e-ISSN 2549-7472
Edumatic: Jurnal Pendidikan Informatika
Volume 10, Issue 1, **150-159.**
https://e-journal.hamzanwadi.ac.id/index.php/edumatic
DOI: 10.29408/edumatic.v10i1.34012

Mutual Information-Driven Feature Selection for Efficient DDoS Detection Using Modern Boosting Ensembles

Muhammad Febrian Fansuri [1,*] **, Kusrini** [1]

[1] Universitas Amikom Yogyakarta, Indonesia
[*] Corresponding author: Muhammad Febrian Fansuri, Universitas Amikom Yogyakarta, Indonesia
✉ mffansuri@students.amikom.ac.id

Received: 25 January 2026 | Revised: 11 February 2026 | Accepted: 7 April 2026 | Published: 17 April 2026

Abstract
Distributed Denial of Service (DDoS) attacks generate high-dimensional network traffic that poses significant challenges for machine learning-based detection systems in terms of predictive accuracy and computational efficiency. This study presents a systematic evaluation of Mutual Information (MI) based feature selection applied to three modern boosting algorithms, namely XGBoost, LightGBM, and CatBoost, using the CIC-DDoS2019 dataset. A controlled experimental design was employed, where data partitioning was performed prior to resampling, and SMOTE was applied exclusively to the training set to prevent data leakage. Feature selection was conducted by identifying the top 25 features based on MI score saturation analysis. The results demonstrate that MI-based feature selection consistently improves classification performance while substantially reducing training time across all models. Among the evaluated methods, LightGBM achieves the best trade-off between accuracy and computational efficiency, reaching an accuracy of 99.88% with significantly reduced training cost. These findings indicate that feature quality plays a critical role in shaping the learning behaviour of boosting algorithms and that MI-based feature selection functions as a structural mechanism for enhancing model stability and scalability in high-dimensional DDoS detection scenarios.

Keywords: ddos detection; ensemble boosting; feature selection; machine learning; mutual information

To cite this article: Fansuri, M. F., & Kusrini, K. (2026). Mutual Information-Driven Feature Selection for Efficient DDoS Detection Using Modern Boosting Ensembles. *Edumatic: Jurnal Pendidikan Informatika*, *10*(1), 150–159. https://doi.org/10.29408/edumatic.v10i1.34012

INTRODUCTION

Digital transformation across critical infrastructure, e-governance, and enterprise systems has intensified reliance on uninterrupted network availability. Within this landscape, Distributed Denial of Service (DDoS) attacks have evolved into increasingly sophisticated and persistent threats (Falowo et al, 2024; Hasan et al., 2024), leveraging large-scale botnets and advanced amplification techniques to disrupt essential services (Hayat et al., 2022; Li et al., 2021; Poonia & Tinker, 2025). Such attacks degrade service reliability and generate substantial financial and operational losses, thereby reinforcing the need for intelligent and adaptive detection mechanisms. Conventional defense approaches based on static thresholds and signature-based detection exhibit limited capability in identifying low-rate, stealthy, and multi-vector attack patterns (Adedeji et al., 2023; Jaafar et al., 2025). Machine learning (ML) has consequently emerged as a viable paradigm due to its ability to model complex traffic behavior

and detect previously unseen attack patterns (Al-Shareeda et al., 2023; Issa & Albayrak, 2023; Tymoshchuk et al., 2024).

A broad spectrum of ML techniques has been investigated for DDoS detection. Traditional single-model approaches, including Support Vector Machines (SVM), have demonstrated moderate predictive performance with accuracy levels around 95% (Ali et al., 2023; Guido et al., 2024; Martinović et al., 2025). Ensemble learning methods have shown superior performance by integrating multiple base learners to improve generalization. In particular, boosting-based algorithms such as XGBoost, LightGBM, and AdaBoost have consistently achieved accuracy levels exceeding 98% in high-dimensional network traffic datasets. These models benefit from iterative error correction, enabling them to refine predictions across successive learning stages.

Feature quality constitutes a critical factor influencing the effectiveness of ML-based detection systems. High dimensional network traffic data frequently contain redundant and noisy attributes that increase computational complexity and degrade model stability. Feature selection methods have therefore been extensively employed to identify relevant attributes and reduce dimensionality. Mutual Information (MI) has been widely recognized as an effective filter-based approach due to its ability to quantify statistical dependency between features and target variables (Mallidi & Ramisetty, 2025; Shah et al., 2026). Empirical studies have demonstrated that MI-based feature selection improves classification accuracy while reducing computational overhead (Alduailij et al., 2022; Yuan & Yang, 2025).

The interaction between feature selection and boosting algorithms presents a critical yet underexplored aspect of DDoS detection modeling (Hirsi et al., 2025; Hossain et al., 2024; MA et al., 2023). Boosting algorithms operate through sequential error correction, where misclassified instances are iteratively reweighted during training. This mechanism inherently increases sensitivity to noisy and redundant features, as such features may be repeatedly emphasized and propagated across boosting iterations. As a result, the presence of low-quality features can lead to increased computational cost and reduced generalization performance (Zhang et al., 2025). A systematic examination of how MI-based feature selection influences the learning behavior and efficiency of modern boosting models remains limited within existing literature.

This study develops a systematic comparative framework that integrates MI-based feature selection with three state-of-the-art boosting algorithms, namely XGBoost, LightGBM, and CatBoost. The analysis focuses on evaluating predictive performance and computational efficiency before and after feature selection, while examining how feature quality influences the learning dynamics of each boosting model. The investigation is conducted within high-dimensional DDoS detection scenarios, including environments such as software-defined networks (SDN) (Ma et al, 2023), where latency and resource efficiency are critical considerations (Han et al., 2024).

The objective of this study is to evaluate the impact of MI-based feature selection on the performance of modern boosting algorithms in high-dimensional DDoS detection. The study examines improvements in predictive performance and computational efficiency, analyzes the interaction between feature quality and boosting learning behavior, and identifies the model that achieves the optimal balance between accuracy and efficiency.

METHOD

This study employs a quantitative experimental methodology designed to develop and evaluate a DDoS attack detection model through the integration of ensemble boosting algorithms with mutual information-based feature selection. The process commences with data acquisition, utilizing the publicly available CIC-DDoS2019 benchmark dataset (Talukder &

Uddin, 2023). The overall research workflow follows a structured sequence of stages, as illustrated in the research framework shown in Figure 1.

Following data collection, preprocessing was conducted to ensure data integrity, including handling missing values, removing duplicates, normalizing numerical features, and encoding categorical variables. The dataset was then split into training and testing sets (80:20) using stratified sampling. The Synthetic Minority Over-Sampling Technique (SMOTE) was applied to the training set to generate synthetic minority samples, while the test set remained unchanged to prevent data leakage.

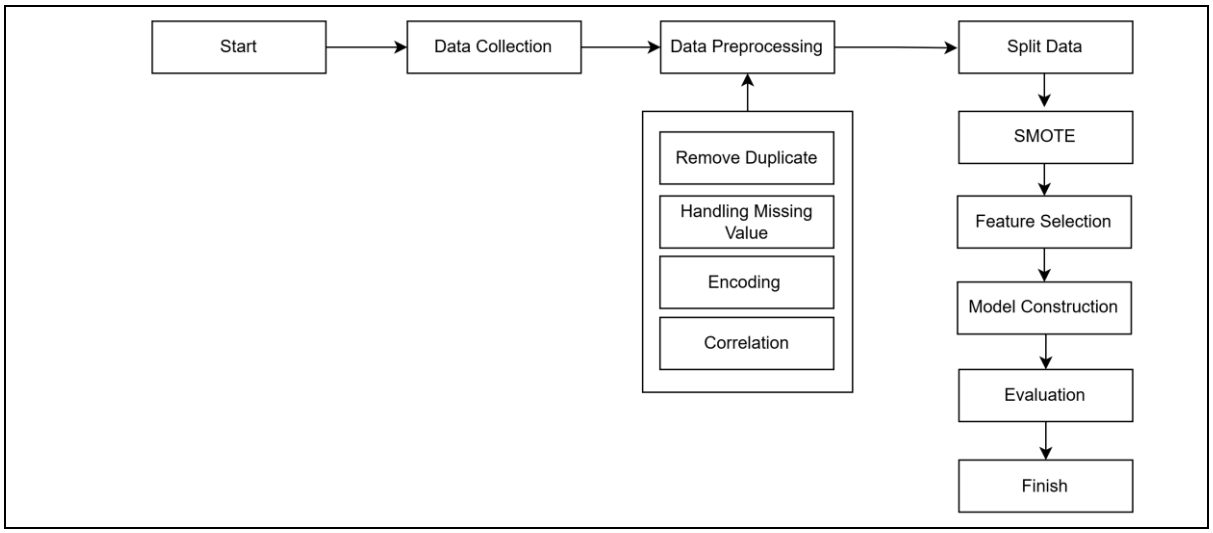

Figure 1. Research workflow

Mutual information based feature selection is applied to identify the most informative features and reduce dimensionality. The value of top_k = 25 was selected based on MI score saturation analysis, where additional features beyond the top-ranked subset contributed marginal information gain. Equation (1) quantifies the amount of information shared between a feature and the target variable. In practice, it measures how much knowing the value of a feature reduces uncertainty about the class label. A higher value indicates that the feature carries more relevant information for distinguishing between attack and benign traffic, whereas a value close to zero suggests little or no contribution to the prediction task.

$$I\ (X;Y) = \sum P\ (x,y) log \frac{P(x,y)}{P(x)P(y)} \quad (1)$$

Ensemble learning has gained attention due to its ability to combine multiple weak learners into a robust predictive model. Three gradient boosting algorithms were evaluated in this study: XGBoost, known for its regularization mechanisms and efficient tree pruning; LightGBM, which employs a leaf-wise tree growth strategy and histogram-based learning for computational efficiency; and CatBoost, which is designed to handle categorical features with minimal preprocessing.

The performance of the proposed DDoS attack detection model is assessed using standard classification metrics derived from the confusion matrix. Accuracy is used to measure the overall correctness of the model, while precision evaluates the reliability of the predicted DDoS attack instances. Recall reflects the model's capability to correctly identify actual attack traffic, and the F-score provides a balanced evaluation by considering both precision and recall.

$$Accuracy = \frac{TP}{TP + TN + FP + FN} \quad (2)$$

The accuracy metric, defined in Equation (2), measures the overall correctness of a classification model as the ratio of correctly predicted instances (TP and TN) to the total number of observations. It provides a high-level evaluation of the model's ability to distinguish between attack and benign traffic in DDoS detection. All experiments were conducted using a fixed random seed to ensure reproducibility and consistency.

RESULTS AND DISCUSSION

Results

The dataset used in this study consists of 431,371 records with 80 features, where 333,540 records are labeled as DDoS attacks and 97,831 records are classified as benign traffic. As illustrated in Figure 2. Before analysis, preprocessing steps were applied, including handling missing values, data cleaning, and normalization. The preprocessing phase is visualized in the provided workflow diagram, beginning with Data Collection and proceeding systematically to Evaluation.

#	Column	Non-Null Count	Dtype
0	Unnamed: 0	431371 non-null	int64
1	Protocol	431371 non-null	int64
2	Flow Duration	431371 non-null	int64
3	Total Fwd Packets	431371 non-null	int64
4	Total Backward Packets	431371 non-null	int64
5	Fwd Packets Length Total	431371 non-null	float64
6	Bwd Packets Length Total	431371 non-null	float64
7	Fwd Packet Length Max	431371 non-null	float64
8	Fwd Packet Length Min	431371 non-null	float64
9	Fwd Packet Length Mean	431371 non-null	float64
10	Fwd Packet Length Std	431371 non-null	float64
11	Bwd Packet Length Max	431371 non-null	float64
12	Bwd Packet Length Min	431371 non-null	float64
13	Bwd Packet Length Mean	431371 non-null	float64
14	Bwd Packet Length Std	431371 non-null	float64
15	Flow Bytes/s	431371 non-null	float64
16	Flow Packets/s	431371 non-null	float64
17	Flow IAT Mean	431371 non-null	float64
18	Flow IAT Std	431371 non-null	float64
19	Flow IAT Max	431371 non-null	float64
20	Flow IAT Min	431371 non-null	float64
21	Fwd IAT Total	431371 non-null	float64
22	Fwd IAT Mean	431371 non-null	float64
23	Fwd IAT Std	431371 non-null	float64
24	Fwd IAT Max	431371 non-null	float64
25	Fwd IAT Min	431371 non-null	float64
26	Bwd IAT Total	431371 non-null	float64
27	Bwd IAT Mean	431371 non-null	float64
28	Bwd IAT Std	431371 non-null	float64
29	Bwd IAT Max	431371 non-null	float64
30	Bwd IAT Min	431371 non-null	float64
31	Fwd PSH Flags	431371 non-null	int64
32	Bwd PSH Flags	431371 non-null	int64
33	Fwd URG Flags	431371 non-null	int64
34	Bwd URG Flags	431371 non-null	int64
35	Fwd Header Length	431371 non-null	int64
36	Bwd Header Length	431371 non-null	int64
37	Fwd Packets/s	431371 non-null	float64
38	Bwd Packets/s	431371 non-null	float64
39	Packet Length Min	431371 non-null	float64
40	Packet Length Max	431371 non-null	float64
41	Packet Length Mean	431371 non-null	float64
42	Packet Length Std	431371 non-null	float64
43	Packet Length Variance	431371 non-null	float64
44	FIN Flag Count	431371 non-null	int64
45	SYN Flag Count	431371 non-null	int64
46	RST Flag Count	431371 non-null	int64
47	PSH Flag Count	431371 non-null	int64
48	ACK Flag Count	431371 non-null	int64
49	URG Flag Count	431371 non-null	int64
50	CWE Flag Count	431371 non-null	int64
51	ECE Flag Count	431371 non-null	int64
52	Down/Up Ratio	431371 non-null	float64
53	Avg Packet Size	431371 non-null	float64
54	Avg Fwd Segment Size	431371 non-null	float64
55	Avg Bwd Segment Size	431371 non-null	float64
56	Fwd Avg Bytes/Bulk	431371 non-null	int64
57	Fwd Avg Packets/Bulk	431371 non-null	int64
58	Fwd Avg Bulk Rate	431371 non-null	int64
59	Bwd Avg Bytes/Bulk	431371 non-null	int64
60	Bwd Avg Packets/Bulk	431371 non-null	int64
61	Bwd Avg Bulk Rate	431371 non-null	int64
62	Subflow Fwd Packets	431371 non-null	int64
63	Subflow Fwd Bytes	431371 non-null	int64
64	Subflow Bwd Packets	431371 non-null	int64
65	Subflow Bwd Bytes	431371 non-null	int64
66	Init Fwd Win Bytes	431371 non-null	int64
67	Init Bwd Win Bytes	431371 non-null	int64
68	Fwd Act Data Packets	431371 non-null	int64
69	Fwd Seg Size Min	431371 non-null	int64
70	Active Mean	431371 non-null	float64
71	Active Std	431371 non-null	float64
72	Active Max	431371 non-null	float64
73	Active Min	431371 non-null	float64
74	Idle Mean	431371 non-null	float64
75	Idle Std	431371 non-null	float64
76	Idle Max	431371 non-null	float64
77	Idle Min	431371 non-null	float64
78	Label	431371 non-null	object
79	Class	431371 non-null	object

Figure 2. Dataset overview

Figure 3 presents the correlation heatmap, which reveals strong dependencies (correlation > 0.90) among packet length statistics, flow-based features, and inter-arrival time metrics, indicating substantial feature redundancy. Such high multicollinearity increases training cost and can lead to unstable split decisions in tree-based boosting models. Therefore, feature selection is necessary to reduce dimensionality, improve computational efficiency, and stabilize model learning for DDoS detection.

A significant class imbalance was observed in the original dataset, with Class 1 markedly outnumbering Class 0; therefore, the SMOTE was employed to address this issue. Prior to resampling, the dataset consisted of 331,507 instances in Class 1 and 97,831 instances in Class 0, indicating a substantial disparity between the two classes (see Figure 4). As illustrated in

Figure 3, the application of SMOTE effectively corrected this imbalance by generating synthetic samples for the minority class. After the resampling process, both Class 1 and Class 0 were balanced with 331,507 instances each, resulting in an equitable class distribution. This resampling process reduces the dominance of the majority class and allows the model to learn decision boundaries more effectively.

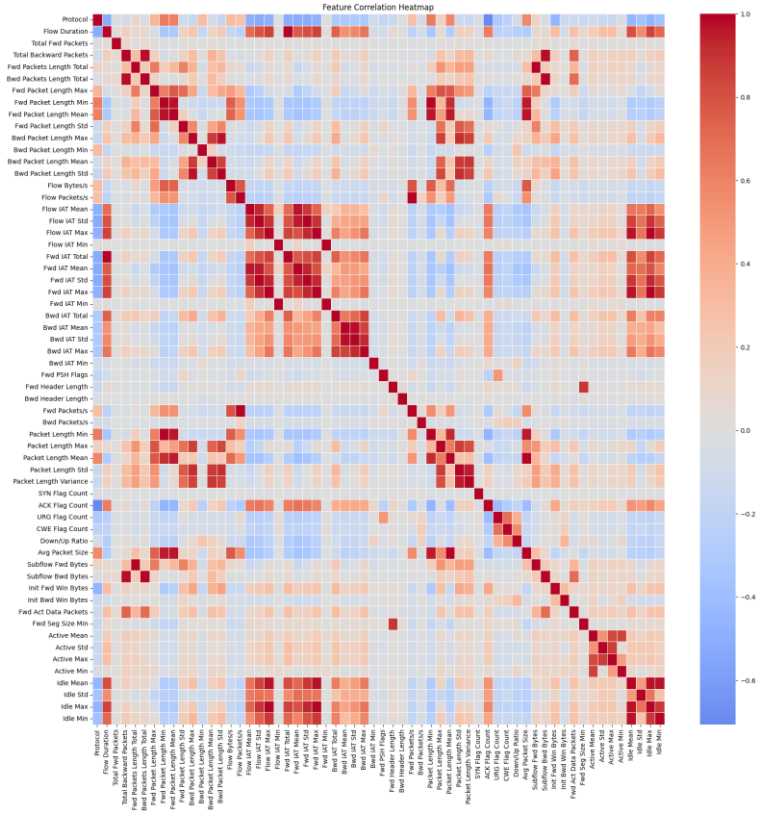

Figure 3. Heatmap correlation

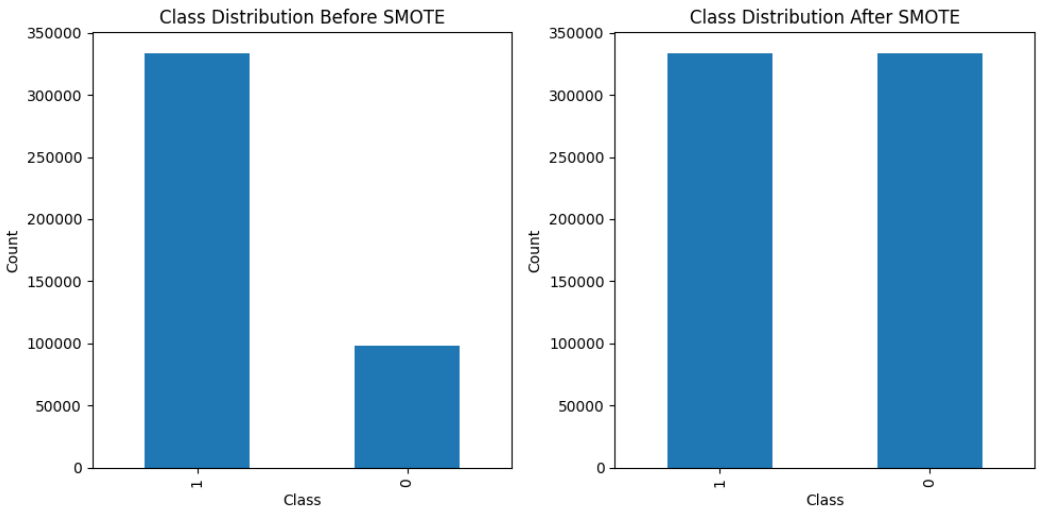

Figure 4. Smote

Mutual Information (MI) was utilized to identify the most informative features relevant to DDoS detection. The top 25 most informative features are visualized in Figure 5. MI quantifies the dependency between each feature and the target label (attack vs. benign). By selecting features with the highest MI scores, the model focuses on the most discriminative

attributes, which improves detection accuracy while reducing computational overhead and the risk of overfitting.

Tables 1 and 2 present a comparative evaluation of the performance of XGBoost, LightGBM, and CatBoost models using all features and with feature selection applied. As shown in Table 1, when all features are utilized, LightGBM achieves the highest performance across most evaluation metrics, including accuracy, precision, recall, and F1-score, while also demonstrating a relatively low computational time. XGBoost exhibits comparable performance, although slightly lower than LightGBM, whereas CatBoost records the lowest F1-score and requires substantially longer execution time.

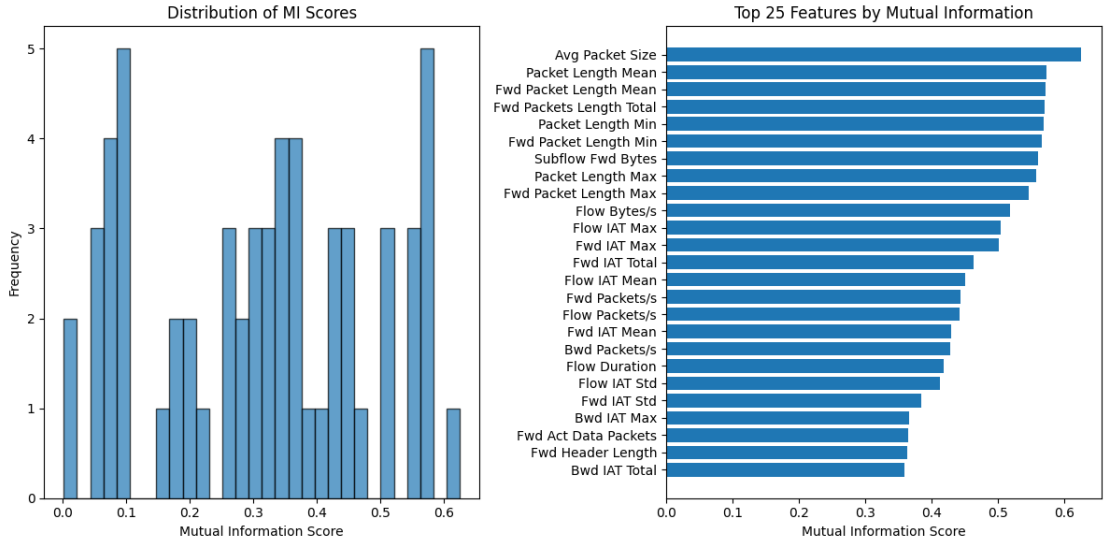

Figure 5. Mutual information top features

Table 1. Model result all feature

Model	Accuracy	Precision	Recall	F1-score	Time (s)
XGBoost	96.85%	96.86%	96.85%	96.62%	47.29
LightGBM	97.63%	97.67%	97.63%	97.55%	39.79
Catboost	96.35%	96.48%	96.35%	95.48%	268.40

Table 2. Model result with feature selection

Model	Accuracy	Precision	Recall	F1-score	Time (s)
XGBoost	99.85%	99.85%	99.85%	99.85%	26.12
LightGBM	99.88%	99.88%	99.85%	99.88%	19.18
Catboost	99.83%	99.83%	99.83%	99.83%	139.82

Table 2 illustrates the results after applying feature selection. Overall, a significant improvement is observed for all models in terms of classification performance. Accuracy, precision, recall, and F1-score values increase to nearly 99.9% across the three algorithms, indicating enhanced discriminative capability. In addition, feature selection leads to a notable reduction in computational time for all models, with LightGBM remaining the most efficient in terms of execution speed. These results indicate that feature selection not only improves predictive performance but also enhances computational efficiency, making the models more suitable for large-scale DDoS detection scenarios.

Figure 5 compares the confusion matrices of the LightGBM model before and after feature selection. The results show that the model after feature selection provides better class separation than the model before feature selection. In particular, the number of false positive

predictions is lower, which indicates improved classification performance. This suggests that the selected features help the model focus on the most relevant information in the dataset. Although there are minor differences in the visual appearance of the figure, these are only related to formatting. They do not affect the numerical evaluation results reported in Tables 1 and 2.

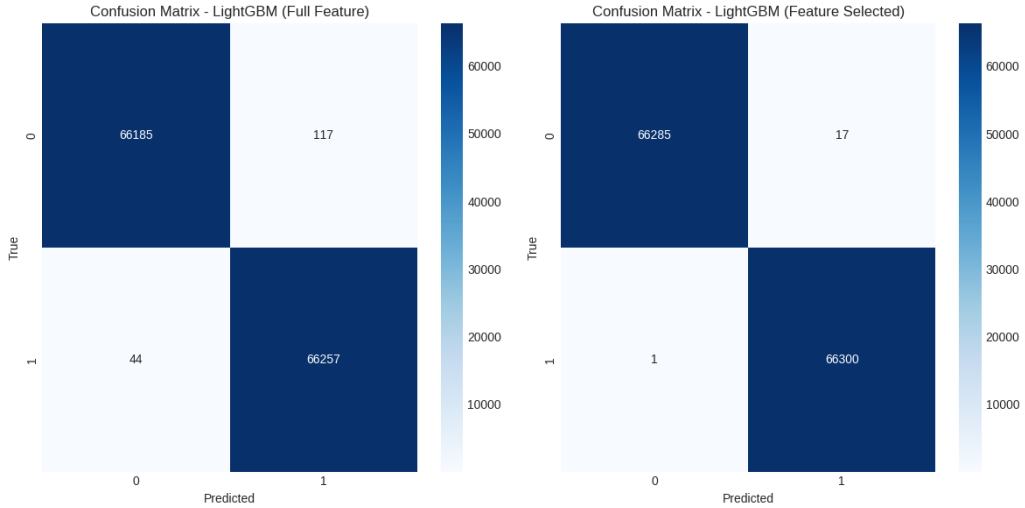

Figure 5. Best model confusion matrix

Discussion

The empirical results indicate that MI-based feature selection produces consistent improvements in both predictive performance and computational efficiency across all evaluated boosting models. As reported in Table 2, accuracy, precision, recall, and F1-score increase to nearly 99.9% following feature selection, accompanied by a substantial reduction in training time. This outcome aligns with recent findings that emphasize the importance of feature quality in high-dimensional intrusion detection tasks, where redundant attributes tend to degrade model stability and increase computational cost (Alduailij et al., 2022; Hossain et al., 2024). The strong multicollinearity observed in the dataset (Figure 2) further supports the necessity of feature filtering to stabilize the learning process and improve model efficiency.

The observed improvements can be theoretically explained through the interaction between feature relevance and the iterative learning mechanism of boosting algorithms. Mutual Information (MI) quantifies the dependency between features and target variables, enabling the selection of highly informative attributes while eliminating noisy or weak predictors (Mallidi & Ramisetty, 2025; Shah et al., 2026). In gradient boosting, models are trained sequentially to minimize residual errors, and the presence of redundant features may lead to unstable split selection and inefficient error propagation. The reduction of irrelevant features constrains the hypothesis space and improves the consistency of decision boundaries across iterations. Similar observations have been reported in recent studies, where feature selection contributes not only to accuracy improvement but also to more stable convergence behavior in ensemble models (Ma et al., 2023; Zhang et al., 2025).

The performance differences among models indicate that LightGBM achieves the most optimal balance between accuracy and computational efficiency. This behavior is consistent with prior studies highlighting the efficiency of LightGBM's leaf-wise tree growth strategy and histogram-based optimization in handling large-scale and high-dimensional datasets (Ma et al., 2023; Alsaffar et al., 2024). With a reduced feature space, LightGBM can more effectively prioritize informative splits, resulting in faster convergence and lower computational cost. XGBoost demonstrates comparable predictive performance, although with higher training time, which is in line with findings that its level-wise growth strategy favors stability but may

limit computational efficiency (Ali et al., 2023). CatBoost achieves competitive classification performance; however, its training time remains substantially higher, which has also been observed in prior studies due to additional processing mechanisms designed to improve robustness and handle categorical features (Hirsi et al., 2025).

These findings extend previous research by demonstrating that feature selection influences not only predictive performance but also the internal learning dynamics of boosting algorithms. Earlier studies primarily reported performance improvements resulting from feature selection techniques (Alduailij et al., 2022; Yuan & Yang, 2025), with limited emphasis on how feature quality affects model behavior during training. The present results provide empirical evidence that feature selection acts as a structural component that shapes model efficiency, stability, and learning capacity. This insight is particularly relevant for real-world deployment scenarios, where detection systems must operate under constraints related to latency, scalability, and resource utilization (Han et al., 2024).

A number of limitations must be acknowledged when interpreting these findings. The high classification performance observed may be influenced by the characteristics of the CIC-DDoS2019 dataset, which may exhibit relatively clear class separability compared to real-world traffic conditions. Similar concerns have been raised in recent studies regarding the generalizability of benchmark datasets in cybersecurity research (Tymoshchuk et al., 2024). In addition, MI was applied as a univariate filter-based method, meaning that interactions among features were not explicitly modeled. This limitation has been noted in prior work, where feature interaction plays a significant role in capturing complex attack patterns (Alsaffar et al., 2024). The evaluation was also conducted in an offline setting, without considering inference latency, memory usage, or robustness under dynamic and adversarial environments.

Future research directions include extending the proposed framework to multiple datasets with varying traffic characteristics, integrating feature selection methods capable of capturing multivariate interactions, and evaluating performance under real-time and streaming conditions. Further investigation into deployment-oriented metrics, including inference efficiency and system-level scalability, is also essential to enhance the practical applicability of ML-based DDoS detection systems.

CONCLUSION

This study demonstrates that the performance of boosting-based models in high-dimensional DDoS detection is strongly influenced by feature quality. The application of MI based feature selection significantly improves predictive performance while reducing computational cost, indicating its effectiveness in filtering redundant and non-informative features. Among the evaluated models, LightGBM achieves the most optimal balance between accuracy and efficiency, highlighting its suitability for large-scale and resource-constrained environments. The findings further reveal that feature selection plays a structural role in shaping the learning behavior of boosting algorithms by enhancing stability and convergence during training. These results provide empirical evidence that integrating feature selection into the modeling pipeline is essential for achieving scalable and efficient DDoS detection systems. Future work should extend this approach to diverse datasets and explore advanced feature selection methods that capture feature interactions under real-world deployment conditions.

REFERENCES

Adedeji, K. B., Abu-Mahfouz, A. M., & Kurien, A. M. (2023). DDoS Attack and Detection Methods in Internet-Enabled Networks: Concept, Research Perspectives, and Challenges. *Journal of Sensor and Actuator Networks*, *12*(4), 51. https://doi.org/10.3390/jsan12040051

Alduailij, M., Khan, Q. W., Tahir, M., Sardaraz, M., Alduailij, M., & Malik, F. (2022).

Machine-Learning-Based DDoS Attack Detection Using Mutual Information and Random Forest Feature Importance Method. *Symmetry*, *14*(6), 1095. https://doi.org/10.3390/sym14061095

Ali, T. E., Chong, Y.-W., & Manickam, S. (2023). Comparison of ML/DL Approaches for Detecting DDoS Attacks in SDN. *Applied Sciences*, *13*(5), 3033. https://doi.org/10.3390/app13053033

Alsaffar, A. M., Nouri-Baygi, M., & Zolbanin, H. M. (2024). Shielding networks: Enhancing intrusion detection with hybrid feature selection and stack ensemble learning. *Journal of Big Data*, *11*(1), 133. https://doi.org/10.1186/s40537-024-00994-7

Al-Shareeda, M. A., Manickam, S., & Saare, M. A. (2023). DDoS attacks detection using machine learning and deep learning techniques: analysis and comparison. *Bulletin of Electrical Engineering and Informatics*, *12*(2), 930-939. https://doi.org/10.11591/eei.v12i2.4466

Falowo, O. I., Ozer, M., Li, C., & Abdo, J. B. (2024). Evolving malware and DDoS attacks: Decadal longitudinal study. *IEEE Access*, *12*, 39221-39237. https://doi.org/10.1109/ACCESS.2024.3376682

Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An overview on the advancements of support vector machine models in healthcare applications: a review. *Information*, *15*(4), 235. https://doi.org/10.3390/info15040235

Han, D., Li, H., Fu, X., & Zhou, S. (2024). Traffic Feature Selection and Distributed Denial of Service Attack Detection in Software-Defined Networks Based on Machine Learning. *Sensors*, *24*(13), 4344. https://doi.org/10.3390/s24134344

Hasan, M. K., Habib, A. A., Islam, S., Safie, N., Abdullah, S. N. H. S., & Pandey, B. (2023). DDoS: Distributed denial of service attack in communication standard vulnerabilities in smart grid applications and cyber security with recent developments. *Energy Reports*, *9*, 1318-1326. https://doi.org/10.1016/j.egyr.2023.05.184

Hayat, R. F., Aurangzeb, S., Aleem, M., Srivastava, G., & Lin, J. C. W. (2022). ML-DDoS: A blockchain-based multilevel DDoS mitigation mechanism for IoT environments. *IEEE Transactions on Engineering Management*, *71*, 12605-12618. https://doi.org/10.1109/TEM.2022.3170519

Hirsi, A., Alhartomi, M. A., Audah, L., Salh, A., Sahar, N. M., Ahmed, S., ... & Farah, A. (2025). Comprehensive analysis of ddos anomaly detection in software-defined networks. *IEEE Access*, *13*, 23013-23071. https://doi.org/10.1109/ACCESS.2025.3535943

Hossain, M. A., & Islam, M. S. (2024). Enhancing DDoS attack detection with hybrid feature selection and ensemble-based classifier: A promising solution for robust cybersecurity. *Measurement: Sensors*, *32*, 101037. https://doi.org/10.1016/j.measen.2024.101037

Issa, A. A., & Albayrak, Z. (2023). DDoS attack intrusion detection system based on hybridization of CNN and LSTM. *Acta Polytechnica Hungarica*, *20*(2), 105-123. https://doi.org/10.12700/APH.20.2.2023.2.6

Jaafar, A. G., Suhaimi, N. H. S., Ghali, A. A., Mansor, H., Samy, G. N., Kama, N., & Hassan, N. H. (2025). A Review of Detection Challenge for Signature and Anomaly-Based Detection in Detecting HTTP DDoS Attacks. *Open International Journal of Informatics*, *13*(2), 1-18. https://doi.org/10.11113/oiji2025.13n2.345

Li, C., Liu, J., Lu, B., & Luo, Y. (2021). Cost-aware automatic scaling and workload-aware replica management for edge-cloud environment. *Journal of Network and Computer Applications*, *180*, 103017. https://doi.org/10.1016/j.jnca.2021.103017

Ma, R., Chen, X., & Zhai, R. (2023). A DDoS Attack Detection Method Based on Natural Selection of Features and Models. *Electronics*, *12*(4), 1059. https://doi.org/10.3390/electronics12041059

Ma, R., Wang, Q., Bu, X., & Chen, X. (2023). Real-Time Detection of DDoS Attacks Based on Random Forest in SDN. *Applied Sciences*, *13*(13), 7872. https://doi.org/10.3390/app13137872

Mallidi, S. K. R., & Ramisetty, R. R. (2025). Embedded-filter ACO using clustering based mutual information for feature selection. *Journal of Combinatorial Optimization*, *49*(2), 27. https://doi.org/10.1007/s10878-025-01259-6

Martinović, M., Dokic, K., & Pudić, D. (2025). Comparative analysis of machine learning models for predicting innovation outcomes: an applied AI approach. *Applied Sciences*, *15*(7), 3636. https://doi.org/10.3390/app15073636

Poonia, L., & Tinker, S. (2025). A comprehensive analysis of the types, impacts, prevention, and mitigation of ddos attacks. *Recent Patents on Engineering*, *19*(9), E18722121322166. https://doi.org/10.2174/0118722121322166240828112546

Shah, S. N. A., Issar, K., & Parveen, R. (2026). A hybrid feature extraction framework combining PCA and mutual information for gene expression based lung cancer classification. *PloS one*, *21*(2), e0342160. https://doi.org/10.1371/journal.pone.0342160

Talukder, M. A., & Uddin, M. A. (2023). Cic-ddos2019 dataset. *Mendeley Data*, *1*. Mendeley. https://doi.org/10.17632/SSNC74XM6R.1

Tymoshchuk, D., Yasniy, O., Mytnyk, M., Zagorodna, N., & Tymoshchuk, V. (2024). Detection and classification of DDoS flooding attacks by machine learning method. *arXiv preprint arXiv:2412.18990*. https://doi.org/10.48550/ARXIV.2412.18990

Yuan, M., & Yang, K. (2025). Deep Learning Network Intrusion Detection Based on MI-XGBoost Feature Selection. *Journal of Cyber Security*, *7*(1), 197–219. https://doi.org/10.32604/jcs.2025.066089

Zhang, R., Li, Z., & Gao, F. (2025). Modeling and diagnosis of industrial system using hybrid deep residual shrinkage network and XGBoost. *Process Safety and Environmental Protection*, *200*, 107438. https://doi.org/10.1016/j.psep.2025.107438