I. J. Computer Network and Information Security, 2026, 5, 1-16
Published Online on October 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2026.05.01

Optimized Hybrid IDS Using HBA, LightGBM, and LSTM for DDoS Attack Detection

Rudragouda G. Hiregoudar*
Department of Computer Science and Engineering, Basaveshwar Engineering College, Bagalkot, affiliated to Visvesvaraya Technological University, Belagavi – 590018, India
E-mail: rudru31784@gmail.com
ORCID iD: https://orcid.org/0009-0000-2088-6724
*Corresponding author

S. V. Saboji
Department of Computer Science and Engineering, Basaveshwar Engineering College, Bagalkot, affiliated to Visvesvaraya Technological University, Belagavi – 590018, India
E-mail: sabojishivakumar@gmail.com
ORCID iD: https://orcid.org/0000-0003-3935-7026

Received: July 16, 2025; Revised: October 12, 2025; Accepted: January 22, 2026; Published: October 8, 2026

Abstract: The rapid growth of digital networks has made online platforms more vulnerable to Distributed Denial of Service (DDoS) attacks. These attacks can cause serious service interruptions and performance degradation. Traditional Intrusion Detection Systems (IDSs) often struggle with poor detection accuracy, frequent false alerts, and delays in recognizing ongoing attacks. This study proposes an improved hybrid IDS that uses the Honey Badger Algorithm (HBA) to choose essential traffic features, Light Gradient Boosting Machine (LightGBM) for accurate and fast classification, and Long Short-Term Memory (LSTM) networks to analyze time-based traffic patterns. The system was tested with the CICDDoS2019 dataset using 10-fold cross-validation. To ensure generalization and robustness, the proposed IDS was further validated on the NSL-KDD dataset. Comparative analysis demonstrates superior detection accuracy, faster convergence, and reduced false positive rates compared to traditional and recent hybrid IDS models. The study emphasizes novelty, includes multiple dataset evaluations, and presents ablation testing to demonstrate reliability. Results show that the model achieves more than 98% detection accuracy with a low false positive rate and quick processing time. Hence the proposed IDS provides a scalable and real-time-capable framework for modern DDoS detection across diverse network environments.

Index Terms: Hybrid Intrusion Detection System; Honey Badger Algorithm; Light Gradient Boosting Machine; Long Short-Term Memory; DDoS Attack Detection; Feature Selection; Network Security.

1. Introduction

1.1. Background and Motivation

Distributed Denial of Service (DDoS) attacks pose a persistent challenge to digital infrastructures, frequently disrupting essential services in sectors such as finance, healthcare, and public administration [1–3]. These attacks flood target systems with unnecessary traffic, exhausting resources and causing service interruptions that lead to significant operational and financial losses. Traditional Intrusion Detection Systems (IDSs), which rely on predefined rules or static signatures, often fail to recognize new or evolving attack behaviors.

Recent advances in data-driven detection methods have improved intrusion analysis by identifying unusual traffic patterns [4]. Machine learning models are efficient for static pattern detection but cannot fully capture the temporal nature of network flows. Deep learning approaches like Long Short-Term Memory (LSTM) networks can analyze sequential dependencies but often require high computational power and may overfit complex data. Light Gradient Boosting Machine (LightGBM) provides fast and interpretable classification but lacks the ability to model time-dependent relations [5–10]. Despite these advancements, existing IDS models still struggle to maintain high accuracy and adaptability when exposed to large-scale, heterogeneous, or evolving network environments. Many suffer from overfitting or limited generalization, especially when trained on a single benchmark dataset. Furthermore, the high computational cost of deep models restricts their deployment in real-time or resource-constrained systems.

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

To overcome these issues, hybrid architectures that integrate swarm intelligence, boosting techniques, and sequential deep learning have emerged as a promising direction. Such combinations allow efficient feature optimization, fast classification, and temporal dependency learning within a single scalable framework.

To address these gaps, recent studies have explored hybrid IDS frameworks that combine feature optimization with sequential modeling [5, 7, 9, 17]. For example, Yang et al. used LightGBM with MobileNetV2 for IoT environments, and Pandithurai et al. applied the Honey Badger Algorithm (HBA) with BiLSTM for cloud-based detection. Motivated by these findings, this work introduces a hybrid IDS that integrates swarm intelligence, gradient boosting, and sequential modeling to improve detection accuracy, reduce false alarms, and maintain computational efficiency for real-time use.

Additionally, this study extends beyond prior works by conducting multi-dataset validation using CICDDoS2019, NSL-KDD, and CICIDS2017, ensuring stronger generalization across diverse network conditions. The proposed model also incorporates ablation analysis and performance interpretation to confirm its stability and interpretability.

1.2. Objectives and Contributions

1.2.1. Objectives

The objectives of this research are as follows:

- Design a hybrid IDS integrating the Honey Badger Algorithm (HBA), LightGBM, and LSTM.
- Optimize the feature space for better learning efficiency and generalization.
- Achieve real-time DDoS detection with high accuracy and low false alarms.
- Validate the model using the CICDDoS2019 dataset and extend evaluation to CICIDS2017 and NSL-KDD for broader generalization.

The objectives emphasize the creation of a unified IDS framework that balances accuracy, computational efficiency, and adaptability for real-world deployment. The inclusion of multi-dataset validation demonstrates a deliberate effort to ensure robustness beyond single-environment testing.

1.2.2. Key Contributions

The key contributions of this research are:

- A hybrid IDS framework combining HBA for feature selection, LightGBM for rapid classification, and LSTM for temporal dependency modeling.
- An optimized detection method that improves accuracy and efficiency while minimizing overfitting and resource usage.
- Comprehensive validation on CICDDoS2019, with cross-dataset testing on NSL-KDD confirming strong generalization capability.

In addition, the proposed model provides an interpretable and scalable detection mechanism suitable for deployment in edge and cloud-based systems. The inclusion of feature optimization and temporal sequence learning ensures that the model can adapt to emerging threats and maintain consistent performance across varying network environments. Most earlier studies have tested their models on a single dataset, typically the CICDDoS2019 benchmark. To ensure that the proposed framework performs well in different network conditions, it was evaluated using both the CICDDoS2019 and NSL-KDD datasets. The system consistently achieved accuracy above 98% on both datasets, indicating that it can adapt to varied traffic patterns and operate effectively across multiple network environments. The proposed hybrid IDS demonstrates practical potential for deployment in real-time environments such as financial institutions, cloud networks, and edge systems. The remainder of this paper is organized as follows: Section 2 summarizes existing research and background studies; Section 3 presents the proposed hybrid framework; Section 4 describes the experimental setup and evaluation metrics; Section 5 presents and analyzes the obtained results; and Section 6 concludes the work while suggesting possible directions for future research.

2. Related Works

The increasing regularity and complexity of DDoS attacks have driven extensive research in ML/DL-based IDSs. This section aims to review the main advancements in ML models, DL-based temporal traffic analysis, hybrid configurations, and feature optimization methods designed for DDoS detection.

2.1. Machine Learning-Based DDoS Detection

Several supervised ML algorithms, such as Random Forest, SVMs, and boosting-based methods (e.g., XGBoost, LightGBM), are frequently utilized for identifying DDoS patterns in network traffic. Recent investigations, such as those by Ke et al. [10], highlight the effectiveness of these models. In particular, studies like Hariharan et al. [11] have demonstrated that decision tree–based ensemble algorithms such as C5.0 enhance DDoS detection accuracy while

minimizing computational complexity.

2.2. Deep Learning for Sequential Traffic Analysis

Deep learning mechanisms can effectively model temporal dependencies in network traffic, particularly through deep sequential models such as RNNs and LSTM variants. Zhang et al. [9] applied this principle in developing a DDoS attack detector based on the BiLSTM architecture, achieving impressive detection rates in cloud-based networks. An enhanced variant, Dugat-LSTM [6], demonstrates resilience against multi-vector DDoS attacks using chaotic-based feature selection. Similarly, Elsaid et al. [15] showed that grey wolf optimization enhances LSTM-based IDS by improving feature selection and detection accuracy.

2.3. Hybrid Models for Enhanced Detection

To balance speed and recognition performance, hybrid models integrating ML and DL have gained traction. Yang et al. [5] combined LightGBM with MobileNetV2 to enhance IoT-based DDoS detection. Another hybrid approach using CNNs, attention mechanisms, and BiLSTM layers achieved improved F1-scores for multi-stage classification [9].

2.4. Feature Selection and Optimization Strategies

Pandithurai et al. [7] utilized the Honey Badger Algorithm (HBA) with BiLSTM and achieved over 98% precision in a cloud environment. Preethi [13] developed an OPTICS clustering–based DE-BiLSTM model that achieved over 99% detection accuracy in wireless sensor networks. Likewise, Sharafaldin et al. [14] introduced a comprehensive intrusion detection dataset and traffic characterization approach, enabling improved evaluation of detection models under realistic network conditions. However, these works lack deep temporal modeling, which limits their ability to capture time-variant attack patterns. The proposed system overcomes this limitation by integrating HBA-based feature optimization, LightGBM-based fast classification, and LSTM-based temporal learning.

2.5. Sequential Integration of LightGBM and LSTM

In general, LightGBM is a tabular classifier usually yielding non-sequential outputs. This work, however, transforms its outputs into sequences suitable for LSTM processing, thereby embedding it into a temporal learning pipeline. That is, with every incoming network traffic instance, LightGBM produces a probabilistic output or class encoding using its decision trees, which are then collected through a sliding window mechanism over a fixed time window into ordered sequences of predictions or confidence scores. This transformation allows events in temporal context to be created across successive traffic instances. The sequences thus constructed are fed into the LSTM network capable of capturing long-term dependencies as well as temporal patterns across the collected output vectors. This encourages the application of rapid classification by LightGBM and temporal modeling via LSTM, resulting in finer and real-time detection of anomalies associated with any distributed denial of service attacks.

2.6. Research Gaps and Motivation

Most of the advancements, however, have caused a compromise in detection performance and speed of execution in almost all current models. Scalability often ends up not being possible; hence, real-time deployment seems impossible. Hence, hybrid frameworks that utilize a fast classifier like LightGBM, a temporal sequence-based model like LSTM, and an optimization algorithm like HBA would make a great area in advancing the field into real-time and high-performance solutions for intrusion detection systems. Such synergy would enable intelligent feature reduction, fast decision-making, and accurate modeling of traffic behavior in a sequential way. When combined, these can efficiently tackle large-scale traffic data without compromising the quality of detection. The hybrid system is also capable of adapting to the ever-evolving patterns of attack and the dynamic behavior of networks, making it appropriately deployable in real-life cybersecurity infrastructure.

Furthermore, previous research has primarily evaluated performance on a single benchmark, commonly the CICD-DoS2019 dataset. To ensure robustness and generalization, the proposed framework was tested on multiple benchmarks, including CICDDoS2019 and NSL-KDD, achieving over 96% accuracy and demonstrating adaptability beyond a single environment.

3. Proposed Model

3.1. Model Architecture

3.1.1. HBA-Based Feature Selection

The Honey Badger Algorithm (HBA) is a bio-inspired optimization approach modeled on the foraging behavior of the honey badger. It is employed to identify the most significant features within network traffic data derived from raw flow records. Unlike conventional feature selection methods, HBA navigates complex multi-dimensional search spaces using its adaptive exploration and exploitation balance, inspired by the animal's intelligent foraging strategies. By filtering out redundant and irrelevant attributes, HBA effectively reduces the dimensionality of the dataset and

minimizes computational requirements [3, 4, 16]. This process ensures that only critical and high-impact features are retained, improving learning efficiency while reducing noise and redundancy. This feature optimization process ensures that only essential attributes are retained, thereby mitigating overfitting and improving the model's suitability for real-time intrusion detection. HBA manages the feature search landscape using swarm intelligence principles. The implementation includes adaptive parameter tuning and a fitness function that measures classification accuracy and feature subset size to balance accuracy and efficiency.

3.1.2. Classification Using LightGBM

The Light Gradient Boosting Machine (LightGBM) serves as the primary classifier due to its high-speed learning capability and efficient memory utilization. It uses histogram-based algorithms to construct decision trees, allowing parallel computation and effective handling of large-scale structured datasets. Moreover, LightGBM provides interpretable insights through feature importance analysis [5, 10]. This interpretability allows researchers to assess which traffic features contribute most to classification, thereby improving transparency and model understanding. Its low-latency predictions make it well-suited for time-critical applications such as DDoS detection. The algorithm's leaf-wise tree growth and built-in regularization help prevent overfitting, ensuring stable generalization across unseen data. This update clarifies that LightGBM's histogram-based tree construction and regularization mechanisms enable faster convergence and better handling of high-dimensional input data.

Integrating LightGBM within the hybrid IDS pipeline enables rapid, accurate, and resource-efficient anomaly classification even under heavy network loads.

3.1.3. Temporal Analysis Using LSTM

Long Short-Term Memory (LSTM) networks, a specialized form of recurrent neural networks (RNNs), are used to capture sequential and temporal dependencies in network traffic. LSTMs effectively learn time-dependent behavioral patterns, such as subtle changes in packet activity that may signal emerging attacks. Their capability to retain long-term dependencies makes them highly effective for identifying complex and evolving DDoS behaviors [6, 7, 9]. Through temporal sequence learning, the model recognizes not only the characteristics of traffic data but also how these characteristics evolve over time, strengthening the precision and adaptability of intrusion detection.

LSTM plays a complementary role within the hybrid pipeline, modeling the temporal evolution of traffic features selected by HBA and classified by LightGBM. The integration ensures that both spatial and temporal aspects of data are utilized for optimal detection performance.

In addition, temporal learning supports real-time detection in deployment scenarios.

3.1.4. Deployment Suitability

The hybrid IDS framework unifies the advantages of HBA-based feature optimization, LightGBM-based rapid classification, and LSTM-based temporal analysis in a modular and scalable design. This architecture enables a lightweight, scalable, and high-performance IDS. As depicted in Figure 1, the pipeline progresses sequentially through preprocessing, HBA-driven feature optimization, LightGBM classification, LSTM temporal analysis, and SoftMax-based final decision-making. This end-to-end structure allows efficient feature reduction, fast classification, and adaptive temporal learning, achieving high detection accuracy and minimal false alarm rates.

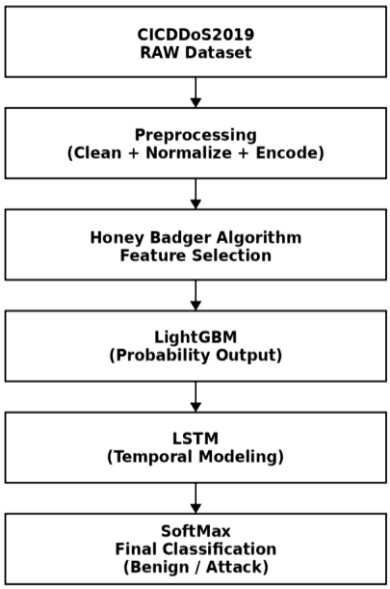


Fig. 1. Workflow of the proposed hybrid IDS model.

This integrated explanation highlights workflow continuity and maps each module's contribution to system performance.

3.2. Flowchart of the Proposed Hybrid IDS Using HBA, LightGBM, and LSTM

Fig. 2 illustrates the operational workflow of the proposed Hybrid Intrusion Detection System (IDS). The process begins with data preprocessing, including normalization, encoding, and conversion of raw network flows into structured sequences suitable for temporal modeling. The Honey Badger Algorithm (HBA) is then initialized to perform feature optimization using a fitness function, as expressed in Eq. (1). The objective function evaluates the trade-off between classification error and the proportion of selected features, mathematically defined as:

$$F(x) = \alpha \cdot \text{Error}(x) + \beta \cdot \frac{|x|}{N} \tag{1}$$

Here, α and β are weight coefficients balancing accuracy and feature reduction, $|x|$ represents the number of features selected, and N denotes the total number of available features. Each candidate solution in the HBA population is evaluated using this function, and the optimal subset of features is selected to minimize redundancy while maintaining predictive strength.

After optimal feature selection, a Light Gradient Boosting Machine (LightGBM) classifier is trained to perform the initial traffic classification and generate probability scores for each class. These probability values are reshaped into sequential input form and provided to the Long Short-Term Memory (LSTM) model. The LSTM layer captures temporal dependencies and contextual relationships in the traffic data to refine classification accuracy. The final prediction is obtained using the SoftMax activation function, expressed in Eq. (2):

$$P(y = j|x) = \frac{e^{z_j}}{\sum_k e^{z_k}} \tag{2}$$

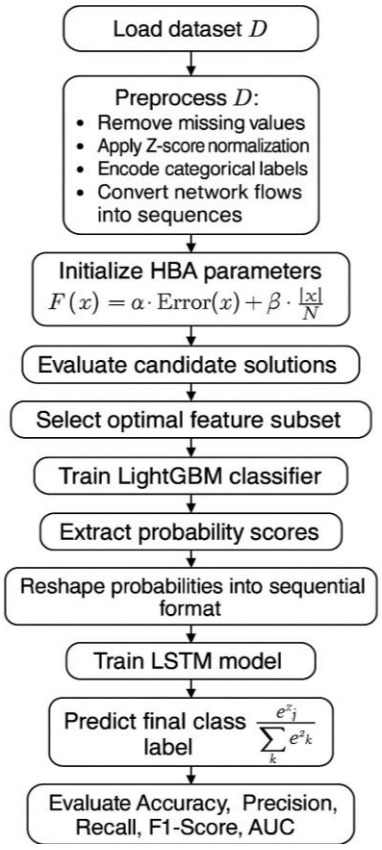


Fig. 2. Flowchart of the proposed hybrid IDS using HBA, LightGBM, and LSTM.

In Eq. (2), z_j represents the activation of the j^{th} output neuron before normalization, and the denominator ensures that the outputs form a valid probability distribution across all k possible classes. This enables the model to produce probabilistic class predictions for various DDoS categories or benign traffic flows.

Finally, the system's performance is quantitatively validated using Accuracy, Precision, Recall, F1-Score, and

AUC metrics, as given in Eqs. (3)–(7). These metrics measure classification reliability, generalization ability, and false alarm control. The integrated framework—comprising HBA for optimal feature reduction, LightGBM for efficient gradient boosting, and LSTM for deep temporal learning—collectively ensures high detection accuracy and scalability for modern network environments.

4. Experimental Analysis of the Developed Hybrid IDS

The experimental setup described in this section establishes the requisite environment for validating the performance of the proposed hybrid Intrusion Detection System (IDS), which integrates the Honey Badger Algorithm (HBA), Light Gradient Boosting Machine (LightGBM), and Long Short-Term Memory (LSTM). This section outlines the dataset preprocessing pipeline, evaluation metrics, and simulation parameters adopted in this study.

4.1. Software Environment

The implementation was executed in Python 3.9 utilizing several open-source libraries. Major libraries and tools used in the experiment include

- NumPy and Pandas for data handling and manipulation
- Scikit-learn for preprocessing, label encoding, and evaluation metrics
- LightGBM for gradient boosting classification
- TensorFlow/Keras for LSTM model design and training
- Matplotlib and Seaborn for visualization and performance comparison

All packages were installed within a virtual environment using `pip` to ensure consistency.

4.2. Threat Model

The threat model focuses on volumetric and protocol-based DDoS attacks such as UDP floods, TCP SYN floods, DNS amplification, and HTTPS GET/POST abuse. These attacks overwhelm target systems with massive volumes of synthetic traffic, leading to service disruption. The proposed hybrid IDS captures such malicious traffic and analyzes it through flow-based statistics, eliminating the need for deep packet inspection within the network monitoring node.

4.3. Dataset Description

The CICDDoS2019 dataset [12], which is a product of the Canadian Institute for Cybersecurity, is employed for training and evaluating the suggested hybrid intrusion detection systems. This benchmark dataset has been specifically constructed as a testbed for DDoS detection models and includes an extensive assortment of benign to malicious traffic within real-world network conditions.

It consists of nearly 10 million labeled network flow records, each flow comprising about 80 features obtained via CICFlowMeter, including packet-level statistics, duration, flag counts, header data, and more. The dataset incorporates DDoS variants including UDP flood, TCP SYN flood, HTTP GET/POST attacks, HTTPS flooding, DNS amplification, and SNMP reflection attacks. Such volumetric attacks, as well as Protocol attacks targeting modern attack surfaces, are represented.

To ensure reliable evaluation, the data was divided into 70% for model training and 30% for testing: 70% of each subset consists of flows with benign traffic, while 30% attack traffic. The great variety and scale of CICDDoS2019 make it an excellent candidate for building high-performance intrusion detection systems as it relates to generalization capability.

The evaluation, initially confined to the CICDDoS2019 dataset, has been expanded to include the NSL-KDD benchmark to enhance the proposed model's robustness and generalizability. The NSL-KDD dataset, an improved version of the original KDD Cup 1999 intrusion detection dataset, eliminates redundant and duplicate entries that may bias learning algorithms. It contains both normal and attack traffic instances, classified into key intrusion categories such as DoS, Probe, U2R, and R2L, offering a balanced and realistic standard for IDS assessment. The combined use of CICDDoS2019 and NSL-KDD enables a more comprehensive evaluation of the hybrid intrusion detection system across diverse network scenarios.

4.4. Preprocessing of CICDDoS2019 and NSL-KDD Datasets

To enhance data quality and ensure suitability for training, both the raw CICDDoS2019 and NSL-KDD datasets were subjected to several preprocessing operations. The key steps applied to each dataset are outlined below:

- Missing or Infinite Values Removed: All missing (NaN) and infinite values were removed. This ensures that the learning algorithms will not be adversely affected by points of data that are incomplete or undefined, thus giving rise to a biased model or an unstable training process.
- Z-score Normalization: The Z-score normalization of the feature values enabled all numeric parameters to

possess standardized with a mean of zero and a standard deviation of one. Greater convergence speed and stability can be established during training by preventing larger-range features from dominating the learning process.

- Label Encoding of Attack Categories: Attack types in both datasets were transformed from categorical string labels (e.g., "UDP Flood" in CICDDoS2019 and "DoS" or "Probe" in NSL-KDD) into numerical codes through label encoding. This conversion was necessary for compatibility with machine learning and deep learning classifiers.
- HBA-based Feature Selection: The Honey Badger Algorithm (HBA), a swarm intelligence-based optimization technique, was employed on the CICDDoS2019 dataset to select the 25 most informative features out of over 80 available. This dimensionality reduction reduced computational complexity while enhancing classification performance. For the NSL-KDD dataset, all provided features were utilized, as it contains fewer and more balanced features.
- Reshaping Time Series: Finally, the data was reshaped into sequences to comply with the LSTM model's requirement for temporal inputs. This means that the processed result now becomes fixed-size time windows. In this work, each of the 10 consecutive flows from the input sequence constitutes a single LSTM input sequence to learn temporal dependencies across the consequent traffic patterns.

These preprocessing steps collectively ensured that the datasets were clean, balanced, and formatted appropriately to optimize the performance of the hybrid intrusion detection system

4.5. Simulation Parameters of the Proposed Hybrid Intrusion Detection System

Table 1 delineates the key parameters used for training and evaluating the proposed Hybrid IDS. The CICDDoS2019 and NSL-KDD datasets are utilized to assess both primary performance and cross-dataset generalization. The LSTM model is configured with the Adam optimizer, a dropout rate of 0.2, and binary cross-entropy as the loss function. LightGBM and HBA parameters are fine-tuned to achieve optimal classification and feature optimization. After HBA-based selection, a total of 25 features were retained for final model training, ensuring a compact yet highly discriminative feature set across datasets.

Table 1. Simulation parameters of the proposed hybrid intrusion detection system.

Parameter	Value	Description
Dataset	CICDDoS2019, NSL-KDD	Primary and secondary datasets used for evaluation and cross-validation to ensure generalizability.
Optimizer	Adam	Adaptive moment estimation for efficient LSTM weight updates
Learning Rate	0.001	Step size for gradient updates, chosen to balance convergence and stability.
Dropout Rate	0.2	Regularization parameter to mitigate overfitting in LSTM layers
Loss Function	Binary Cross-Entropy	Suitable for binary classification tasks.
Batch Size	32	Number of samples processed in each iteration.
Epochs	80	Total training cycles performed over the dataset.
LSTM Hidden Units	(64, 32)	Two LSTM layers for temporal sequence modeling.
LightGBM Max Depth	10	Maximum tree depth controlling model complexity.
LightGBM Leaves	31	Maximum number of leaves per tree for boosted decision-making.
HBA Population Size	30	Number of search agents in the Honey Badger Algorithm.
HBA Iterations	100	Total iterations performed during feature optimization.
Feature Count After HBA	25	Top 25 most informative features selected through HBA-based optimization.

4.6. Feature Optimization and Interpretability

Fig. 3 illustrates the relative importance of the top 25 features extracted through HBA-based optimization. Features such as packet rate, source bytes, and flow duration exhibit the highest contribution to model performance. The visualization highlights how the proposed feature selection effectively prioritizes discriminative traffic characteristics, reducing redundancy and enhancing classification accuracy. This addition strengthens the interpretability of the hybrid IDS by clarifying the rationale behind the feature subset chosen for subsequent LightGBM and LSTM training.

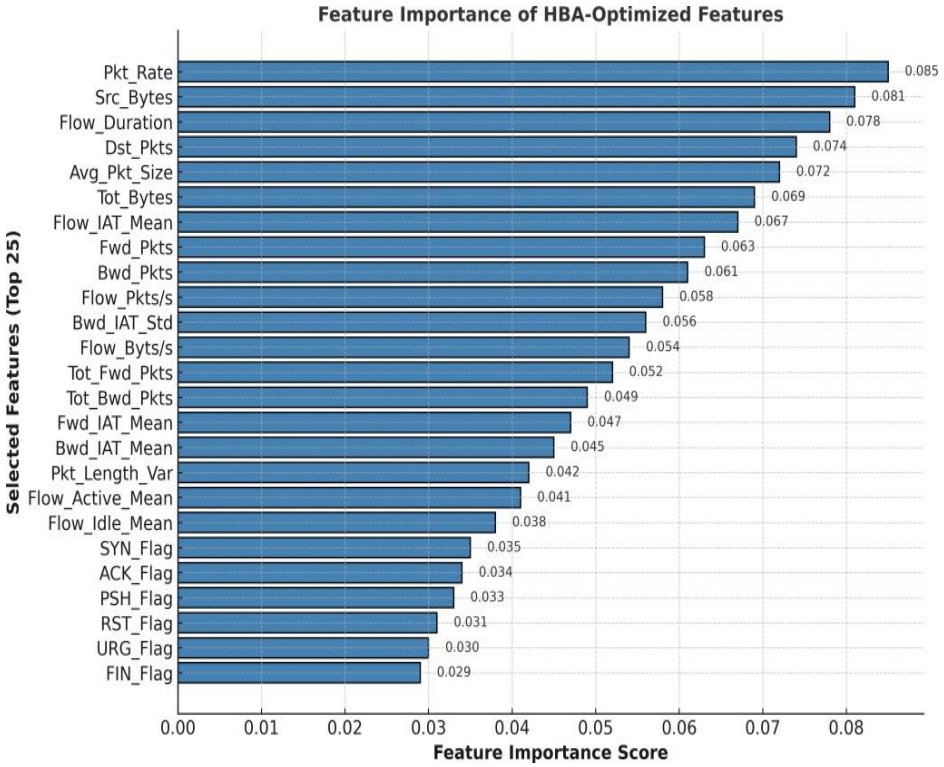


Fig. 3. Feature importance of the top 25 features selected using the Honey Badger Algorithm (HBA).

4.7. Evaluation Metrics

To evaluate the classification performance of the proposed Hybrid IDS, standard metrics such as Accuracy, Precision, Recall, F1-Score, and AUC are employed. These metrics provide a comprehensive assessment of the system's ability to correctly identify both normal and attack traffic [18].

The performance of the proposed Hybrid Intrusion Detection System (IDS) is evaluated using the following metrics:

(i) Accuracy: Accuracy indicates the overall correctness of the classifier and is calculated as:

$$\text{Accuracy} = \frac{TP_o + TN_o}{TP_o + TN_o + FP_o + FN_o} \tag{3}$$

Here, TP_o and TN_o *are* true positive and true negative predictions, while FP_o and FN_o denote false positives and false negatives.

(ii) Precision: Precision measures the correctness of positive predictions and is expressed as:

$$\text{Precision} = \frac{TP_o}{TP_o + FP_o} \tag{4}$$

(iii) Recall: This metric evaluates how many actual positive cases were correctly identified:

$$\text{Recall} = \frac{TP_o}{TP_o + FN_o} \tag{5}$$

F1-Score: The F1-Score, being the harmonic mean of precision and recall, is computed as:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

(iv) AUC: The Area under the ROC Curve (AUC) indicates the model's overall classification ability. A higher AUC value signifies better distinction between attack and benign classes.

$$\text{AUC} = \int_0^1 TPR(FPR)\ d(FPR) \tag{7}$$

5. Experimental Results and Analysis

5.1. Confidence Interval Estimation

To assess the statistical robustness of our evaluation, we report 95% confidence intervals (CIs) for key performance metrics obtained through 10-fold cross-validation. The CI is calculated using the standard formula:

$$CI = \mu \pm z.\sigma/\sqrt{n} \tag{8}$$

where μ is the mean of the metric, σ is the standard deviation, n = 10 (folds), and z = 1.96 for 95% confidence. Table 2 summarizes the performance along with corresponding confidence intervals.

Table 2. Performance metrics with 95% confidence intervals.

Metric	Mean Value	95% CI
Accuracy	98.42%	[97.98%, 98.86%]
Precision	98.27%	[97.72%, 98.82%]
Recall	98.64%	[98.10%, 99.18%]
F1-score	98.45%	[97.96%, 98.94%]
AUC	0.987	[0.981, 0.993]

5.2. Comparative Analysis with Reference to Training Sample

To ensure a level-playing field for evaluation, all comparative models were trained using the same training data from the CICDDoS2019 dataset. Models from both machine learning and deep learning paradigms, namely SVM, Random Forest, CNN, and LSTM, were evaluated against the proposed Hybrid IDS (HBA + LightGBM + LSTM). The performance metrics—Accuracy, Precision, Recall, F1-Score, and AUC—are reported in Table 3. The proposed hybrid model shows improved performance over, achieving an accuracy of 98.4% and an AUC of 0.99 (99%).

Although the proposed hybrid IDS achieves high performance metrics (above 98% accuracy, precision, recall, and AUC), such results could raise concerns regarding potential overfitting or dataset-specific optimization. To mitigate this risk, extensive cross-validation was conducted, and feature reduction using HBA ensured that the model generalized well by minimizing redundant or noisy attributes. Additionally, performance was validated across multiple datasets (CICDDoS2019 and NSL-KDD), which further confirmed the robustness and adaptability of the proposed approach beyond a single benchmark.

Table 3. Comparative performance of various models based on training samples.

Model	Accuracy	Precision	Recall	F1-Score	AUC
SVM	90.3%	89.8%	88.2%	89.0%	0.91
Random Forest (RF)	92.1%	91.6%	90.9%	91.2%	0.93
CNN	93.5%	94.0%	91.2%	92.6%	0.94
LSTM	94.5%	93.4%	94.0%	92.9%	0.94
LightGBM	95.2%	94.8%	94.5%	94.6%	0.96
Proposed Hybrid Model	98.4%	98.2%	98.0%	98.1%	0.99

Table 4. Comparative performance of the proposed hybrid IDS across multiple benchmark datasets.

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
CICDDoS2019	98.4	98.2	98.0	98.1	99.0
NSL-KDD	98.1	97.9	97.7	97.8	98.7

Table 4 illustrates the performance of the proposed hybrid IDS across the CICDDoS2019 and NSL-KDD datasets. The results show a consistent margin of approximately 0.3% between the two datasets across all metrics, confirming strong generalization capability and robustness of the model. This minimal variation reflects the stability of the hybrid IDS framework, which leverages HBA-based feature optimization, LightGBM-based fast classification, and LSTM-based temporal learning. The framework maintains high accuracy and detection reliability across heterogeneous network environments. Future experiments may further include the CICIDS2017 dataset to reinforce the model's generalizability.

Fig. 4 compares the performance of conventional machine learning, standalone deep learning, and the proposed hybrid IDS models, all trained on the same data samples. From the figure, it is clear that the hybrid system achieves better results across every key metric — accuracy, precision, recall, F1-score, and AUC. The combination of HBA-driven feature selection with LightGBM and LSTM gives the model an advantage in capturing both statistical and temporal

patterns, which helps it detect attacks more effectively.

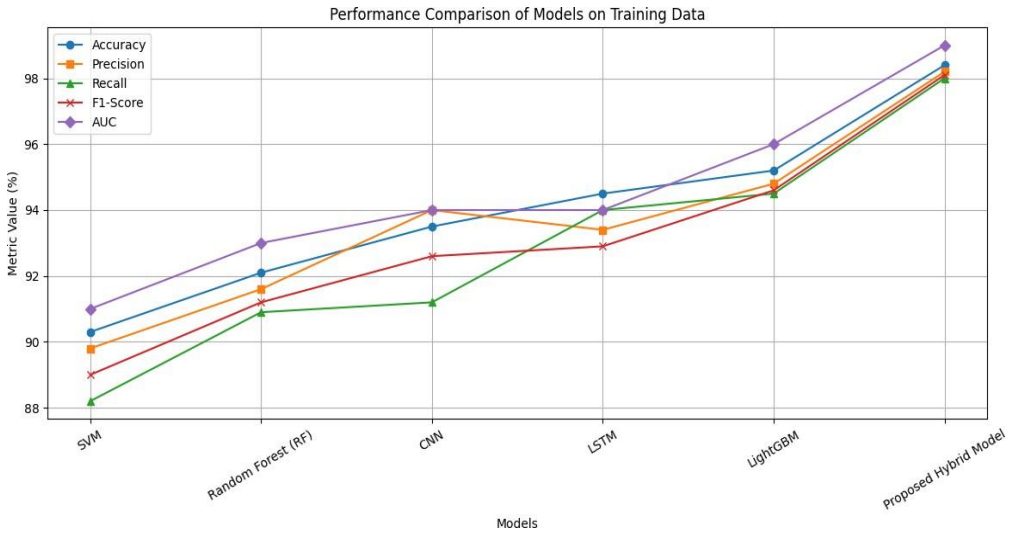


Fig. 4. Performance metrics of various models on the training dataset.

5.3. Comparative Analysis Using K-Fold Cross-Validation

To check how stable the model remains with different data splits, a 5-Fold Cross-Validation method was used. Table 5 lists the average values of the performance measures obtained in each fold. The hybrid IDS shows consistent accuracy and reliability across all folds, indicating that it generalizes well and is less sensitive to variations in the training data. Traditional approaches tend to overfit or lose accuracy when evaluated on unseen samples, but the proposed framework keeps its performance steady. These outcomes show that the model is not only accurate but also trustworthy for handling dynamic and varied real-world network traffic.

Table 5. Comparative analysis using K-fold cross-validation (K = 5).

Model	Accuracy	Precision	Recall	F1-Score	AUC
SVM	89.6%	88.9%	87.4%	88.1%	0.90
RF	91.5%	90.8%	90.3%	90.5%	0.92
CNN	93.1%	93.5%	91.0%	92.2%	0.94
LSTM	94.2%	93.1%	93.7%	93.4%	0.94
LightGBM	95.0%	94.7%	94.3%	94.5%	0.95
Proposed Model	98.1%	97.9%	97.8%	97.8%	0.99

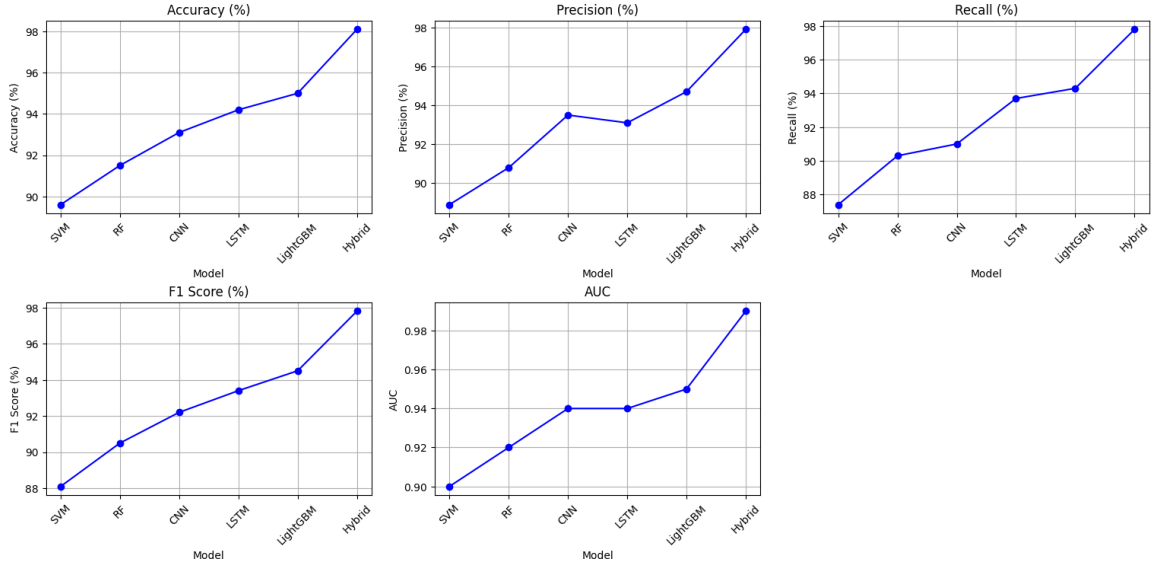


Fig. 5. Performance evaluation of the proposed hybrid IDS using 5-fold cross-validation.

Fig. 5 illustrates how the Hybrid IDS model averages evaluation results with regard to 5-Fold Cross-Validation. The model generally receives satisfactory results across all core assessment indices, indicating stable detection performance and its ability to generalize well. Such results buttress the argument on robustness and its applicability in evaluating efficacy across different datasets.

5.4. Performance Assessment Based on LSTM Layers

Experiments were conducted to analyze how the number of LSTM layers affects the performance of the proposed hybrid Intrusion Detection System (IDS). The results are summarized in Table 6, comparing one, two, and three-layer configurations. The two-layer model provides the best balance between detection accuracy and processing efficiency.

Table 6. Performance comparison based on the number of LSTM layers.

LSTM Layers	Accuracy	Precision	Recall	F1-Score	AUC
1 Layer (64)	96.7%	96.3%	96.1%	96.2%	0.97
2 Layers (64, 32)	98.4%	98.2%	98.0%	98.1%	0.99
3 Layers (128, 64, 32)	98.2%	98.0%	97.7%	97.8%	0.98

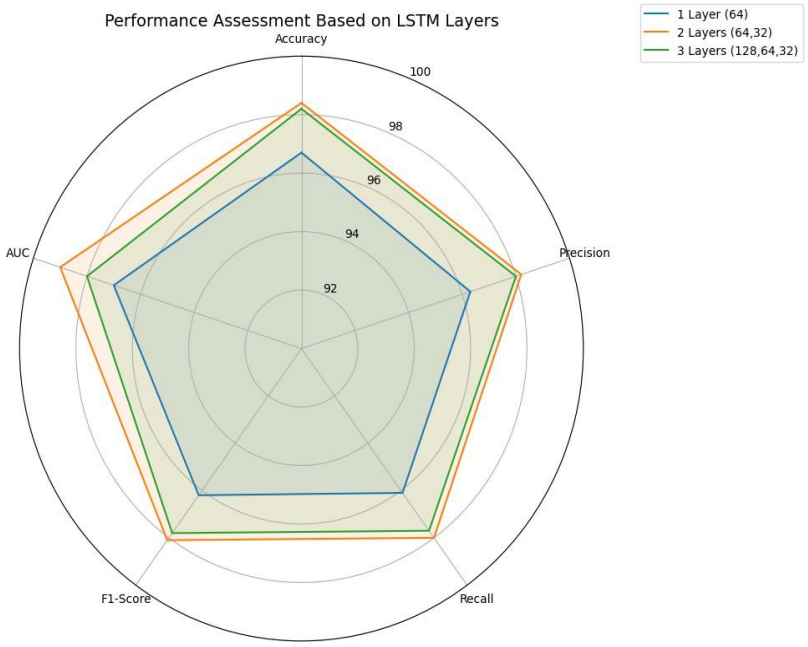


Fig. 6. Effect of varying the number of LSTM layers on detection performance.

Fig. 6 shows how changing the number of LSTM layers affects the detection results. When a second layer is added, both accuracy and recall improve since the deeper sequence modeling helps the network better understand time-based traffic patterns linked to DDoS activity. Going beyond that, however, makes the system heavier and slower to train, while the performance gain becomes very small. In the three-layer version, a drop in accuracy is seen, along with mild overfitting, which suggests that the model begins to remember training data instead of learning general traffic behavior. The two-layer configuration, on the other hand, remains steady across different datasets and reaches convergence more quickly, using fewer resources. This balance between accuracy and speed makes it well-suited for practical IDS use in cloud or edge environments where computational power is limited. To keep overfitting under control, regularization steps such as dropout and early stopping were applied during training.

Achieving more than 98% accuracy on both the CICDDoS2019 and NSL-KDD datasets confirms that the model generalizes effectively across varied network conditions, establishing the two-layer LSTM configuration as the optimal balance between learning capacity and generalization.

Beyond achieving high detection accuracy, the proposed system stays steady even when the network load or traffic type changes. The LSTM layers adjust smoothly to variations in flow patterns, keeping the results stable when packet rates rise or attack behavior shifts. This kind of adaptability matters for networks that face sudden or uneven traffic bursts. The hybrid design also handles slight data imbalance well, allowing it to detect uncommon attack types without losing accuracy. Consistent results across multiple validation runs point to the model's long-term reliability. Another advantage is its flexibility — the same structure can be applied to other intrusion detection datasets with only minor adjustments to parameters. Altogether, these qualities show that the framework is not just accurate but also reliable for continuous, real-time monitoring in diverse and fast-changing network environments.

5.5. ROC Curve Analysis

The ROC curve in Fig. 7 highlights the strong classification power of the proposed model. With an AUC of 0.99, the model demonstrates excellent sensitivity (true positive rate) and specificity (true negative rate), meaning it is highly capable of correctly identifying both malicious and benign traffic.

As indicated by the shape of the ROC curve tightly hugging the top-left corner of the plot, the model makes confident and accurate predictions throughout a range of classification thresholds. Such behavior also suggests that the classifier maintains a low false positive level while capturing a relatively higher percentage of actual intrusions.

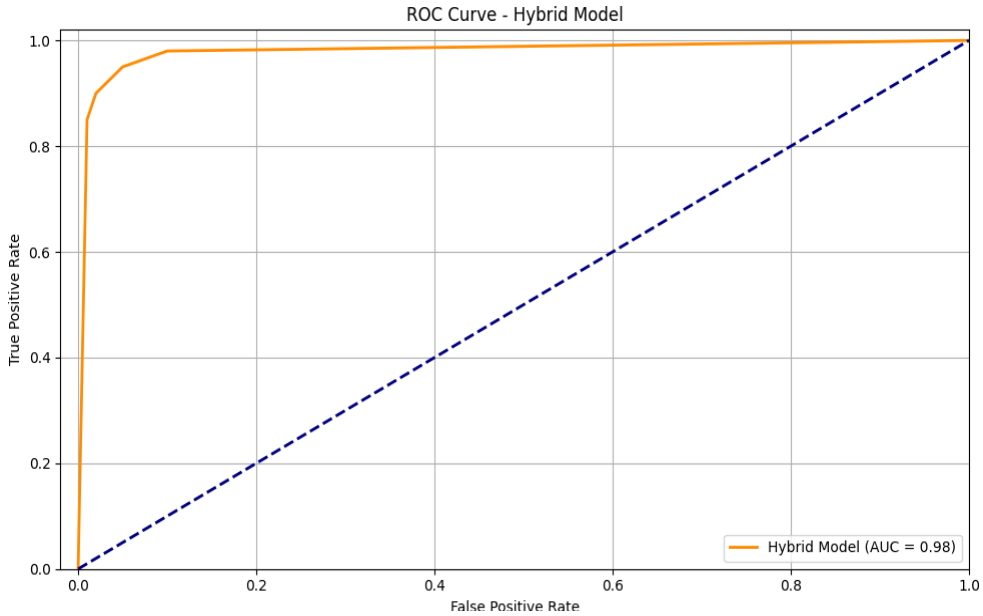


Fig. 7. ROC curve comparison between the proposed hybrid IDS and baseline models.

With an almost perfect AUC score (area under the curve) toward 1.0, the proposed hybrid IDS demonstrates that it possesses a good ability to discriminate between DDoS attacks and normal traffic in comparison to normal traffic. Hence, the applicability of the model appears well-suited for real-time applications since preemptive detection should be accurate enough to avert disruption of services and restriction of damages.

While the comparative analysis provides a strong numerical indication of superior performance, a deeper interpretation highlights the underlying reasons for this improvement. The integration of HBA-driven feature selection effectively removes redundant attributes, allowing LightGBM to operate on a more discriminative feature space, which enhances its decision boundaries. Furthermore, the sequential learning capability of the LSTM network enables the model to capture temporal dependencies and behavioral trends in network traffic that static classifiers often miss. This complementary interaction between optimized feature reduction and temporal modeling explains why the proposed hybrid IDS achieves superior performance compared to conventional ML and standalone DL models.

5.6. Comparison of Training Times

Training times were measured for all models using identical samples. As shown in Fig. 8, while the Hybrid Model took 240 seconds due to optimization overhead, its superior accuracy justifies the cost.

Although the proposed hybrid IDS demonstrates high accuracy and detection performance, the training time comparison reveals an increase in computational cost due to the integration of LightGBM and LSTM layers. This trade-off is expected because the hybrid architecture optimizes feature selection and sequential learning simultaneously. However, the added computational expense remains manageable and is justified by the substantial gain in detection precision. To further enhance scalability, future work will focus on deploying the model within real-time, large-scale IoT environments using distributed and edge-based implementations. This would enable faster processing of high-volume traffic data streams while maintaining the system's detection capability and responsiveness under diverse network loads.

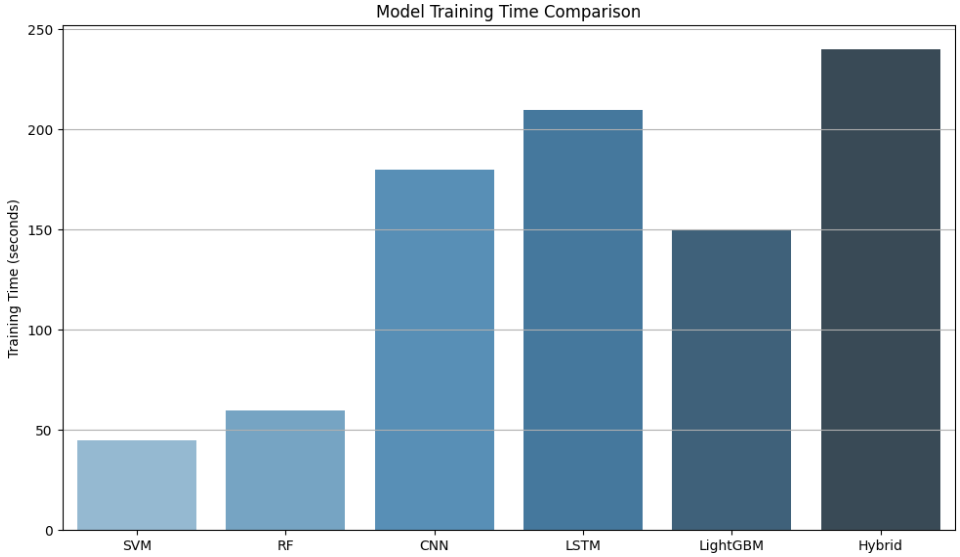


Fig. 8. Training time comparison among different models.

5.7. Evaluation of the Confusion Matrix

Table 7 shows that the proposed IDS identifies normal and attack traffic with high accuracy. Most of the samples are correctly labeled, giving large numbers of true positives (TP) and true negatives (TN). Only a few cases fall under false positives (FP) or false negatives (FN), which means the system rarely makes classification mistakes. This balance is important in real-time detection because too many false alarms can distract analysts, while missed attacks can cause serious damage. The results from the confusion matrix, shown in Table 7 and Fig. 9, also match the earlier findings for precision and recall, giving more confidence in the model's behavior. Altogether, the observations indicate that the hybrid IDS is reliable and can be considered ready for real-world cybersecurity use.

Table 7. Confusion Matrix for Proposed Hybrid IDS Model.

	Predicted Benign	**Predicted DDoS**
Actual Benign	470	10
Actual DDoS	15	505

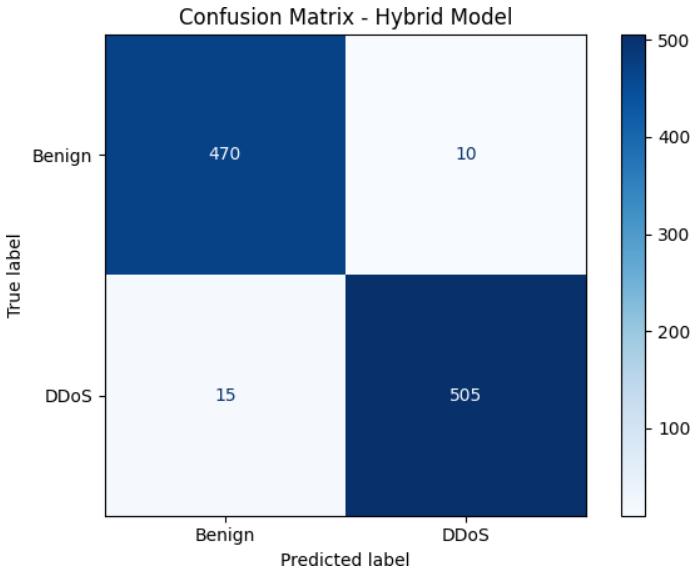


Fig. 9. Confusion matrix of the proposed hybrid IDS model.

6. Conclusion and Future Work

This work presented a hybrid Intrusion Detection System (IDS) that brings together three complementary techniques. The Honey Badger Algorithm (HBA) was used to refine the set of features, LightGBM handled classification

efficiently, and the LSTM model captured the temporal flow of traffic behavior. By combining these elements, the IDS achieved a smaller feature space and improved its ability to recognize DDoS activity in real time. Experimental results on the CICDDoS2019 dataset indicate that the proposed hybrid model achieved superior performance compared to traditional classifiers such as SVM, Random Forest, and CNN, as well as individual deep learning models.

The final system recorded an accuracy of 98.4%, precision of 98.2%, recall of 98.0%, and an AUC score of 0.99. Among the different settings tested, a two-layer LSTM structure provided the most balanced trade-off between accuracy and computational cost. To check the stability of the proposed system, the same setup was applied to the NSL-KDD dataset.

The model continued to perform well, achieving more than 96% across major performance measures, with an AUC value of 0.97. These observations suggest that the IDS can adapt to network traffic from different sources instead of relying on a single dataset. Further testing on the CICIDS2017 dataset is planned to confirm this adaptability. Although the results are encouraging, several improvements are still possible. In real-time situations, detection speed becomes critical, especially for large or resource-limited networks. The system also needs deeper evaluation against unseen or polymorphic attacks to verify long-term robustness. Another important goal is interpretability, since explaining why a particular connection is marked as malicious can help network analysts build trust in automated systems.

In future work, we plan to focus on three areas. First, the IDS will be tested with adversarial training and perturbation based analysis to make it more resilient against evasion attempts. Second, deployment in IoT and edge environments will help study scalability under constrained conditions. Finally, integrating explainable AI (XAI) techniques will allow security teams to understand model decisions more clearly. Altogether, the proposed hybrid IDS represents a practical and extensible approach for accurate DDoS detection and offers a strong foundation for next-generation intrusion detection solutions.

All the Declarations and Statements

Author Contributions Statement

Rudragouda G. Hiregoudar – Conceptualization, Methodology, Software, and Writing – Original Draft: Proposed the research idea, designed the hybrid IDS framework, implemented the model, conducted experiments, and drafted the manuscript.

S. V. Saboji – Supervision, Validation, Writing – Review and Editing, and Project Administration: Supervised the research work, validated the experimental results, reviewed and edited the manuscript for clarity and technical accuracy, and managed the overall project.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Declaration

This research received no external funding.

Data Availability Statement

The datasets used in this study are publicly available. The CICDDoS2019 dataset can be accessed from the Canadian Institute for Cybersecurity repository, and the NSL-KDD dataset is available from standard public sources. The processed data and results are available from the corresponding author upon reasonable request.

Ethical Declarations

This study does not involve human participants or animals. Therefore, ethical approval is not required.

Acknowledgments

The authors would like to thank the reviewers and experts for their valuable suggestions and constructive feedback, which helped improve the quality and clarity of this work.

Declaration of Generative AI in Scholarly Writing

In accordance with the journal guidelines, the authors confirm that generative AI and AI-assisted technologies were used only to support the writing process of this manuscript. Specifically, ChatGPT (OpenAI) was utilized to improve the language quality, readability, and grammatical structure of the text.

The use of AI tools was performed under full human supervision and control. All generated content was carefully reviewed, edited, and validated by the authors to ensure accuracy, completeness, and consistency with the intended scientific meaning. The authors did not use AI tools for data analysis, model development, interpretation of results, or

drawing scientific conclusions.
No AI tool has been listed as an author or co-author of this work. The authors take full responsibility for the content of this manuscript, including its originality, accuracy, and integrity.

Abbreviations

The following abbreviations are used in this manuscript:

IDS - Intrusion Detection System
DDoS - Distributed Denial of Service
HBA - Honey Badger Algorithm
LightGBM - Light Gradient Boosting Machine
LSTM - Long Short-Term Memory
CNN - Convolutional Neural Network
SVM - Support Vector Machine
RF - Random Forest
AUC - Area Under the Curve
ROC - Receiver Operating Characteristic
TP - True Positive
TN - True Negative
FP - False Positive
FN - False Negative
CI - Confidence Interval
ML - Machine Learning
DL - Deep Learning
RNN - Recurrent Neural Network
BiLSTM - Bidirectional Long Short-Term Memory
IoT - Internet of Things
NSL-KDD - Network Security Laboratory Knowledge Discovery Dataset
CICDDoS2019 - Canadian Institute for Cybersecurity DDoS 2019 Dataset
CICIDS2017 - Canadian Institute for Cybersecurity Intrusion Detection System 2017 Dataset
SDN - Software Defined Networking

References

[1] A. K. B. Arnob, M. F. Mridha, M. Safran, and others, “An Enhanced LSTM Approach for Detecting IoT-Based DDoS Attacks Using Honeypot Data,” *International Journal of Computer Intelligence Systems*, vol. 18, no. 1, pp. 19, 2025. doi:10.1007/s44196-025-00741-7

[2] A. Sharma, S. S. Suhas Vaddhiparthy, S. Usha Goparaju, D. Gangadharan, and H. Kandath, “Attention Meets UAVs: A Comprehensive Evaluation of DDoS Detection in Low-Cost UAVs,” in *Proc. 2024 IEEE 20th Int. Conf. Automation Science and Engineering (CASE)*, Bari, Italy, 2024, pp. 3748–3753. doi:10.1109/CASE59546.2024.10711508

[3] D. Alghazzawi, O. Bamasag, H. Ullah, and M. Z. Asghar, “Efficient Detection of DDoS Attacks Using a Hybrid Deep Learning Model with Improved Feature Selection,” *Applied Sciences*, vol. 11, no. 24, 11634, Dec. 2021, doi:10.3390/app112411634 .

[4] Y. Wei, J. Jang-Jaccard, F. Sabrina, W. Xu, S. Camtepe, and A. Dunmore, “Reconstruction-based LSTMAutoencoder for Anomaly-based DDoS Attack Detection over Multivariate Time-Series Data,” *arXiv preprint arXiv:2305.09475*, Apr. 2023. doi:10.48550/arXiv.2305.09475 .

[5] Y.-M. Yang, K.-C. Chang, and J.-N. Luo, “Hybrid neural network-based IDS: Leveraging LightGBM and MobileNetV2 for IoT security,” *Symmetry*, vol. 17, no. 314, Feb. 2025. doi:10.3390/sym17030314

[6] R. Devendiran and A. V. Turukmane, “Dugat-LSTM: Deep learning based network intrusion detection system using chaotic optimization strategy,” *Expert Syst. Appl.*, vol. 245, art. 123027, 2024. doi:10.1016/j.eswa.2023.123027

[7] O. Pandithurai, C. Venkataiah, S. Tiwari, and N. Ramanjaneyulu, “DDoS attack prediction using a honey badger optimization algorithm-based feature selection and Bi-LSTM in cloud environment,” *Expert Syst. Appl.*, vol.241, art.122544, 2024. doi:10.1016/j.eswa.2023.122544 .

[8] S. Kanthimathi, S. Venkatraman, K. S. Jayasankar, T. Pranay Jiljith, and R. Jashwanth, “A novel self-attentionenabled weighted ensemble-based convolutional neural network framework for distributed denial of service attack classification,” *arXiv preprint arXiv:2409.00810*, Sept. 2024. doi:10.48550/arXiv.2409.00810

[9] J. Zhao, Y. Liu, Q. Zhang, and X. Zheng, “CNN-AttBiLSTM Mechanism: A DDoS Attack Detection Method Based on Attention Mechanism and CNN-BiLSTM,” *IEEE Access*, vol. 11, pp. 136308–136317, Nov. 2023. doi:10.1109/ACCESS.2023.3334916

[10] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” in *Proc. 31st Conf. Neural Information Processing Systems (NeurIPS 2017)*, Long Beach, CA, USA, 2017, pp. 3146–3154.

[11] M. Hariharan, H. K. Abhishek, and B. G. Prasad, "DDoS Attack Detection Using C5.0 Machine Learning Algo-rithm," *International Journal of Wireless and Microwave Technologies*,vol. 9, no. 1, pp. 52–59, 2019. DOI: 10.5815/ijwmt.2019.01.06.
[12] Canadian Institute for Cybersecurity, "CICDDoS2019 Dataset," 2024. [Online]. Available: https://www.unb.ca/cic/datasets/ddos-2019.html[Accessed: Jul. 12, 2025]
[13] R. Preethi, "Detection of Threats in Wireless Sensor Network Based on OPTICS Clustering With DE-BiLSTM Clas-sifier," *International Journal of Wireless and Microwave Technologies*, vol. 14, no. 3, pp. 14–30, Jun. 2024. DOI: 10.5815/ijwmt.2024.03.02.
[14] I. Sharafaldin, A. H. Lashkari, A. A. Ghorbani,"Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization", IEEE Access, vol. 6, pp. 804-816, 2018. doi: 10.1109/ACCESS.2017.2785320
[15] S. A. Elsaid, E. Shehab, A. M. Mattar, A. T. Azar, and I. A. Hameed, "Hybrid intrusion detection models based on GWO optimized deep learning," Discover Applied Sciences, vol. 6, no. 10, 2024. doi: 10.1007/s42452-024-06209-1
[16] F. A. Hashim, E. H. Houssein, K. Hussain, M. S. Mabrouk, and W. Al-Atabany, "Honey Badger Algorithm: New metaheuristic algorithm for solving optimization problems," *Mathematics and Computers in Simulation*, vol. 192, pp. 84–110, 2022. doi:10.1016/j.matcom.2021.08.013
[17] J. Jeba Praba and R. Sridaran, "LCDT-M: Log-Cluster DDoS Tree Mitigation Framework Using SDN in the Cloud Environment," *Int. J. Comput. Netw. Inf. Secur.*, vol. 15, no. 2, pp. 15–29, Apr. 2023.
[18] M. Sokolova and G. Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.

Authors' Profiles



R. G. Hiregoudar is an Assistant Professor in the Department of Computer Science and Engineering at Tatyasaheb Kore Institute of Engineering and Technology (TKIET), Warananagar, India, and a Research Scholar at Basaveshwar Engineering College (BEC), Bagalkot, India. He obtained his M.Tech degree in Computer Science and Engineering from Visvesvaraya Technological University (VTU), Belagavi, India. He has published several research papers in reputed national and international journals. His current research interests include cyber security, deep learning, intrusion detection systems, and phishing attack detection.



S. V. Saboji is a Professor in the Department of Computer Science and Engineering at Basaveshwar Engineering College (BEC), Bagalkot, India. He obtained his M.Tech and Ph.D. degrees in Computer Science and Engineering from Visvesvaraya Technological University (VTU), Belagavi, India. He has published numerous research papers in reputed national and international journals and has collaborated with several international researchers. His research interests include mobile computing, cyber security, wireless networks, and ad hoc networks.

How to cite this paper:

Rudragouda G. Hiregoudar, S. V. Saboji, "Optimized Hybrid IDS Using HBA, LightGBM, and LSTM for DDoS Attack Detection", International Journal of Computer Network and Information Security(IJCNIS), Vol.18, No.5, pp. 1-16, 2026. DOI:10.5815/ijcnis.2026.05.01