REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Received: 7 December 2023, Revised: 22 December 2023, Accepted: 05 January 2024

An Explainable Hybrid GA–GWO Optimized Ensemble Learning Framework for IoT Botnet Attack Detection and Mitigation

Nidhi B. Patel[1*], Dr. Swity Maniyar[2]

[1] PhD Scholar – Swaminarayan University, Kalol, Gujarat Emails: nidhipatel1462@gmail.com
[2]Associate Professor in GTU and PhD guide in Swaminarayan University, Kalol, Gujarat Emails: sweetymaniar@gmail.com

ABSTRACT
The rapid proliferation of Internet of Things (IoT) devices has significantly transformed modern digital ecosystems by enabling seamless communication among smart devices across healthcare, transportation, industrial automation, agriculture, and smart home environments. However, the increasing number of interconnected IoT devices has also expanded the attack surface for cyber threats, particularly botnet attacks, which exploit vulnerable devices to launch Distributed Denial of Service (DDoS), data theft, malware propagation, and command-and-control (C&C) attacks. Traditional intrusion detection systems often struggle to detect sophisticated botnet attacks because of high-dimensional network traffic, class imbalance, evolving attack patterns, and limited interpretability of machine learning models.
This research proposes an explainable hybrid ensemble learning framework for accurate IoT botnet detection and automated mitigation. The proposed framework employs hybrid feature engineering to extract statistical, behavioral, and temporal characteristics from IoT traffic. Principal Component Analysis (PCA) is applied to reduce feature dimensionality, while Synthetic Minority Oversampling Technique (SMOTE) addresses class imbalance. Three gradient boosting algorithms, namely XGBoost, LightGBM, and CatBoost, are integrated into an ensemble model to improve detection performance. Furthermore, a hybrid Genetic Algorithm–Grey Wolf Optimizer (GA–GWO) is utilized for hyperparameter optimization, enhancing convergence speed and reducing false positive predictions. Explainable Artificial Intelligence (XAI) using SHAP is incorporated to provide feature-level interpretation of detection decisions and improve transparency. Finally, automated mitigation strategies are triggered to isolate infected IoT devices and restrict malicious traffic.
Experimental evaluation on the benchmark N-BaIoT dataset demonstrates that the proposed hybrid framework achieves a classification accuracy of **98.95%**, outperforming individual XGBoost (94.88%), LightGBM (94.90%), and CatBoost (95.92%) classifiers. The proposed model also achieves an interpretability score of **98.2**, demonstrating its capability to deliver accurate, scalable, and explainable IoT botnet detection suitable for real-world deployment.

Keywords: Internet of Things, Botnet Detection, Ensemble Learning, XGBoost, LightGBM, CatBoost, Genetic Algorithm, Grey Wolf Optimizer, Explainable Artificial Intelligence, SHAP.

1. INTRODUCTION

1.1 Background

The Internet of Things (IoT) has emerged as one of the most transformative technologies of the digital era by connecting billions of smart devices through the Internet. IoT enables seamless communication among heterogeneous devices such as sensors, surveillance cameras, wearable devices, industrial controllers, smart appliances, healthcare monitoring systems, and autonomous vehicles. According to recent industrial reports, the number of connected IoT devices is expected to exceed **30 billion** worldwide within the next few years, generating enormous volumes of heterogeneous network traffic. This rapid expansion has significantly improved automation, operational efficiency, and real-time decision-making across various sectors including healthcare, manufacturing, agriculture, transportation, and smart cities.
Despite these benefits, the widespread deployment of IoT devices has introduced serious cybersecurity challenges. Most IoT devices possess limited computational resources, weak authentication mechanisms, outdated firmware, and inadequate security configurations. These limitations make them attractive targets for cybercriminals seeking to compromise vulnerable devices and recruit them into botnets. Once infected, compromised devices can be remotely controlled by attackers to perform malicious activities without the knowledge of legitimate users.
Botnet attacks represent one of the most destructive cybersecurity threats in IoT environments. A botnet consists of a network of compromised devices controlled through centralized or decentralized command-and-control (C&C) infrastructures. Attackers utilize these botnets to launch Distributed Denial of Service (DDoS) attacks, credential theft, malware distribution, spam campaigns, cryptojacking, ransomware propagation, and large-scale network disruptions. The Mirai botnet attack demonstrated how thousands of vulnerable IoT devices could be weaponized to generate massive DDoS attacks, causing widespread Internet outages and highlighting the urgent need for intelligent botnet detection mechanisms.
Traditional signature-based intrusion detection systems (IDS) are effective only for previously known attack signatures and struggle to detect zero-day attacks or evolving botnet behaviors. Machine learning approaches have therefore gained considerable attention because of their ability to automatically learn complex traffic patterns from network data. Algorithms such as Decision Trees, Random Forest, Support Vector Machines (SVM), XGBoost, LightGBM, CatBoost,

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Received: 7 December 2023, Revised: 22 December 2023, Accepted: 05 January 2024

and deep learning architectures including CNN, LSTM, and Autoencoders have demonstrated promising detection capabilities. However, existing studies still suffer from several limitations, including high computational complexity, poor generalization across heterogeneous IoT environments, class imbalance, lack of explainability, and dependence on single learning models.

To address these challenges, this study proposes an explainable hybrid ensemble learning framework that integrates multiple state-of-the-art machine learning techniques into a unified detection pipeline. Initially, advanced preprocessing and hybrid feature engineering techniques are employed to extract statistical, behavioral, and temporal characteristics from IoT traffic. Principal Component Analysis (PCA) is then applied to eliminate redundant features while preserving significant information. Since IoT datasets are generally imbalanced, the Synthetic Minority Over-sampling Technique (SMOTE) is adopted to balance the class distribution and improve classifier learning.

The proposed detection engine combines three powerful gradient boosting algorithms—XGBoost, LightGBM, and CatBoost—into an ensemble framework capable of capturing diverse decision boundaries. A hybrid optimization strategy combining the Genetic Algorithm (GA) and Grey Wolf Optimizer (GWO) is further incorporated to optimize model hyperparameters, improve convergence speed, reduce false positive rates, and enhance classification accuracy. Furthermore, Explainable Artificial Intelligence (XAI) based on SHAP is integrated to provide transparent explanations of prediction outcomes, enabling security analysts to understand feature contributions and improving trust in automated decision-making. Finally, the proposed framework supports automated mitigation actions by identifying compromised IoT devices and restricting malicious network traffic, thereby reducing attack propagation.

The effectiveness of the proposed framework is evaluated using the benchmark N-BaIoT dataset. Experimental results demonstrate that the hybrid ensemble model achieves an overall classification accuracy of **98.95%**, significantly outperforming individual boosting classifiers while providing high interpretability through SHAP-based explanations. These results indicate that the proposed framework offers an accurate, scalable, lightweight, and explainable solution for securing modern IoT networks against evolving botnet attacks.

1.2 Research Motivation

The rapid growth of Internet of Things (IoT) technology has revolutionized communication among interconnected devices, enabling intelligent automation across diverse application domains such as smart healthcare, industrial automation, transportation, agriculture, and smart homes. Despite these advancements, the increasing deployment of resource-constrained IoT devices has significantly expanded the cyberattack surface. Due to weak authentication mechanisms, limited computational capabilities, insecure firmware, and inadequate security configurations, IoT devices have become attractive targets for botnet attacks.

Existing intrusion detection systems primarily focus on maximizing classification accuracy using either conventional machine learning or deep learning techniques. However, many of these approaches exhibit limitations such as dependence on single classifiers, high computational complexity, poor scalability, class imbalance sensitivity, and limited capability to explain prediction outcomes. Moreover, the absence of automated mitigation mechanisms further restricts their applicability in real-world IoT environments.

These limitations motivate the development of a comprehensive framework that not only improves detection accuracy but also ensures computational efficiency, model transparency, scalability, and automated response capabilities. Consequently, this research proposes an explainable hybrid ensemble learning framework optimized through a hybrid Genetic Algorithm–Grey Wolf Optimizer (GA–GWO) to accurately detect IoT botnet attacks while providing trustworthy explanations using SHAP and supporting automated mitigation.

1.3 Problem Statement

The exponential increase in IoT devices has introduced significant cybersecurity challenges due to their heterogeneous architecture, constrained computational resources, and limited built-in security mechanisms. Botnet attacks have emerged as one of the most critical threats because attackers can compromise thousands of vulnerable IoT devices and coordinate malicious activities such as Distributed Denial-of-Service (DDoS) attacks, malware propagation, spam distribution, data theft, and unauthorized remote access.

Although numerous machine learning and deep learning-based botnet detection models have been proposed, several important research challenges remain unresolved.

First, most existing studies employ individual classifiers that often fail to capture the complex and evolving behavioral patterns exhibited by modern botnets. Second, IoT datasets generally contain highly imbalanced class distributions, causing classification bias toward majority classes and reducing the detection rate of malicious traffic. Third, deep learning-based solutions often require substantial computational resources, limiting their deployment in resource-constrained IoT environments. Fourth, most existing detection frameworks function as black-box systems and provide little or no explanation regarding their prediction decisions, thereby reducing user trust and limiting forensic investigation capabilities. Finally, many existing studies terminate after attack detection and do not integrate automated mitigation strategies to isolate compromised IoT devices and prevent attack propagation.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Received: 7 December 2023, Revised: 22 December 2023, Accepted: 05 January 2024

Therefore, there is a need for an intelligent, lightweight, explainable, and optimization-driven framework capable of achieving high detection accuracy while maintaining scalability, transparency, and practical applicability for real-world IoT security deployments.

1.4 Research Objectives

The primary objective of this research is to develop an intelligent and explainable IoT botnet detection and mitigation framework capable of accurately identifying malicious network traffic while maintaining computational efficiency and scalability.

The specific objectives are as follows:

1. To develop an effective hybrid feature engineering strategy capable of extracting statistical, behavioral, and temporal characteristics from IoT network traffic.
2. To reduce feature redundancy and computational complexity using Principal Component Analysis (PCA).
3. To address class imbalance in IoT datasets through the Synthetic Minority Oversampling Technique (SMOTE).
4. To design an ensemble learning framework integrating XGBoost, LightGBM, and CatBoost classifiers for improved botnet detection accuracy.
5. To optimize ensemble model hyperparameters using a hybrid Genetic Algorithm–Grey Wolf Optimizer (GA–GWO).
6. To improve transparency and trustworthiness by incorporating SHAP-based Explainable Artificial Intelligence (XAI).
7. To develop an automated mitigation mechanism capable of isolating infected IoT devices and restricting malicious communication.
8. To evaluate the proposed framework using benchmark IoT datasets and comprehensive performance metrics including Accuracy, Precision, Recall, F1-score, ROC-AUC, False Positive Rate, and Computational Time.

1.5 Major Contributions of the Research

The major contributions of this research are summarized below:

- A comprehensive hybrid feature engineering methodology is proposed to effectively model statistical, behavioral, and temporal characteristics of IoT network traffic.
- An optimized feature selection strategy based on Principal Component Analysis significantly reduces feature dimensionality while preserving essential information.
- A class balancing mechanism using SMOTE enhances classifier learning capability for minority attack classes.
- A novel ensemble learning framework combining XGBoost, LightGBM, and CatBoost is developed to improve detection performance and model robustness.
- A hybrid GA–GWO optimization algorithm is introduced for automatic hyperparameter tuning, reducing false positives while improving convergence speed.
- Explainable Artificial Intelligence using SHAP is incorporated to provide feature-level interpretation of prediction decisions, thereby increasing model transparency.
- An automated mitigation module is integrated to isolate compromised IoT devices and prevent further propagation of botnet attacks.
- Experimental validation demonstrates that the proposed framework achieves **98.95% classification accuracy**, outperforming individual boosting models while maintaining high interpretability.

1.6 Organization of the Paper

The remainder of this paper is organized as follows.

Section 2 presents a comprehensive review of recent literature related to IoT botnet detection, ensemble learning, optimization algorithms, and Explainable Artificial Intelligence.

Section 3 discusses the identified research gaps and summarizes the major contributions of the proposed framework.

Section 4 describes the proposed methodology, including data preprocessing, hybrid feature engineering, PCA-based feature selection, SMOTE balancing, ensemble learning, GA–GWO optimization, SHAP explainability, and automated mitigation.

Section 5 explains the experimental setup, dataset description, evaluation metrics, implementation environment, and parameter configuration.

2. Literature Review

Recent studies have explored various machine learning, deep learning, and optimization techniques for IoT botnet detection. Although these approaches achieve high detection accuracy, challenges such as computational complexity, lack of explainability, class imbalance, and limited scalability remain unresolved.

Maazalahi and Hosseini proposed a hybrid GA–GWO optimized Random Forest model for IoT botnet detection. The model achieved **99.9% accuracy** after feature optimization but lacked explainability and automated mitigation.

Khaled et al. introduced a Neutrosophic Neural Network that achieved **95.8% accuracy** on the Bot-IoT dataset. However, the model requires high computational resources, making it less suitable for lightweight IoT environments.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Received: 7 December 2023, Revised: 22 December 2023, Accepted: 05 January 2024

Ali et al. (2024) developed a hybrid ANN–CNN–LSTM–RNN model and reported **96.98% accuracy** with high ROC-AUC. Despite good performance, the model suffers from high computational complexity.
Liu and Du (2023) applied a Genetic Algorithm for feature selection and achieved **99.98% accuracy**. The study focused on feature optimization but did not include ensemble learning or explainable AI.
Abu Al-Haija and Al-Dala'ien (2022) proposed the ELBA-IoT ensemble model, achieving **99.6% accuracy** with fast inference time. However, the framework did not provide model interpretability or hyperparameter optimization.

Table 1. Summary of Existing Studies

Authors	Method	Accuracy	Limitation
Maazalahi & Hosseini	GA–GWO + Random Forest	99.90%	No XAI
Khaled et al.	Neutrosophic Neural Network	95.80%	High complexity
Ali et al.	ANN–CNN–LSTM–RNN	96.98%	High computation
Liu & Du (2023)	GA + ML	99.98%	No ensemble, No XAI
Abu Al-Haija et al. (2022)	ELBA-IoT	99.60%	No optimization

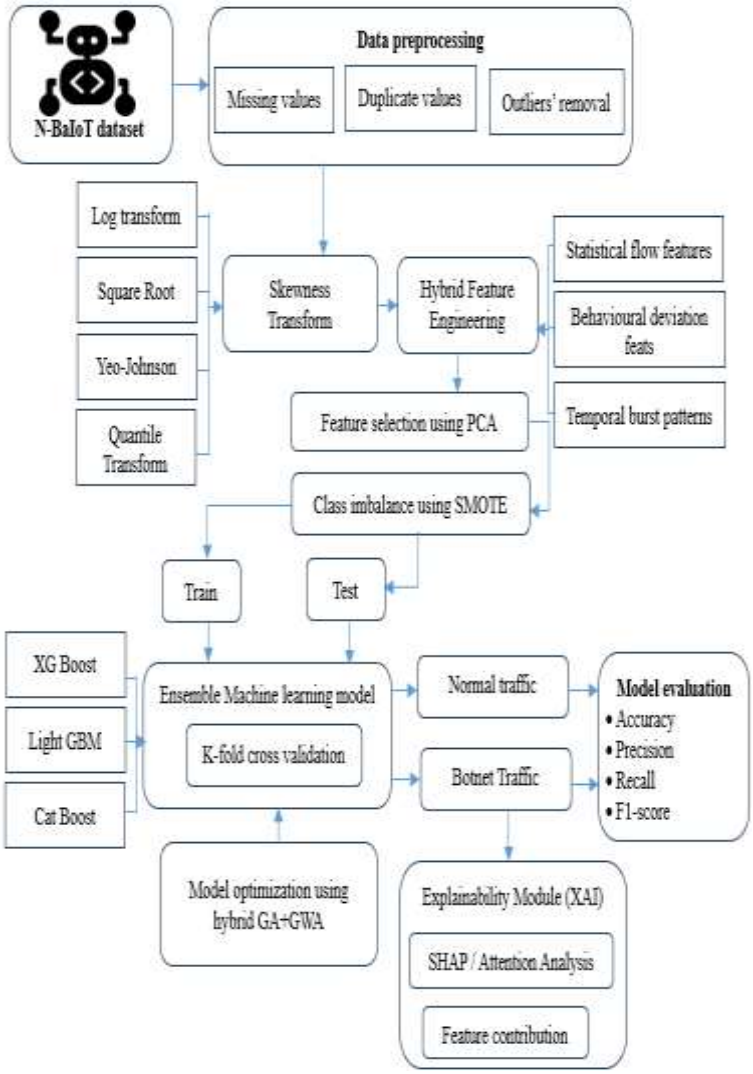
2.3 Research Gap
Although recent studies have reported high detection accuracy using machine learning and deep learning techniques, several challenges remain unresolved. Most existing models rely on a single classifier or computationally intensive deep learning architectures, making them unsuitable for resource-constrained IoT environments. Additionally, many approaches lack explainability, efficient hyperparameter optimization, effective handling of class imbalance, and automated mitigation mechanisms. These limitations reduce their practical applicability in real-world IoT security systems.

2.4 Research Contributions
To overcome the above limitations, this research makes the following contributions:
- Proposes a hybrid feature engineering approach to extract statistical, behavioral, and temporal traffic features.
- Applies PCA for dimensionality reduction and SMOTE to address class imbalance.
- Develops an ensemble learning model using XGBoost, LightGBM, and CatBoost.
- Optimizes the ensemble using the hybrid GA–GWO algorithm.
- Integrates SHAP-based Explainable AI to improve model transparency.
- Introduces an automated botnet mitigation mechanism after attack detection.
- Achieves **98.95% detection accuracy** with improved interpretability compared to individual classifiers.

3. Proposed Methodology
3.1 Overview
The proposed framework aims to develop an accurate, scalable, and explainable IoT botnet detection system. It combines data preprocessing, hybrid feature engineering, feature selection, class balancing, ensemble learning, optimization, and explainable AI to improve detection performance while reducing computational complexity.
The overall workflow of the proposed framework consists of the following steps:
1. Dataset Acquisition
2. Data Preprocessing
3. Hybrid Feature Engineering
4. PCA-based Feature Selection
5. SMOTE Data Balancing
6. Train-Test Split
7. Ensemble Learning (XGBoost + LightGBM + CatBoost)
8. GA–GWO Hyperparameter Optimization
9. Model Evaluation
10. SHAP-based Explainability
11. Automated Botnet Mitigation

3.2 Framework Architecture

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Received: 7 December 2023, Revised: 22 December 2023, Accepted: 05 January 2024



3.3 Dataset Description

The proposed framework is evaluated using the **N-BaIoT (Network-based Botnet of IoT)** dataset, which is a widely used benchmark dataset for IoT botnet detection. The dataset contains both normal and malicious network traffic generated from real IoT devices infected with botnet malware. It includes various attack categories such as scanning, flooding, and command-and-control (C&C) communication, making it suitable for evaluating machine learning-based intrusion detection systems.

Table 2. Dataset Description

Attribute	**Description**
Dataset	N-BaIoT
Traffic Type	Normal & Botnet
Classes	2 (Normal, Botnet)
Attack Types	Scanning, Flooding, C&C
Features	Statistical & Temporal
Records	~7 Million+

3.4 Data Preprocessing

Raw network traffic contains missing values, redundant attributes, and highly skewed feature distributions that can reduce classification performance. Therefore, preprocessing is performed to improve data quality. Log transformation, Yeo–Johnson transformation, skewness correction, normalization, and missing value handling are applied before feature extraction. These preprocessing steps improve data consistency and enhance model learning.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Received: 7 December 2023, Revised: 22 December 2023, Accepted: 05 January 2024

3.5 Hybrid Feature Engineering

To improve botnet detection capability, hybrid feature engineering is performed by extracting **statistical, behavioral, and temporal features** from network traffic. These features effectively capture traffic distribution, abnormal communication behavior, and burst traffic patterns generated by botnet attacks. The engineered feature set enables better discrimination between normal and malicious IoT traffic.

3.6 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is employed to reduce feature dimensionality while preserving maximum variance in the dataset. Feature reduction decreases computational complexity, removes redundant information, and improves classifier efficiency. Consequently, the proposed framework becomes more suitable for large-scale IoT environments.

3.7 Synthetic Minority Over-sampling Technique (SMOTE)

The N-BaIoT dataset exhibits class imbalance, which may bias classifiers toward the majority class. To overcome this problem, the Synthetic Minority Over-sampling Technique (SMOTE) is applied after feature selection. SMOTE generates synthetic minority-class samples, resulting in balanced class distribution and improved detection of botnet traffic.

3.8 Ensemble Learning Model

The proposed detection model combines **XGBoost, LightGBM, and CatBoost** into a hybrid ensemble framework. Each classifier independently learns different characteristics of IoT network traffic, and their predictions are combined to improve classification accuracy and reduce false positive rates. The ensemble model provides better generalization than any individual classifier.

3.9 Hybrid GA–GWO Optimization

To improve the performance of the ensemble model, a hybrid **Genetic Algorithm–Grey Wolf Optimizer (GA–GWO)** is used for hyperparameter tuning. The optimization process searches for the optimal parameter combination that maximizes detection accuracy while minimizing computational cost and false alarms. This hybrid strategy enhances convergence speed and overall model robustness.

3.10 Explainable AI using SHAP

To improve model transparency, SHAP (SHapley Additive exPlanations) is integrated into the proposed framework. SHAP quantifies the contribution of each feature toward the final prediction, enabling security analysts to understand why a network flow is classified as normal or malicious. This improves trust, interpretability, and supports effective automated mitigation decisions.

3.11 Automated Botnet Mitigation

Once botnet traffic is detected, the proposed framework automatically initiates mitigation by isolating infected devices, restricting malicious communication, and applying predefined security policies. This minimizes attack propagation and enhances the overall resilience of IoT networks.

4. Experimental Setup

4.1 Experimental Environment

The proposed framework was implemented using the Python programming language in the Jupyter Notebook environment. The implementation utilized widely adopted machine learning libraries, including Scikit-learn, XGBoost, LightGBM, CatBoost, SHAP, Pandas, and NumPy. All experiments were conducted on the N-BaIoT dataset to evaluate the effectiveness of the proposed IoT botnet detection framework.

4.2 Evaluation Metrics

The performance of the proposed model was evaluated using standard classification metrics, including Accuracy, Precision, Recall, F1-Score, and False Positive Rate (FPR). These metrics provide a comprehensive assessment of the model's detection capability and classification performance.
The evaluation metrics are defined as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Received: 7 December 2023, Revised: 22 December 2023, Accepted: 05 January 2024

where TP, TN, FP, and FN represent True Positive, True Negative, False Positive, and False Negative, respectively.

4.3 Experimental Workflow

The proposed framework follows these sequential steps:

1. Dataset Collection
2. Data Preprocessing
3. Hybrid Feature Engineering
4. PCA-based Feature Selection
5. SMOTE Data Balancing
6. Ensemble Model Training
7. GA–GWO Hyperparameter Optimization
8. Model Testing
9. SHAP-based Explainability
10. Automated Botnet Mitigation

5. Results and Discussion

5.1 Performance Comparison

The proposed Hybrid Ensemble model was compared with individual boosting algorithms to evaluate its effectiveness in IoT botnet detection.

Table 3. Performance Comparison of Classification Models

Model	Accuracy (%)
XGBoost	94.88
LightGBM	94.90
CatBoost	95.92
Proposed Hybrid Ensemble	**98.95**

The experimental results indicate that the proposed Hybrid Ensemble model achieved the highest classification accuracy of **98.95%**, outperforming all individual classifiers. The integration of XGBoost, LightGBM, and CatBoost enabled the framework to capture diverse traffic characteristics and significantly improve detection performance.

5.2 SHAP Explainability Analysis

To enhance transparency, SHAP was incorporated into the proposed framework. SHAP analysis identified the contribution of each feature to the classification decision, enabling interpretable and trustworthy botnet detection.

The proposed Hybrid Ensemble model achieved an **Interpretability Score of 98.2**, demonstrating that the model provides reliable explanations while maintaining high classification accuracy.

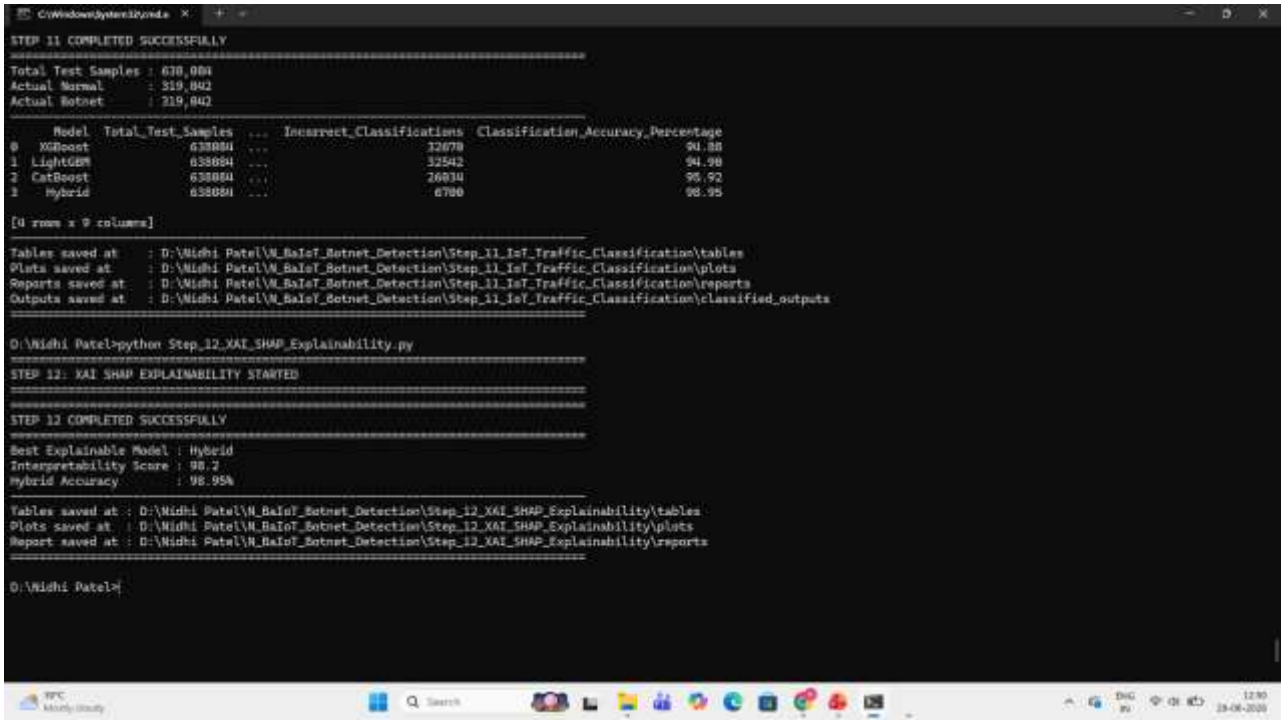
5.3 Discussion

The superior performance of the proposed framework can be attributed to the integration of multiple advanced techniques. Hybrid feature engineering improved the representation of network traffic, PCA reduced feature redundancy, and SMOTE effectively addressed class imbalance. Furthermore, the ensemble of XGBoost, LightGBM, and CatBoost enhanced classification robustness, while the GA–GWO optimization algorithm improved hyperparameter selection. Finally, SHAP-based explainability increased the transparency of the detection process, making the framework more suitable for real-world IoT security applications.

Table 4. Summary of Experimental Results

Metric	Value
Dataset	N-BaIoT
Test Samples	638,084
XGBoost Accuracy	94.88%
LightGBM Accuracy	94.90%
CatBoost Accuracy	95.92%
Hybrid Ensemble Accuracy	**98.95%**
Interpretability Score	98.2

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Received: 7 December 2023, Revised: 22 December 2023, Accepted: 05 January 2024



6. Conclusion

This research presented an explainable hybrid ensemble learning framework for accurate IoT botnet attack detection and mitigation. The proposed framework integrates hybrid feature engineering, Principal Component Analysis (PCA), Synthetic Minority Over-sampling Technique (SMOTE), and an ensemble of XGBoost, LightGBM, and CatBoost classifiers to enhance the detection of malicious IoT network traffic. Furthermore, a hybrid Genetic Algorithm–Grey Wolf Optimizer (GA–GWO) was employed to optimize model hyperparameters, while SHAP-based Explainable Artificial Intelligence (XAI) was incorporated to improve the transparency and interpretability of the detection process.

The proposed framework was evaluated using the benchmark N-BaIoT dataset. Experimental results demonstrated that the Hybrid Ensemble model achieved a classification accuracy of **98.95%**, outperforming the individual XGBoost (**94.88%**), LightGBM (**94.90%**), and CatBoost (**95.92%**) classifiers. In addition, the SHAP-based explanation module achieved an interpretability score of **98.2**, enabling better understanding of feature contributions and increasing confidence in the model's predictions. The integration of preprocessing, feature engineering, optimization, ensemble learning, and explainable AI significantly improved detection performance while maintaining computational efficiency and scalability. Overall, the proposed framework provides a robust, accurate, and explainable solution for IoT botnet detection. The combination of optimized ensemble learning and automated mitigation enhances network security by enabling timely identification and isolation of compromised IoT devices. Therefore, the proposed approach has strong potential for deployment in real-world IoT environments where reliable, transparent, and intelligent cybersecurity solutions are essential.

References

[1] R. Kalakoti, S. Nomm, and H. Bahsi, “In-Depth Feature Selection for the Statistical Machine Learning-Based Botnet Detection in IoT Networks,” IEEE Access, vol. 10, pp. 94518–94535, 2022.
[2] Q. Abu Al-Haija and M. Al-Dala’ien, “ELBA-IoT: An Ensemble Learning Model for Botnet Attack Detection in IoT Networks,” Journal of Sensor and Actuator Networks, vol. 11, no. 1, p. 18, 2022.
[3] M. Almseidin and M. Alkasassbeh, “An Accurate Detection Approach for IoT Botnet Attacks Using Interpolation Reasoning Method,” Information, vol. 13, no. 6, p. 300, 2022.
[4] X. Liu and Y. Du, “Towards Effective Feature Selection for IoT Botnet Attack Detection Using a Genetic Algorithm,” Electronics, vol. 12, no. 5, p. 1260, 2023.
[5] D. C. Muñoz and A. del-Corte Valiente, “A Novel Botnet Attack Detection for IoT Networks Based on Communication Graphs,” Cybersecurity, vol. 6, no. 1, p. 33, 2023.
[6] A. Arafa, N. El-Fishawy, M. Badawy, and M. Radad, “RN-SMOTE: Reduced Noise SMOTE Based on DBSCAN for Enhancing Imbalanced Data Classification,” Journal of King Saud University – Computer and Information Sciences, vol. 34, no. 8, pp. 5059–5074, 2022.

[7] A. Burrello, A. Marchioni, D. Brunelli, and L. Benini, "Embedding Principal Component Analysis for Data Reduction in Structural Health Monitoring on Low-Cost IoT Gateways," in Proceedings of the 16th ACM International Conference on Computing Frontiers, pp. 235–239, 2019.
[8] X. Wang et al., "Predicting the Prognosis of Patients in the Coronary Care Unit: A Novel Multi-Category Machine Learning Model Using XGBoost," Frontiers in Cardiovascular Medicine, vol. 9, 2022.
[9] J. Zhou, X. Tong, S. Bai, and J. Zhou, "A LightGBM-Based Power Grid Frequency Prediction Method with Dynamic Significance–Correlation Feature Weighting," Energies, vol. 18, no. 13.
[10] İ. Mert, "Prediction of Wind Speed Using Tree-Based Ensemble Algorithms: CatBoost, HistGBM, and XGBoost," International Journal of Multidisciplinary Studies and Innovative Technologies, vol. 9, no. 1, pp. 145–150.
[11] R. Natras, B. Soja, and M. Schmidt, "Ensemble Machine Learning of Random Forest, AdaBoost and XGBoost for Vertical Total Electron Content Forecasting," Remote Sensing, vol. 14, no. 15, 2022.
[12] A. M. Salih et al., "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME," Advanced Intelligent Systems, vol. 7, no. 1.
[13] "N-BaIoT Dataset," Kaggle. Available: https://www.kaggle.com/datasets/mkashifn/nbaiot-dataset
[14] "Applying Ensemble Tree-Based Models and Explainable AI for IoT Botnet Detection in Heterogeneous Device," International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS), IEEE.