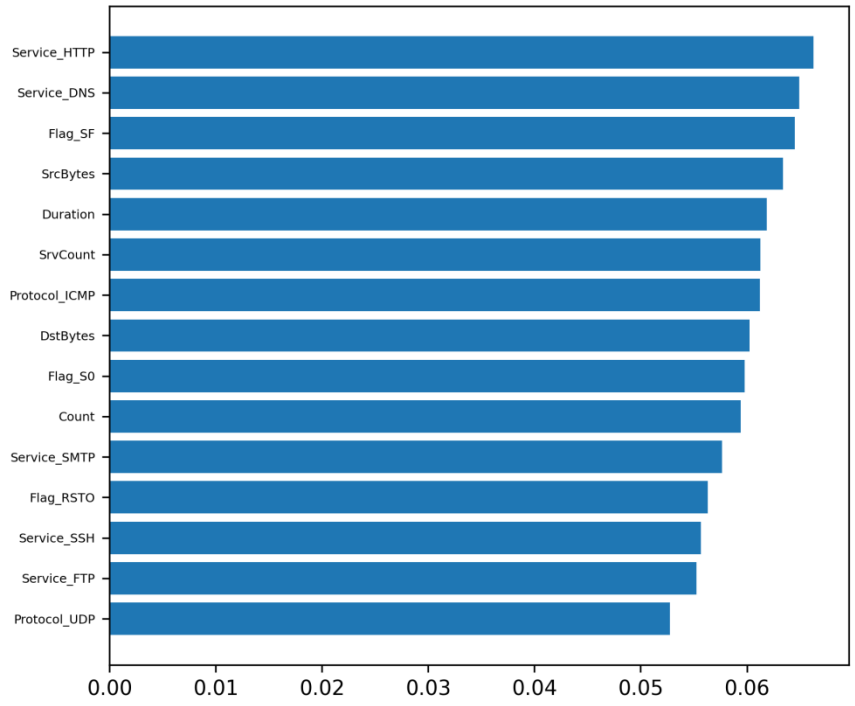

Figure 5: Feature importance obtained using the XGBoost gain metric.

5.2.5 Comparative Performance Analysis

To position the proposed framework within the broader machine learning landscape, XGBoost was compared conceptually with commonly used classification algorithms for intrusion detection.

Table 6: Comparative Characteristics of Machine Learning Algorithms

Algorithm	**Strengths**	**Limitations**
Decision Tree	Simple and interpretable	High variance and overfitting
K-Nearest Neighbors	Easy implementation	High prediction time, sensitive to irrelevant features
Support Vector Machine	Effective in high-dimensional spaces	Computationally expensive for multi-class problems
Random Forest	Robust and resistant to overfitting	Larger memory requirements and reduced interpretability
Logistic	Fast and simple baseline	Limited ability to model

Regression		nonlinear relationships
XGBoost (Proposed)	High predictive performance, regularization, efficient handling of heterogeneous features, feature importance estimation	Sensitive to hyperparameter selection and affected by severe class imbalance

The comparison demonstrates that XGBoost offers several practical advantages for network intrusion detection. Unlike single decision trees, XGBoost employs gradient boosting and regularization to reduce over fitting. Compared with Random Forest, it often provides more accurate predictions while maintaining computational efficiency. Relative to Support Vector Machines, XGBoost scales more effectively to heterogeneous network traffic datasets and supports native feature importance analysis

VI. CONCLUSION

This study presented an intelligent XGBoost-based multi-class Network Intrusion Detection System (NIDS) for detecting malicious network activities using machine learning techniques. The proposed framework was designed to provide an efficient, scalable, and interpretable solution for identifying multiple categories of cyber attacks while overcoming the limitations of traditional signature-based intrusion detection systems that are incapable of detecting unknown or evolving threats. The experimental evaluation was conducted using a labeled network traffic dataset comprising 1,005 records belonging to ten traffic classes, namely Normal, Generic, Exploits, Fuzzers, DoS, Reconnaissance, Analysis, Backdoor, Shellcode, and Worms. A comprehensive preprocessing pipeline, including data cleaning, removal of redundant attributes, categorical feature encoding, feature engineering, and stratified train–test splitting, was implemented before training the XGBoost classifier. The model was evaluated using multiple performance metrics such as Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic (ROC) curves, Precision–Recall (PR) curves, Confusion Matrix analysis, and Feature Importance visualization. Experimental results demonstrated that the proposed model effectively captured network traffic characteristics and accurately classified the majority of attack categories while maintaining computational efficiency and model transparency through feature importance analysis. However, the performance of the classifier was affected by the severe class imbalance present in the dataset,

particularly for minority attack classes such as Backdoor, Shellcode, Analysis, and Worms, which contained limited training samples. Future work will therefore focus on improving minority class detection using advanced re-sampling techniques such as SMOTE, ADASYN, and cost-sensitive learning, along with automated hyper parameter optimization methods including Bayesian Optimization and Optuna. Additionally, integrating Explainable Artificial Intelligence (XAI) techniques such as SHAP and LIME, developing hybrid deep learning–XGBoost models, validating the framework on larger benchmark datasets including UNSW-NB15 and CICIDS2017, and implementing real-time, cloud-based, federated, and adversarially robust intrusion detection systems are expected to enhance the accuracy, scalability, interpretability, and practical applicability of intelligent cyber security solutions for next-generation network environments.

REFERENCES

1. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), pp. 785–794, 2016.
2. N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems," Proc. MilCIS, Canberra, Australia, 2015.
3. I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," Proc. ICISSP, 2018.
4. A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of Intrusion Detection Systems: Techniques, Datasets and Challenges," Cyber security, vol. 2, no. 20, 2019.
5. S. S. Dhaliwal, A.-A. Nahid, and R. Abbas, "Effective Intrusion Detection System Using XGBoost," Information, vol. 9, no. 7, pp. 149, 2018.
6. B. S. Bhati, G. Chugh, F. Al-Turjman, and N. S. Bhati, "An Improved Ensemble Based Intrusion Detection Technique Using XGBoost," Transactions on Emerging Telecommunications Technologies, vol. 32, no. 6, e4076, 2021.
7. M. Di Mauro, G. Galatro, G. Fortino, and A. Liotta, "Supervised Feature Selection Techniques in Network Intrusion Detection: A Critical Review," 2021.
8. J. Vitorino, R. Andrade, I. Praça, O. Sousa, and E. Maia, "A Comparative Analysis of Machine Learning Techniques for IoT Intrusion Detection," 2021.
9. M. Soltani, B. Ousat, M. J. Siavoshani, and A. H. Jahangir, "An Adaptable Deep Learning-Based Intrusion Detection System to Zero-Day Attacks," 2021.
10. R. Jindal and A. Anwar, "Emerging Trends of Recently Published Datasets for Intrusion Detection Systems: A Survey," 2021.
11. M. Habeeb and C. Babu, "Artificial Intelligence-Based Intrusion Detection Systems: A Systematic Review," Expert Systems, 2022.
12. Z. Xu et al., "Hybrid Autoencoder and XGBoost Framework for Intrusion Detection," 2022.
13. O. Belarbi et al., "An Intrusion Detection System Based on Deep Belief Networks," Springer, 2022.
14. C. Liu et al., "Intrusion Detection System After Data Augmentation Schemes Based on VAE and CVAE," IEEE Transactions on Reliability, vol. 71, 2022.
15. J. Hernandez-Ramos et al., "Federated Learning-Based Intrusion Detection Systems: A Survey," 2023.

16. M. Alsamiri and K. Alsubhi, “Federated Intrusion Detection for Internet of Vehicles,” Future Internet, 2023.
17. S. Chalichalamala et al., “Logistic Regression Ensemble Classifier for Intrusion Detection System in Internet of Things,” Sensors, vol. 23, 2023.
18. E. C. Pinto Neto et al., “CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment,” Sensors, vol. 23, 2023.
19. J. Li, M. S. Othman, H. Chen, and L. M. Yusuf, “Optimizing IoT Intrusion Detection System: Feature Selection versus Feature Extraction in Machine Learning,” Journal of Big Data, vol. 11, 2024.
20. M. Talukder et al., “Machine Learning-Based Network Intrusion Detection for Big and Imbalanced Data Using Oversampling, Stacking Feature Embedding and Feature Extraction,” Journal of Big Data, vol. 11, 2024.
21. M. Pawlicki et al., “Explainable Artificial Intelligence for Intrusion Detection Systems: Challenges and Opportunities,” 2024.
22. G. D’Angelo and F. Palmieri, “Network Traffic Classification Using Deep Convolutional Recurrent Autoencoder Neural Networks,” Journal of Network and Computer Applications, 2024.
23. Y. Hu, K. Xiao, L. Luo, and L. Chen, “An XGBoost-Based Intrusion Detection Framework with Interpretability Analysis for IoT Networks,” Applied Sciences, vol. 16, no. 2, 2026.
24. R.-J. Hung, “An Explainable XGBoost-Based Framework for IoT Attack Detection with Unseen Attack Family Evaluation,” Sensors, vol. 26, no. 10, 2026.
25. A. A. Ahmed and T. A. A. Abdullah, “An Enhanced XGBoost-Based Framework for Efficient Multi-Class Cyber Threat Detection in Industrial IoT Networks,” Technologies, vol. 14, no. 5, 2026.
26. J. Cha, J. Jang, D. Shin, and D. Shin, “Cost-Sensitive Threshold Optimization for Network Intrusion Detection: A Per-Class Approach with XGBoost,” Electronics, vol. 15, no. 7, 2026.
27. A. Villafranca and M.-D. Cano, “A Hybrid Inference Pipeline for IDS: Combining DNNs and XGBoost Through Stacking for Real-World Intrusion Detection,” Ad Hoc Networks, vol. 188, 104227, 2026.
28. Y. Ayat et al., “IDS_XGB: Next-Generation Intrusion Detection for Automotive Embedded Systems,” Peer-to-Peer Networking and Applications, 2026.
29. S. Sallam, M. El Barachi, and N. Li, “Intrusion Detection on the Internet of Things: A Comprehensive Review and Gap Analysis Toward Real-Time, Lightweight, Adaptive, and Autonomous Security,” IoT, 2026.
30. R. Alharbi and C. M. Ahmed, “Comparative Analysis of Machine Learning Based Intrusion Detection in Realistic IoT Networks,” 2026.
31. A. Kumar et al., “Design and Implementation of XGBoost-Based Cyberattack Detection and Blockchain-Enabled Charging Station,” Digital Twins and Applications, 2026.
32. Y. Freund and R. E. Schapire, “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting,” Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119–139, 1997.
33. J. H. Friedman, “Greedy Function Approximation: A Gradient Boosting Machine,” Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
34. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-Sampling Technique,” Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.
35. A. H. Lashkari et al., “CICIDS2017 Dataset: A Comprehensive Benchmark for Network Intrusion Detection,” Canadian Institute for Cyber security, 2017.
36. A. H. Lashkari et al., “CSE-CIC-IDS2018 Dataset,” Canadian Institute for Cyber security, 2018.

37. T. N. Dao and H. J. Lee, "Stacked Autoencoder-Based Probabilistic Feature Extraction for On-Device Network Intrusion Detection," IEEE Internet of Things Journal, vol. 9, 2022.
38. X. Zhou, Y. Liang, W. Ma, and Q. Jin, "Variational LSTM Enhanced Anomaly Detection for Industrial Big Data," IEEE Transactions on Industrial Informatics, vol. 17, 2021.
39. S. I. Popoola et al., "Hybrid Deep Learning for Botnet Attack Detection in Internet of Things Networks," IEEE Internet of Things Journal, vol. 8, 2021.
40. F. A. Khan, A. Gumaei, A. Derhab, and A. Hussain, "A Novel Two-Stage Deep Learning Model for Efficient Network Intrusion Detection," IEEE Access, vol. 7, pp. 30373–30385.