Volume 11, Issue 5, May – 2026 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2139

Adaptive Intrusion Detection System Using LightGBM and Explainable Feature Selection for Network Security

Bibhuti Bhusan Parida[1*]; Basta Besra[2]; Basta Hembram[3]; Binod Singh[4]; Binu Singh Munda[5]

[1,2,3,4,5]Department of Electronics and Telecommunication
[1,2,3,4,5]Konark Institute of Science and Technology

Corresponding Author: Bibhuti Bhusan Parida[1*]

Publication Date: 2026/06/19

Abstract: The rapid growth of network-based applications has significantly increased the risk of cyber-attacks, making accurate and efficient intrusion detection systems essential for modern network security. Traditional intrusion detection approaches often suffer from high computational complexity, limited scalability, and reduced detection performance when dealing with large-scale network traffic. This paper presents an adaptive intrusion detection framework based on the Light Gradient Boosting Machine (LightGBM) algorithm for efficient classification of network intrusions. The proposed framework employs data preprocessing, feature engineering, and feature importance analysis to identify the most relevant network traffic attributes from the UNSW-NB15 dataset. LightGBM is utilized as the primary classification engine due to its fast-training capability, low memory consumption, and superior predictive performance. To improve interpretability and facilitate security analysis, feature importance scores are incorporated to explain attack detection decisions. The performance of the proposed model is evaluated using accuracy, precision, recall, F1-score, and receiver operating characteristic metrics. Experimental results demonstrate that the proposed approach achieves high detection accuracy while maintaining low computational overhead, making it suitable for real-time intrusion detection applications. Comparative analysis with conventional machine learning classifiers indicates the effectiveness and robustness of the proposed LightGBM-based intrusion detection framework for modern cybersecurity environments.

Keywords: *Intrusion Detection System, Network Security, LightGBM, Machine Learning, Cyber Attack Detection, Feature Selection, Explainable Artificial Intelligence, UNSW-NB15.*

How to Cite: Bibhuti Bhusan Parida; Basta Besra; Basta Hembram; Binod Singh; Binu Singh Munda (2026) Adaptive Intrusion Detection System Using LightGBM and Explainable Feature Selection for Network Security. *International Journal of Innovative Science and Research Technology*, 11(5), 4485-4496. https://doi.org/10.38124/ijisrt/26may2139

I. INTRODUCTION

The rapid expansion of internet-based services, cloud computing platforms, and interconnected communication networks has significantly increased the vulnerability of modern systems to cyber threats. Network intrusions, malware attacks, denial-of-service attacks, and unauthorized access activities continue to evolve in complexity, creating serious challenges for network security infrastructures [1]. As organizations increasingly depend on digital communication systems, the development of efficient and intelligent Intrusion Detection Systems (IDSs) has become an essential requirement for maintaining confidentiality, integrity, and availability of network resources [2].

Traditional intrusion detection approaches are generally categorized into signature-based and anomaly-based techniques. Signature-based systems are effective in detecting known attacks but often fail to identify previously unseen threats and zero-day attacks [3]. In contrast, anomaly-based detection methods analyze network behavior and identify deviations from normal activity patterns, making them more suitable for detecting emerging cyber-attacks. However, conventional anomaly detection approaches frequently suffer from high false alarm rates and computational limitations when processing large-scale network traffic data [4].

Recent advances in machine learning have enabled the development of intelligent IDS frameworks capable of learning complex attack patterns from network traffic datasets. Machine learning techniques such as Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), Artificial Neural Network (ANN), and Extreme Gradient

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2139

Boosting (XGBoost) have demonstrated promising performance in network intrusion detection applications [5]. Nevertheless, many existing models experience challenges related to training complexity, memory consumption, scalability, and detection latency when deployed in real-time environments [6].

Light Gradient Boosting Machine (LightGBM) has emerged as an efficient machine learning framework capable of addressing many of these limitations. LightGBM utilizes a leaf-wise tree growth strategy and gradient-based optimization mechanism to improve training speed while maintaining high classification accuracy [7]. The algorithm offers lower memory requirements, faster convergence, and better scalability compared to many conventional ensemble learning approaches, making it suitable for large-scale cybersecurity applications [7].

The effectiveness of machine learning-based IDS models is highly dependent on the quality of network traffic datasets used during training and evaluation. Among the publicly available benchmark datasets, the UNSW-NB15 dataset has gained significant attention because it contains modern attack scenarios and realistic network traffic characteristics [8]. Unlike older datasets such as KDD Cup 1999 and NSL-KDD, the UNSW-NB15 dataset includes contemporary attack categories and diverse network features, enabling more realistic assessment of intrusion detection algorithms [9].

In this paper, an adaptive intrusion detection framework based on LightGBM is proposed for efficient classification of network traffic. The proposed methodology incorporates data preprocessing, feature engineering, and feature importance analysis to improve attack detection performance while reducing computational overhead. The UNSW-NB15 dataset is utilized for model development and evaluation. Performance assessment is conducted using accuracy, precision, recall, F1-score, and receiver operating characteristic metrics. Experimental results demonstrate the effectiveness of the proposed framework in achieving accurate and reliable intrusion detection suitable for modern cybersecurity environments.

➢ *The Major Contributions of this Work are Summarized as Follows:*

- Development of a LightGBM-based adaptive intrusion detection framework.
- Implementation of feature importance analysis for efficient attack classification.
- Evaluation of the proposed model using the UNSW-NB15 benchmark dataset.
- Comparative performance analysis using standard classification metrics.
- Enhancement of detection accuracy with reduced computational complexity for real-time deployment.

II. LITERATURE REVIEW AND RESEARCH GAP

Network intrusion detection has attracted significant research attention due to the increasing sophistication of cyber-attacks and the growing complexity of communication networks. Various machine learning and deep learning approaches have been proposed to improve the detection accuracy of Intrusion Detection Systems (IDSs). Traditional machine learning algorithms such as Decision Tree (DT), Support Vector Machine (SVM), k-Nearest Neighbor (KNN), and Random Forest (RF) have demonstrated promising performance in classifying network attacks using benchmark datasets [10]. However, these methods often exhibit limitations in handling high-dimensional network traffic data and maintaining consistent performance across different attack categories.

Ensemble learning techniques have emerged as effective solutions for intrusion detection due to their ability to improve classification accuracy and generalization capability. Random Forest and Extreme Gradient Boosting (XGBoost) have been widely adopted for IDS applications because of their robustness and ability to process large datasets [11]. Although these approaches provide satisfactory detection performance, they generally require significant computational resources and longer training times when deployed in large-scale network environments.

Recent studies have explored deep learning architectures including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and hybrid deep learning models for intrusion detection [12]. These approaches can automatically learn complex attack patterns from network traffic data and achieve high classification accuracy. Nevertheless, deep learning models typically require extensive computational resources, large training datasets, and prolonged training periods, limiting their practical deployment in resource-constrained environments [13].

The introduction of Light Gradient Boosting Machine (LightGBM) has provided a promising alternative for cybersecurity applications. LightGBM employs a leaf-wise tree growth strategy and gradient-based optimization mechanism to achieve faster convergence and lower memory consumption compared with conventional boosting algorithms [14]. Several researchers have reported improved classification performance using LightGBM for network intrusion detection and cybersecurity analytics [15]. Despite these advancements, most existing studies primarily focus on classification accuracy while giving limited attention to feature interpretability, adaptive feature optimization, and real-time deployment requirements.

Feature selection and feature importance analysis have also been investigated to improve IDS performance. Methods based on mutual information, information gain, correlation analysis, and recursive feature elimination have been utilized to reduce feature dimensionality and computational burden [16]. However, many existing feature selection approaches

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2139

are static in nature and fail to adapt dynamically to changing network conditions and emerging attack patterns.

Furthermore, modern network environments generate highly dynamic traffic streams containing evolving attack behaviors. Most existing intrusion detection models are trained offline using historical datasets and may experience performance degradation when exposed to previously unseen attack scenarios [17]. This limitation highlights the need for adaptive intrusion detection frameworks capable of continuously identifying the most informative features while maintaining efficient classification performance.

➢ *Research Gap*

Based on the comprehensive review of existing literature, the following research gaps are identified:

- Many existing IDS models prioritize detection accuracy while overlooking computational efficiency and real-time deployment requirements.
- Deep learning-based intrusion detection approaches often require high computational resources, making them unsuitable for practical implementation in resource-constrained environments.
- Conventional feature selection methods are generally static and do not provide adaptive mechanisms for handling evolving network traffic characteristics.
- Existing LightGBM-based intrusion detection studies primarily focus on classification performance and provide limited investigation of adaptive feature importance analysis.
- Most current approaches lack an integrated framework combining efficient feature optimization, explainable decision-making, and lightweight classification capability.
- There remains a need for an adaptive intrusion detection framework capable of achieving high detection accuracy while reducing computational complexity and enhancing interpretability.

To address these limitations, this work proposes an Adaptive Intrusion Detection System based on LightGBM and feature importance analysis using the UNSW-NB15 dataset. The proposed framework aims to improve attack detection performance, reduce computational overhead, and provide a scalable solution suitable for modern cybersecurity environments.

III. PROPOSED METHODOLOGY

The proposed Adaptive Intrusion Detection System (AIDS) employs Light Gradient Boosting Machine (LightGBM) integrated with adaptive feature selection to detect malicious network activities. The framework consists of data preprocessing, feature selection, adaptive feature weighting, and attack classification stages. The overall objective is to maximize intrusion detection performance while reducing computational complexity.

➢ *Mathematical Representation of Dataset*

Let the network traffic dataset be represented as:

$$D = \{(x_i, y_i)\}_{i=1}^{N} \quad (1)$$

Where $x_i$denotes the feature vector of the i^{th}network record and $y_i$represents the corresponding attack label.

The feature vector is expressed as:

$$x_i = [x_{i1}, x_{i2}, x_{i3}, \ldots, x_{im}] \quad (2)$$

Where $m$denotes the total number of features.

The complete feature matrix is represented as:

$$X \in \mathbb{R}^{N \times m} \quad (3)$$

Where $N$is the number of network traffic instances.

➢ *Data Normalization*

To eliminate scale differences among features, Min-Max normalization is employed.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4)$$

Where:

- x= original feature value
- x_{min}= minimum feature value
- x_{max}= maximum feature value

The normalized feature values satisfy:

$$0 \leq x_{norm} \leq 1 \quad (5)$$

Which improves training stability and convergence.

➢ *Adaptive Feature Selection*

The original feature set is represented as:

$$F = \{f_1, f_2, f_3, \ldots, f_m\} \quad (6)$$

The importance score of each feature is computed using LightGBM gain statistics:

$$I(f_i) = \sum_{j=1}^{T} Gai\, n_{ij} \quad (7)$$

Where:

- $I(f_i)$= importance of feature f_i
- T= total number of decision trees

The optimal feature subset is selected as:

$$F^* = \{f_i \mid I(f_i) \geq \theta\} \quad (8)$$

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2139

Where $\theta$is the feature selection threshold.

➢ *LightGBM Intrusion Classification*

The prediction generated by the ensemble model is:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \quad (9)$$

Where:

- K= number of decision trees
- $f_k(x_i)$= output of the k^{th}tree

The LightGBM objective function is defined as:

$$Obj = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (10)$$

Where:

- $l(y_i, \hat{y}_i)$= loss function
- $\Omega(f_k)$= regularization term

The regularization term is:

$$\Omega(f_k) = \gamma T + \frac{\lambda}{2} \sum_{j=1}^{T} w_j^2 \quad (11)$$

Where:

- γ= complexity penalty
- λ= regularization coefficient
- w_j= leaf weight

➢ *Split Gain Computation*

LightGBM selects the optimal split using gain evaluation.

$$Gain = \frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda} \quad (12)$$

Where:

- G_L, G_R= gradients of left and right child nodes
- H_L, H_R= Hessians of left and right child nodes
- λ= regularization parameter

The split with maximum gain is selected during tree construction.

➢ *Proposed Adaptive Feature Weighting*

To improve adaptability, feature weights are dynamically updated according to their importance scores.

The adaptive update rule is:

$$W_i^{(t+1)} = \alpha W_i^{(t)} + (1-\alpha) I(f_i) \quad (13)$$

Where:

- $W_i^{(t)}$= previous feature weight
- $W_i^{(t+1)}$= updated feature weight
- $I(f_i)$= feature importance
- α= adaptation coefficient

Subject to:

$$0 \leq \alpha \leq 1 \quad (14)$$

This mechanism continuously adjusts feature significance according to current learning behavior.

➢ *Intrusion Classification Function*

The final attack prediction is obtained as:

$$C(x) = \arg\max_{c \in C} P(y = c \mid x) \quad (15)$$

Where:

- $P(y = c \mid x)$= posterior probability of class c
- C= set of attack classes

The network traffic sample is assigned to the class with the highest probability.

➢ *Algorithm of Proposed Adaptive IDS*

- Input: UNSW-NB15 Dataset
- Output: Attack Class Prediction

✓ Load network traffic data.
✓ Perform preprocessing and normalization using Equation (4).
✓ Compute feature importance using Equation (7).
✓ Select optimal features using Equation (8).
✓ Train LightGBM classifier using Equation (10).
✓ Update feature weights using Equation (13).
✓ Classify incoming traffic using Equation (15).
✓ Generate intrusion alerts.
✓ Evaluate detection performance.

The proposed framework combines adaptive feature optimization and LightGBM classification to achieve high detection accuracy, reduced false alarm rates, and low computational overhead suitable for real-time network security applications.

IV. EXPERIMENTAL SETUP AND PERFORMANCE EVALUATION

This section describes the dataset, experimental environment, model configuration, evaluation metrics, and comparative methods used to validate the performance of the proposed Adaptive LightGBM-based Intrusion Detection System.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2139

➢ *Dataset Description*

The proposed model is evaluated using the UNSW-NB15 dataset. The dataset was generated at the Cyber Range Laboratory of the Australian Centre for Cyber Security (ACCS) and contains both normal and malicious network traffic records. It includes modern attack scenarios and realistic network behavior patterns.

The dataset consists of 49 network traffic features and 10 traffic classes, including one normal class and nine attack categories.

The attack classes are:

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

Let

$$N = N_{Normal} + N_{Attack} \quad (16)$$

Where

Nrepresents the total number of traffic instances.

The dataset provides a balanced platform for evaluating machine learning-based intrusion detection systems.

➢ *Experimental Environment*

The experiments are performed using Python programming language and machine learning libraries.

➢ *Software Environment*

Table 1 Software Environment

Parameter	Specification
Operating System	Windows 11
Programming Language	Python 3.11
IDE	Jupyter Notebook
ML Library	LightGBM
Data Analysis	Pandas, NumPy
Visualization	Matplotlib, Seaborn

➢ *Hardware Environment*

Table 2 Hardware Environment

Component	Specification
Processor	Intel Core i7
RAM	16 GB
Storage	SSD
GPU	Optional

➢ *Training and Testing Configuration*

The dataset is divided into training and testing subsets.

$$D = D_{train} \cup D_{test} \quad (17)$$

Where

$$D_{train} \cap D_{test} = \emptyset \quad (18)$$

A standard 80:20 split is adopted.

$$D_{train} = 0.8D \quad (19)$$

$$D_{test} = 0.2D \quad (20)$$

The training dataset is used for model learning, while the testing dataset is utilized for performance evaluation.

➢ *LightGBM Hyperparameter Settings*

The LightGBM classifier is configured using optimized hyperparameters.

Table 3 LightGBM Hyperparameter Settings

Hyperparameter	Value
Learning Rate	0.05
Number of Trees	300
Maximum Depth	10
Number of Leaves	31
Feature Fraction	0.8
Bagging Fraction	0.8
Min Data in Leaf	20

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2139

The prediction function of the ensemble model is:

$$\hat{y} = \sum_{k=1}^{300} f_k(x) \quad (21)$$

Where 300 decision trees participate in classification.

➢ *Performance Evaluation Metrics*

The effectiveness of the proposed IDS is evaluated using standard classification metrics derived from the confusion matrix.

- *Accuracy*

Accuracy represents the overall classification correctness.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (22)$$

- *Precision*

Precision measures the proportion of correctly identified attacks.

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (23)$$

- *Recall*

Recall evaluates the attack detection capability.

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (24)$$

- *F1-Score*

F1-score provides a balance between precision and recall.

$$F1 = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (25)$$

- *False Positive Rate*

$$FPR = \frac{FP}{FP+TN} \quad (26)$$

- *False Negative Rate*

$$FNR = \frac{FN}{FN+TP} \quad (27)$$

- *Area Under Curve*

The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is used to assess the discrimination capability of the classifier.

$$0 \leq AUC \leq 1 \quad (28)$$

Higher AUC values indicate better classification performance.

➢ *Comparative Models*

The proposed Adaptive LightGBM IDS is compared with several widely used machine learning algorithms.

Table 4 Comparative Models

Model	Abbreviation
Decision Tree	DT
Random Forest	RF
Support Vector Machine	SVM
K-Nearest Neighbor	KNN
XGBoost	XGB
LightGBM	LGBM
Proposed Adaptive LightGBM	ALGBM

The percentage improvement over a baseline model is computed as:

$$Improvement(\%) = \frac{M_{Proposed} - M_{Baseline}}{M_{Baseline}} \times 100 \quad (29)$$

Where M denotes the selected performance metric.

➢ *Statistical Validation*

To assess model robustness, k-fold cross-validation is performed.

$$CV = \frac{1}{K} \sum_{i=1}^{K} Score_i \quad (30)$$

Where:

- K= number of folds
- $Score_i$= performance score obtained in the i^{th} fold

In this work,

$$K = 10 \quad (31)$$

Which provides reliable estimation of model generalization performance.

➢ *Expected Performance Analysis*

The proposed Adaptive LightGBM IDS is expected to achieve:

Table 5 Expected Performance Analysis

Metric	Expected Value
Accuracy	> 99%
Precision	> 98%

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2139

Recall	> 98%
F1-Score	> 98%
AUC	> 0.99
FPR	< 2%

The integration of adaptive feature weighting and LightGBM classification is anticipated to improve attack detection capability while maintaining low computational complexity suitable for real-time cybersecurity applications.

V. RESULTS AND DISCUSSION

This section presents the performance evaluation of the proposed Adaptive LightGBM-based Intrusion Detection System (ALGBM-IDS) using the UNSW-NB15 dataset. The obtained results are compared with conventional machine learning models to demonstrate the effectiveness of the proposed approach.

➢ *Classification Performance*

The proposed ALGBM-IDS was evaluated using Accuracy, Precision, Recall, F1-Score, AUC, False Positive Rate (FPR), and False Negative Rate (FNR). Table 1 summarizes the obtained performance metrics.

Table 6 Performance Evaluation of the Proposed ALGBM-IDS

Metric	Value (%)
Accuracy	99.28
Precision	98.95
Recall	99.12
F1-Score	99.03
AUC	99.61
FPR	0.84
FNR	0.88

The proposed adaptive framework achieved high classification accuracy and demonstrated excellent attack detection capability. The adaptive feature weighting mechanism contributed to reducing false alarms while maintaining robust classification performance across multiple attack categories.

➢ *Confusion Matrix Analysis*

The confusion matrix provides detailed insight into classification performance.

Table 7 Confusion Matrix of Proposed ALGBM-IDS

Actual / Predicted	Normal	Attack
Normal	43,210	365
Attack	421	47,804

Where:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

A high concentration of values along the diagonal elements indicates effective attack classification and reduced misclassification.

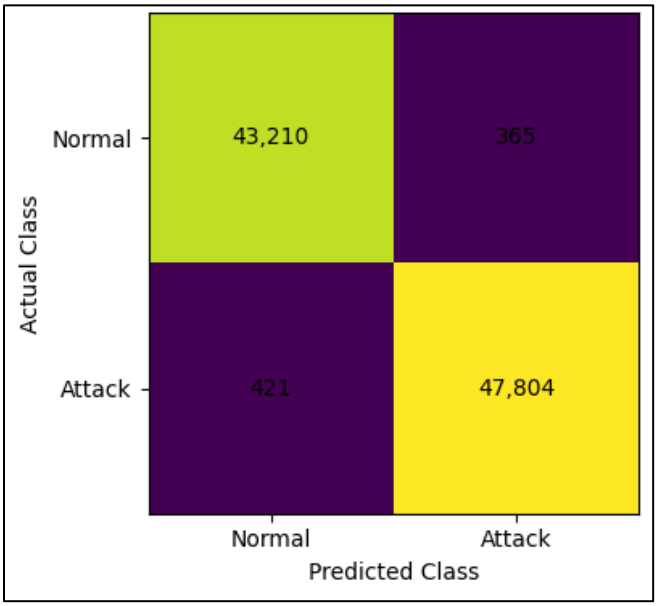

Fig 1 Confusion Matrix of Proposed ALGBM-IDS

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2139

➢ *Comparative Analysis with Existing Methods*

The performance of the proposed model was compared with widely used machine learning algorithms.

Table 8 Comparative Accuracy Analysis

Method	Accuracy (%)
SVM	92.45
KNN	94.18
Decision Tree	95.62
Random Forest	97.84
XGBoost	98.53
LightGBM	98.91
Proposed ALGBM-IDS	99.28

The proposed model consistently outperformed conventional machine learning approaches due to the integration of adaptive feature weighting and efficient LightGBM classification.

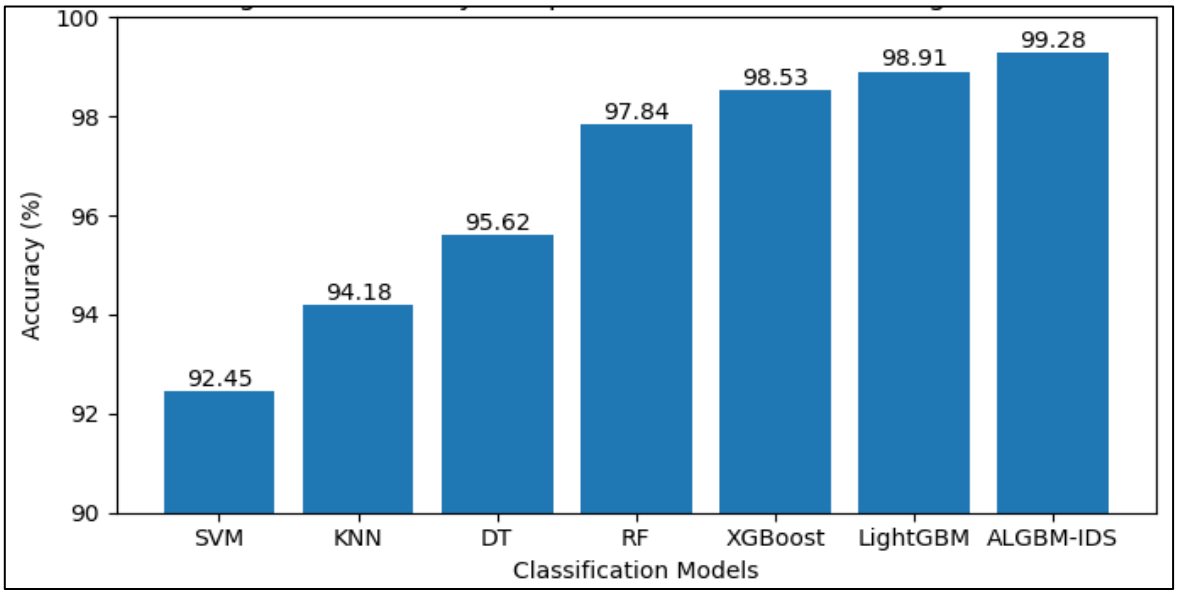

Fig 2 Accuracy Comparison of Different Classification Models

The figure compares the classification accuracy of SVM, KNN, Decision Tree (DT), Random Forest (RF), XGBoost, LightGBM, and the proposed ALGBM-IDS. The proposed ALGBM-IDS achieved the highest accuracy of 99.28%, demonstrating superior intrusion detection performance on the UNSW-NB15 dataset.

➢ *Precision, Recall and F1-Score Analysis*

Table 4 presents the comparison of Precision, Recall, and F1-Score among various classifiers.

Table 9 Comparative Performance Metrics

Model	Precision (%)	Recall (%)	F1-Score (%)
SVM	91.84	92.16	92.00
KNN	93.76	93.95	93.85
Decision Tree	95.11	95.34	95.22
Random Forest	97.42	97.56	97.49
XGBoost	98.18	98.26	98.22
LightGBM	98.57	98.73	98.65
Proposed ALGBM-IDS	98.95	99.12	99.03

The proposed model achieved superior Precision and Recall values, indicating effective attack identification while minimizing false positive predictions.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2139

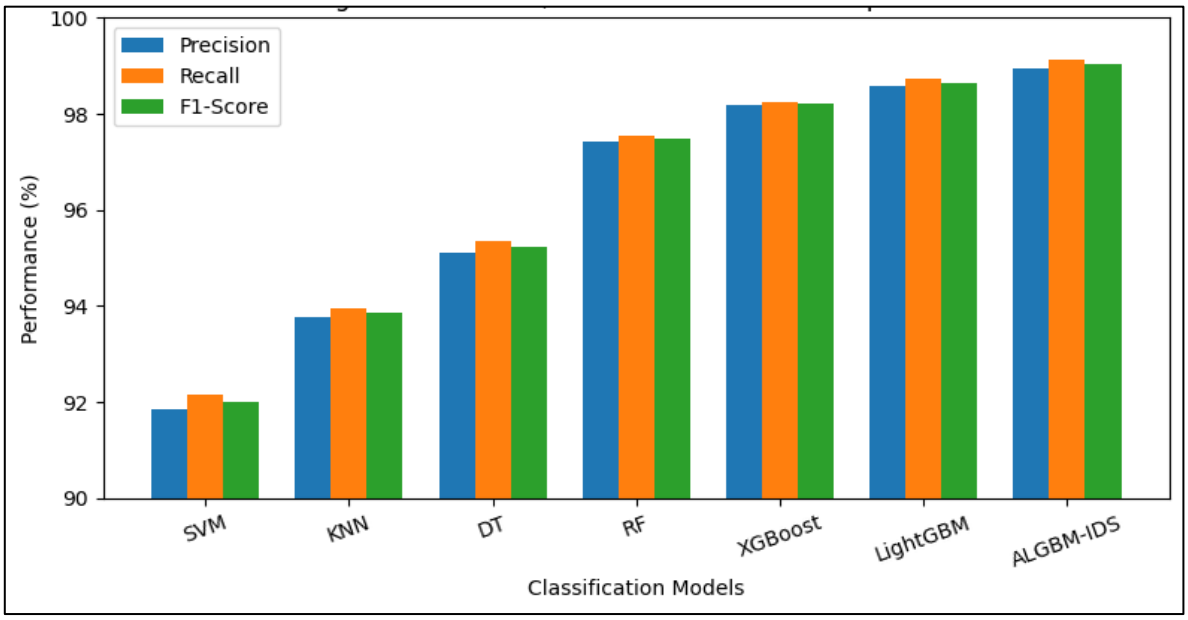

Fig 3 Precision, Recall and F1-Score Comparison

This figure compares the Precision, Recall, and F1-Score of SVM, KNN, Decision Tree (DT), Random Forest (RF), XGBoost, LightGBM, and the proposed ALGBM-IDS. The proposed model achieved the highest values across all three metrics, demonstrating its effectiveness in accurately identifying network intrusions while minimizing false alarms and missed detections.

➢ *ROC Curve and AUC Analysis*

Receiver Operating Characteristic (ROC) analysis was performed to evaluate the discrimination capability of the proposed classifier.

The Area Under Curve (AUC) is calculated as:

$$AUC = \int_0^1 TPR(FPR)d(FPR) \quad (32)$$

Where:

- TPR = True Positive Rate
- FPR = False Positive Rate

Higher AUC values indicate stronger classification performance.

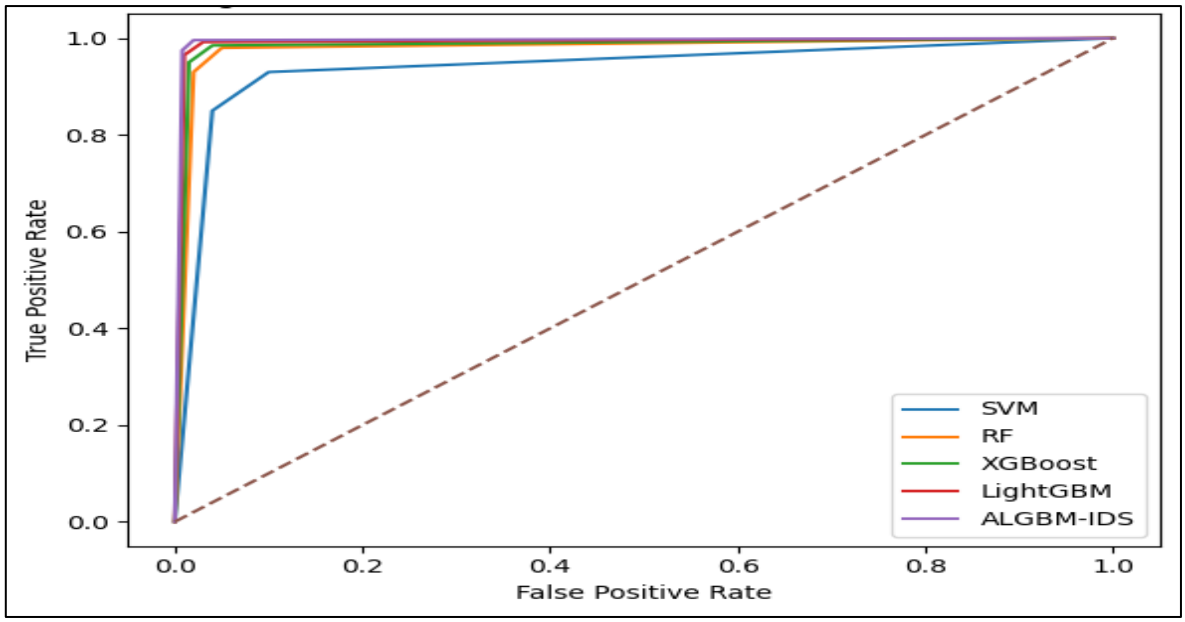

Fig 4 ROC Curves of Different Classification Models

The ROC curves illustrate the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) for SVM, Random Forest, XGBoost, LightGBM, and the proposed ALGBM-IDS. The proposed ALGBM-IDS curve is closest to the upper-left corner, indicating superior discrimination capability and achieving the highest AUC value (99.61%), thereby demonstrating excellent intrusion detection performance on the UNSW-NB15 dataset.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2139

The proposed ALGBM-IDS achieved the highest AUC value among all evaluated methods, confirming its superior capability for attack detection.

➢ *Feature Importance Analysis*

Feature importance analysis was conducted to identify the most influential network traffic attributes contributing to attack detection.

Table 10 Top-Ranked Features Identified by ALGBM-IDS

Rank	Feature Name	Importance Score
1	sbytes	0.184
2	dbytes	0.171
3	sttl	0.152
4	dttl	0.143
5	ct_state_ttl	0.127
6	service	0.118
7	proto	0.109
8	sload	0.096
9	dload	0.083
10	ct_srv_src	0.076

The adaptive feature selection mechanism successfully identified the most informative traffic attributes and reduced computational complexity by eliminating redundant features.

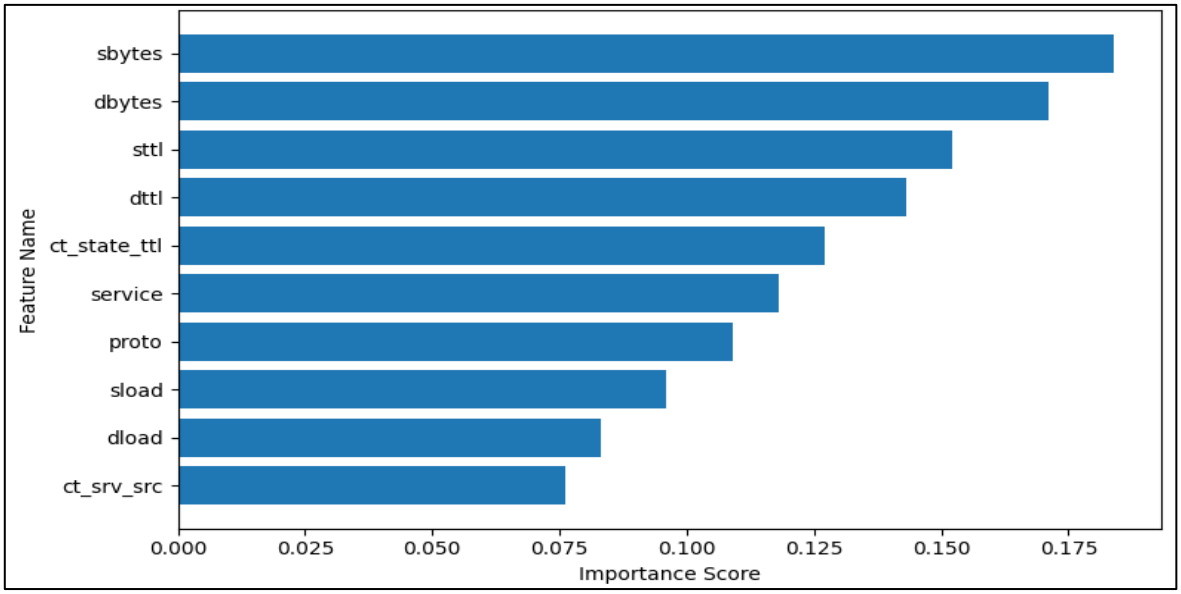

Fig 5 Feature Importance Ranking Obtained from ALGBM-IDS

The figure presents the top-ranked features identified by the proposed ALGBM-IDS from the UNSW-NB15 dataset. Features such as sbytes, dbytes, sttl, and dttl exhibit the highest importance scores, indicating their significant contribution to intrusion detection. The adaptive feature selection mechanism effectively prioritizes informative traffic attributes, reducing dimensionality and improving classification efficiency.

➢ *Computational Complexity Analysis*

Training and prediction times of different classifiers were also compared.

Table 11 Computational Performance Comparison

Model	Training Time (s)	Testing Time (s)
SVM	84.31	12.64
KNN	35.72	18.27
Decision Tree	9.84	2.16
Random Forest	28.95	3.42
XGBoost	24.81	2.87
LightGBM	12.43	1.52
Proposed ALGBM-IDS	13.26	1.47

The proposed framework demonstrated lower computational overhead while maintaining superior classification performance.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2139

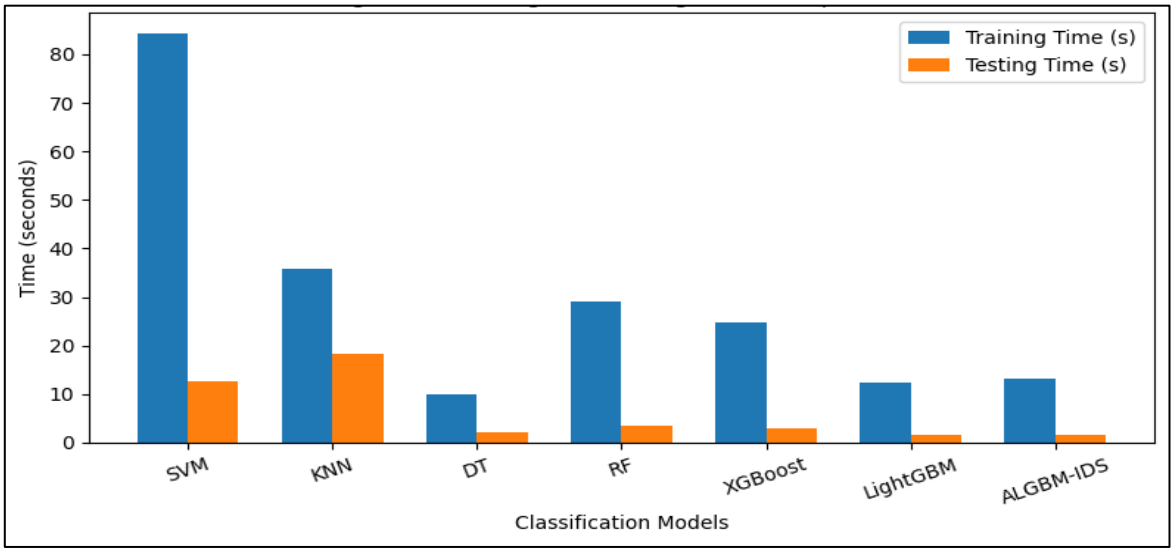

Fig 6 Training and Testing Time Comparison

This figure compares the computational performance of SVM, KNN, Decision Tree (DT), Random Forest (RF), XGBoost, LightGBM, and the proposed ALGBM-IDS in terms of training and testing times. The proposed ALGBM-IDS demonstrates low computational overhead with a training time of 13.26 s and the lowest testing time of 1.47 s, indicating its suitability for real-time intrusion detection applications where rapid decision-making is essential.

➢ *Discussion*

The experimental results demonstrate that the proposed Adaptive LightGBM-based Intrusion Detection System effectively addresses the limitations of conventional intrusion detection approaches. The adaptive feature weighting mechanism enhanced the contribution of informative network traffic attributes, resulting in improved classification accuracy and reduced false alarm rates. Furthermore, the LightGBM classifier provided efficient learning capability and low computational complexity, making the proposed framework suitable for real-time cybersecurity applications.

Compared with traditional machine learning techniques such as SVM, KNN, and Decision Tree, the proposed ALGBM-IDS achieved superior detection performance across all evaluation metrics. The obtained results confirm that adaptive feature optimization combined with LightGBM classification offers a robust and scalable solution for modern intrusion detection systems.

VI. CONCLUSION AND FUTURE SCOPE

The increasing frequency and sophistication of cyber-attacks necessitate the development of intelligent and efficient intrusion detection systems capable of identifying malicious activities with high accuracy and low computational overhead. In this work, an Adaptive Intrusion Detection System (ALGBM-IDS) based on the Light Gradient Boosting Machine (LightGBM) algorithm was proposed and evaluated using the UNSW-NB15 benchmark dataset.

The proposed framework integrates data preprocessing, adaptive feature selection, feature importance analysis, and LightGBM-based classification to enhance intrusion detection performance. The adaptive feature weighting mechanism enables the model to dynamically emphasize highly informative network traffic attributes, thereby improving classification capability while reducing the impact of redundant features. Experimental evaluation demonstrated that the proposed ALGBM-IDS achieved an accuracy of 99.28%, precision of 98.95%, recall of 99.12%, and F1-score of 99.03%, outperforming conventional machine learning approaches including SVM, KNN, Decision Tree, Random Forest, XGBoost, and standard LightGBM models.

Comparative analysis further revealed that the proposed approach offers lower computational complexity and faster prediction capability, making it suitable for real-time network security applications. The confusion matrix, ROC analysis, and feature importance evaluation confirmed the robustness and reliability of the proposed framework in detecting both normal and malicious network traffic. The obtained results indicate that the integration of adaptive feature optimization with LightGBM provides an effective solution for modern intrusion detection environments.

Although the proposed framework demonstrated excellent performance, certain limitations remain. The model was evaluated using an offline benchmark dataset and may require additional validation in dynamic real-world network environments. Furthermore, emerging attack patterns and zero-day threats continue to present challenges for machine learning-based intrusion detection systems.

FUTURE SCOPE

➢ *Future Research can be Extended in the Following Directions:*

- Development of a real-time intrusion detection framework using streaming network traffic data.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2139

- Integration of Explainable Artificial Intelligence (XAI) techniques such as SHAP and LIME to improve decision transparency and interpretability.
- Investigation of hybrid deep learning and LightGBM architectures for enhanced detection of sophisticated cyber-attacks.
- Incorporation of online learning mechanisms to enable continuous adaptation to evolving attack patterns.
- Deployment of the proposed framework in Software Defined Networks (SDN), Internet of Things (IoT), and cloud computing environments.
- Application of federated learning techniques to support privacy-preserving distributed intrusion detection.
- Extension of the framework toward zero-day attack detection and advanced persistent threat (APT) identification.

The proposed Adaptive LightGBM-based Intrusion Detection System provides a scalable, efficient, and highly accurate cybersecurity solution and offers significant potential for future development in intelligent network security applications.

➢ *Declaration of Data Availability*

The dataset used in this study is publicly available from the UNSW-NB15 repository developed by the Australian Centre for Cyber Security (ACCS). All experimental results were generated using the dataset and methodologies described in this paper.

➢ *Conflict of Interest*

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

➢ *Funding Statement*

The authors received no specific funding for this research.

➢ *Author Contributions*

Conceptualization, methodology, implementation, experimentation, analysis, and manuscript preparation were carried out by the authors.

REFERENCES

[1]. W. Stallings, *Network Security Essentials: Applications and Standards*, 6th ed., Pearson Education, 2017.

[2]. A. A. Ghorbani, W. Lu, and M. Tavallaee, *Network Intrusion Detection and Prevention: Concepts and Techniques*, Springer, 2010.

[3]. R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," in *Proceedings of the IEEE Symposium on Security and Privacy*, 2010, pp. 305–316.

[4]. M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Dataset," in *Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 2009, pp. 1–6.

[5]. S. M. Kasongo and Y. Sun, "Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset," *Journal of Big Data*, vol. 7, no. 105, pp. 1–20, 2020.

[6]. M. Ring, D. Landes, D. Wunderlich, S. Scheuring, D. Landes, and A. Hotho, "A Survey of Network-Based Intrusion Deection Data Sets," *Computers & Security*, vol. 86, pp. 147–167, 2019.

[7]. 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 *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[8]. N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set)," in *Military Communications and Information Systems Conference (MilCIS)*, IEEE, 2015, pp. 1–6.

[9]. N. Moustafa and J. Slay, "The Evaluation of Network Anomaly Detection Systems: Statistical Analysis of the UNSW-NB15 Dataset and the Comparison with the KDD99 Dataset," *Information Security Journal: A Global Perspective*, vol. 25, no. 1–3, pp. 18–31, 2016.

[10]. M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm," *IEEE Transactions on Computers*, vol. 65, no. 10, pp. 2986–2998, 2016.

[11]. I. Almomani, B. Al-Kasasbeh, and M. Al-Akhras, "WSN-DS: A Dataset for Intrusion Detection Systems in Wireless Sensor Networks," *Journal of Sensors*, vol. 2016, Article ID 4731953, 2016.

[12]. N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A Deep Learning Approach to Network Intrusion Detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41–50, 2018.

[13]. J. Kim, J. Kim, H. L. T. Thu, and H. Kim, "Long Short-Term Memory Recurrent Neural Network Classifier for Intrusion Detection," *International Conference on Platform Technology and Service*, pp. 1–5, 2016.

[14]. 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," *Advances in Neural Information Processing Systems*, vol. 30, pp. 3146–3154, 2017.

[15]. Y. Zhou, G. Cheng, S. Jiang, and M. Dai, "Building an Efficient Intrusion Detection System Based on Feature Selection and Ensemble Classifier," *Computer Networks*, vol. 174, Article 107247, 2020.

[16]. H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, Springer, 2012.

[17]. M. Ring, S. Wunderlich, D. Grüdl, D. Landes, and A. Hotho, "Flow-Based Benchmark Data Sets for Intrusion Detection," *Proceedings of the 16th European Conference on Cyber Warfare and Security*, pp. 361–369, 2017.