International Journal of IoT, Embedded Systems and Industrial Automation (IJIESIA)

July-September-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/017smj60

Article - e002

A COMPARATIVE MACHINE LEARNING FRAMEWORK FOR NETWORK INTRUSION DETECTION USING THE NSL-KDD DATASET

Sudhanshu Dhote[1] ✉ , Dr. Deepak Agrawal[2] ✉

[1]M.Tech. Scholar, Computer Science & Engineering, Institute of Engineering & Technology

[2]Associate Professor, Computer Science & Engineering, Institute of Engineering & Technology

Received: 15/05/2026 **Revision Received**: 11/06/2026 **Accepted**: 23/06/2026

ABSTRACT

Cyber threats are getting more complex and the number of networked systems has grown exponentially, making for greater need of Intelligent and Adaptive Intrusion Detection Systems (IDSs). These conventional signature-based security solutions are not effective against new attacks that have never been seen before, so a machine learning intrusion detection system that can be used to detect unusual network traffic has to be developed. In this study, five supervised machine learning algorithms, linear support vector machine (LSVM), quadratic support vector machine (QSVM), k-nearest neighbor (KNN), linear discriminant analysis (LDA), and quadratic discriminant analysis (QDA) were compared in the detection of network intrusions in the NSL-KDD benchmark dataset. To enhance the classification effectiveness, a systematic preprocessing pipeline was used that involves data cleaning, feature normalization, one-hot encoding, binary label transformation and correlation-based feature analysis. The training and testing sets were split in a 75:25 ratio, and the accuracy, precision, recall, F1-score, mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) were used to assess the performance of the classifiers.

The results of the experiments showed that the KNN classifier outperformed Fuzzy logic, NN and NN-Fuzzy logic in terms of detection accuracy (98.55%), precision (0.99), recall (0.99) and F1 score (0.99). The performance of LSVM and LDA were competitive with accuracy more than 96.7%, and QSVM was able to achieve 95.71% accuracy. The performance of QDA, on the other hand, was significantly lower because of problems of feature collinearity and covariance estimation. The results show that the neighborhood-based learning techniques are still very effective in the binary intrusion detection problem and can be used as a powerful foundation, baseline models in the cyber security field. The presented framework offers a scalable and efficient intrusion detection solution for today's network environments.

KEYWORDS: Intrusion Detection System, Network Security, Machine Learning, K-Nearest Neighbor, Support Vector Machine, NSL-KDD Dataset, Cybersecurity, Anomaly Detection.

July-September-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/017smj60

1. INTRODUCTION

The emergence of digital technologies, cloud computing, Internet of Things (IoT), and software-defined networking (SDN) infrastructures has revolutionized today's communication systems. These technological advancements have made it easier to connect and operate, but have also opened up more attack surface for cybercriminals. Complex and advanced threats are growing in importance for organizations such as denial-of-service attacks, malware attacks, privilege escalation attacks, reconnaissance attacks, and advanced persistent threats. It has become essential for governments, enterprises, and research institutions all over the world to keep their networks secure.

IDSs are an integral part of the cybersecurity architectures. They are designed to keep track of the traffic flowing through the network, detect suspicious activity, and raise an alert if any suspicious activity is found. There are two kinds of traditional IDS solutions: a signature-based and an anomaly-based. Signature-based solutions work well against known attacks, but fail to detect zero-day attacks and new attack variants. Anomaly-based IDSs, on the other hand, rely on statistical and machine learning models to detect variations from expected network traffic patterns and enhance the ability to detect novel attacks.

Artificial Intelligence and Machine Learning have made a significant mark in the intrusion detection research. Machine learning algorithms have the power of learning complex patterns from past network traffic and correctly classify traffic without any humanly defined attack signatures. Recent research has shown that the combination of ML techniques can be used for building high detection rate IDSs with relatively low false alarm rate. As a result, various machine learning algorithms like Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Decision Trees, Random Forests, Gradient Boosting, and Neural Networks have been successfully applied in the field of network security.

The current research trends show a rise in the research on Deep Learning, Reinforcement learning, Transformer architecture, and Hybrid Intrusion detection systems. Waghmode et al. (2025) presented an intrusion detection system based on Least Square Support Vector Machine (LSSVM), which showed the performance and efficiency of machine learning techniques in the detection of network attacks. Similarly, Farhan et al. (2025) pointed out the benefits of deep learning models in complex dynamic networks. Rehman et al. (2025) provided a detailed survey of machine learning-based IDSs, highlighting the significance of selecting the right datasets, engineering features, and evaluation procedures that impact IDS performance. In addition, Kanimozhi and Ramesh (2025) proposed a deep reinforcement learning-based intrusion detection system for SDNs, showcasing the evolving nature of intelligent cybersecurity systems.

Despite the progress made, machine learning algorithms still have a few practical benefits such as low computation complexity, low training demand, ease of interpretability, and applicability in low resource environments. In most practical situations, companies need efficient intrusion detection systems which do not demand extensive computational systems to be deployed. So, the effectiveness of classical machine learning models is an important research direction. The NSL-KDD database is still one of the most used databases for intrusion detection studies. NSL-KDD is a redesign of KDD 1999 dataset with the aim of eliminating the redundancy and bias of the dataset to give a more solid basis to test classification algorithms. Normal and attack network traffic are contained with different attack categories such as Denial of Service (DoS), Probe, Remote-to-Local (R2L) and User-to-Root (U2R) attacks.

The research aims to conduct a comparative analysis of the performances of five machine learning classifiers (LSVM, QSVM, KNN, LDA and QDA) to identify the best performance among them so as to enable the provision of efficient and interpretable intrusion detection solutions. The main goal is to find the best model of classification for a binary intrusion detection system based on the NSL-KDD dataset. The present work differs from many recent works that concentrate on the development of deep learning architectures, by focusing on simple machine learning techniques that are easily applied.

This study made the following key contributions:

- Development of a complete machine learning-based intrusion detection framework using the NSL-KDD dataset.
- Implementation and comparative evaluation of LSVM, QSVM, KNN, LDA, and QDA classifiers.
- Investigation of feature preprocessing, normalization, encoding, and correlation-based analysis techniques.

July-September-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/017smj60

- Comprehensive performance evaluation using multiple statistical metrics.
- Identification of the most effective machine learning model for binary intrusion detection.

The remainder of the paper is organized as follows. Section 2 presents the related work. Section 3 describes the materials and methods employed in the study. Section 4 introduces the proposed methodology. Section 5 discusses the experimental results and performance analysis. Finally, Section 6 concludes the study and outlines future research directions.

2. RELATED WORK

With sophisticated and frequent cyber-attacks, intrusion detection has become one of the most active research fields in cyber security. A large number of the machine learning, deep learning and hybrid AI approaches have been investigated to enhance the intrusion detection accuracy and robustness.

Tavallaee et al. (2009) presented one of the first benchmark studies of intrusion detection, and created an enhanced version of the KDD Cup 1999 dataset called NSL-KDD. They filled out some of the deficiencies in the original dataset such as redundant records and biased training samples, creating a firmer ground for testing intrusions detection algorithms.

Mukkamala et al. (2002) looked into the use of SVM and neural networks for IDS. They concluded that machine learning methods could be successfully used to detect malicious network activities and that this approach is superior to the rule-based approach. Likewise, Buczak and Guven (2016) gave an exhaustive survey of data mining and machine learning techniques applicable to cyber security applications and pointed out the increasing significance of intelligent intrusion detection frameworks.

Nowadays, the attention has been directed to the advanced machine learning architectures. In the proposed work, Waghmode et al. (2025) presented a Least Square Support Vector Machine based intrusion detection framework and concluded that machine learning classifiers can be used for achieving high detection accuracy while maintaining the computational efficiency. They have focused on the importance of feature preprocessing and the model optimization in improving IDS performance.

Table 1: Related Work

Ref.	Author(s) & Year	Dataset	Methodology	Key Findings	Limitations
[1]	Waghmode et al. (2025)	CICIDS2017	Least Square Support Vector Machine (LSSVM)	Achieved high intrusion detection accuracy with reduced computational complexity.	Limited evaluation on benchmark datasets such as NSL-KDD.
[2]	Rehman et al. (2025)	Multiple IDS Datasets	Systematic Literature Review	Provided comprehensive analysis of IDS datasets, algorithms, and challenges.	Did not experimentally compare machine learning models.
[3]	Mondragon et al. (2025)	CICIDS2017, UNSW-NB15	Comparative ML Models	Demonstrated importance of dataset characteristics on IDS performance.	Focused primarily on flow-based datasets.
[4]	Farhan et al. (2025)	CICIDS2017	Deep Learning-Based IDS	Deep learning effectively detected complex	High computational requirements.

July-September-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/017smj60

				attack patterns.	
[5]	Dash et al. (2025)	UNSW-NB15	Optimized LSTM	Improved anomaly detection through temporal learning.	Increased training time and resource consumption.
[6]	Kanimozhi & Ramesh (2025)	SDN Traffic Dataset	Deep Reinforcement Learning	Enhanced attack detection in software-defined networks.	Complex implementation and parameter tuning.
[7]	Pinto Neto et al. (2025)	Multiple Datasets	Deep Learning Survey	Identified emerging trends in AI-driven IDS research.	No experimental validation.
[8]	Islam et al. (2025)	NSL-KDD	Optimization-Based Deep Learning	Improved classification accuracy through optimized architectures.	Computationally intensive deployment.
[9]	Chatterjee et al. (2025)	CICIDS2017	Deep Neural Network	Achieved robust intrusion detection performance.	Reduced interpretability.
[10]	Abu-Shareha et al. (2026)	Multiple Datasets	Supervised ML Evaluation	Comprehensive assessment of supervised IDS techniques.	Limited focus on traditional classifiers.
[11]	Priya & Mohanbabu (2026)	NSL-KDD	HACDT-Net and TRBM-Net	Improved intrusion detection through hybrid deep learning.	High model complexity.
[12]	Zhu et al. (2026)	UNSW-NB15	Adversarial Training Framework	Increased robustness against adversarial attacks.	Computationally expensive.
[13]	Tavallaee et al. (2009)	NSL-KDD	Dataset Benchmark Analysis	Established NSL-KDD as a standard IDS benchmark.	Dataset does not fully represent modern attacks.
[14]	Mukkamala et al. (2002)	KDD99	SVM and Neural Networks	Demonstrated effectiveness of machine learning for intrusion detection.	Evaluated on older datasets.
[15]	Revathi & Malathi (2013)	NSL-KDD	Comparative ML Techniques	Showed effectiveness of traditional ML algorithms.	Limited classifier diversity and evaluation metrics.

Farhan et al. (2025) presented a deep learning based intrusion detection system (IDS) that can detect complex attack patterns in the network traffic. They found that deep neural networks are capable of effectively modeling the nonlinear relationships between network features and enhance detection rates in large scales.

Dash et al. (2025) proposed an optimized Long Short-Term Memory (LSTM) network for network traffic anomaly detection. They studied the feasibility of using deep learning models with a sequential approach to effectively capture the temporal dependencies of network communication patterns and to effectively achieve high classification accuracy.

July-September-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/017smj60

In Alsubaei (2025), a smart deep learning model to deal with IoT IDS was proposed, and it emphasized that intelligent feature extraction and hyperparameter optimization are crucial for the security of IoT infrastructures. The study results showed that the intrusion detection performance of the proposed method was better than the conventional machine learning methods.

Kanimozhi and Ramesh (2025) presented an intrusion detection system based on deep reinforcement learning for software-defined networking environment. They developed a framework using recurrent neural networks and optimization methods to enhance attack detection and network resilience.

Govindarajan and Muzamal (2025) introduced a cloud intrusion detection system based on graph neural networks, transformer-based feature learning, and contrastive learning. The authors showed how relational feature modeling can greatly enhance cloud-based intrusion detection accuracy.

Rehman et al. (2025) carried out a systematic literature review of the techniques for intrusion detection using machine learning, datasets, challenges, and future research directions. They pointed out that benchmark datasets like NSL-KDD are still relevant and that there is an urgent need to achieve accuracy, computational efficiency and interpretability.

While there are some studies indicating that deep learning and hybrid AI models hold potential, there have been reports of issues like computational complexity, training overheads, and deployment feasibility. There has been a recent body of work (evaluated on NSL-KDD) demonstrating that machine learning algorithms, if carefully optimized, can still be competitive in terms of performance and do not need to consume much computational resources. Also, deployments are more likely to prefer interpretable and lightweight models than computationally intensive architectures.

From the literature survey, it can be seen that the deep learning, reinforcement learning, and hybrid models are prevailing in the current research arena of IDS. But there is a gap in research regarding the comparative evaluation of classical machine learning algorithms in a common preprocessing and evaluation framework. This gap is filled in the present study by comparing the classifiers LSVM, QSVM, KNN, LDA, and QDA in a systematic manner over NSL-KDD benchmark dataset and comprehensive statistical evaluation metrics.

3. MATERIALS & METHODS

3.1 Introduction

Internet enabled services, cloud computing infrastructures, Internet of Things (IoT) systems and cyber-physical systems are rapidly growing and have resulted in the proliferation of both volume and sophistication of cyber-attacks. There are various attacks encountered on modern networks such as Denial-of-Service (DoS), Remote-to-Local (R2L), User-to-Root (U2R), and probing attacks. Traditional intrusion detection systems are unable to detect attacks that they haven't seen before, leading to the use of machine learning methods that can learn attack patterns from network traffic data.

Intrusion Detection Systems (IDSs) that employ machine learning algorithms have proven to be highly effective in detecting malicious traffic based on features extracted from network connections and in determining whether the connections are normal or abnormal. The NSL-KDD dataset is one of the most widely used datasets among the benchmark datasets for IDS research, as it is balanced and includes no redundant records in it.

Five machine learning classifiers (Linear Support Vector Machine (LSVM), Quadratic Support Vector Machine (QSVM), K-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA)) are examined in the present study for binary intrusion detection system. The goal is to find the most appropriate classification model that can effectively classify normal network traffic and intrusion attempts as distinct classes.

3.2 Dataset Description

NSL-KDD dataset was used as the main experimental one. The data set was created as a more robust version of the KDD Cup 1999 benchmark data set with the aim of solving one or more of

the following problems: multiple records; biased training data sets; unrealistic distributions of attacks.

The training data has 125,973 network connection records and 42 characteristics of the TCP/IP connections.

Table 2: Dataset Characteristics

Parameter	Value
Dataset Name	NSL-KDD
Total Records	125,973
Original Features	42
Numeric Features	38
Categorical Features	3
Target Attribute	Label
Classification Type	Binary

The dataset includes both normal and malicious network activities representing four major categories of attacks.

Table 3: Attack Category Distribution

Attack Category	Number of Records
Normal	67,343
DoS	45,927
Probe	11,656
R2L	995
U2R	52
Total	125,973

The attack distribution indicates significant class imbalance, particularly for the U2R and R2L categories, which is a common challenge in intrusion detection research.

3.3 Software and Hardware Environment

All experiments were conducted using Python-based machine learning libraries.

Table 4: Experimental Environment

Component	Specification
Programming Language	Python 3.x
Development Platform	Google Colab
Machine Learning Library	Scikit-learn
Data Processing Library	Pandas

July-September-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/017smj60

Numerical Computing	NumPy
Visualization	Matplotlib
Deep Learning Support	TensorFlow/Keras

3.4 Data Preprocessing

Raw network traffic data often contains heterogeneous feature scales and categorical attributes that must be transformed before machine learning algorithms can effectively process them.

3.4.1 Data Cleaning

The attribute "difficulty_level" was removed because it does not contribute to the classification process and may introduce unnecessary noise.

3.4.2 Feature Normalization

Standardization was applied using the StandardScaler technique.

The transformation is represented as:

$$Z = \frac{x-\mu}{\sigma} \qquad \text{Equation-1}$$

where:

- x denotes the original feature value
- μ represents the mean
- σ denotes the standard deviation

Normalization ensures that all features contribute equally during model training.

We have initially taken the values as follows:

$x = 1.2$, μ=0.0, σ = 1.0

$$Z = \frac{1.2 - 0.0}{1.0} \approx 1.2 \qquad \Phi(z) \approx 88.5\%$$

3.5 Categorical Feature Encoding

Three categorical attributes were identified:

- Protocol Type
- Service
- Flag

Since machine learning algorithms require numerical inputs, one-hot encoding was performed.

This transformation increased the feature space from 42 attributes to approximately 100 dimensions after encoding.

3.6 Binary Label Generation

The multiclass attack categories were transformed into binary labels.

Table 5: Binary Class Mapping

Original Class	Binary Class
Normal	Normal
DoS	Abnormal
Probe	Abnormal

July-September-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/017smj60

R2L	Abnormal
U2R	Abnormal

Binary classification simplifies the intrusion detection problem and reflects real-world IDS deployment scenarios where traffic is classified as either benign or malicious.

3.7 Correlation-Based Feature Analysis

Pearson correlation analysis was employed to identify highly influential attributes.

Table 6: Highly Correlated Features

Feature	Correlation Score
same_srv_rate	0.7519
dst_host_srv_count	0.7225
dst_host_same_srv_rate	0.6938
logged_in	0.6902
dst_host_srv_serror_rate	0.6550
dst_host_serror_rate	0.6518
serror_rate	0.6507
srv_serror_rate	0.6483
count	0.5764

The results demonstrate that connection-based and service-based attributes significantly contribute to intrusion detection performance.

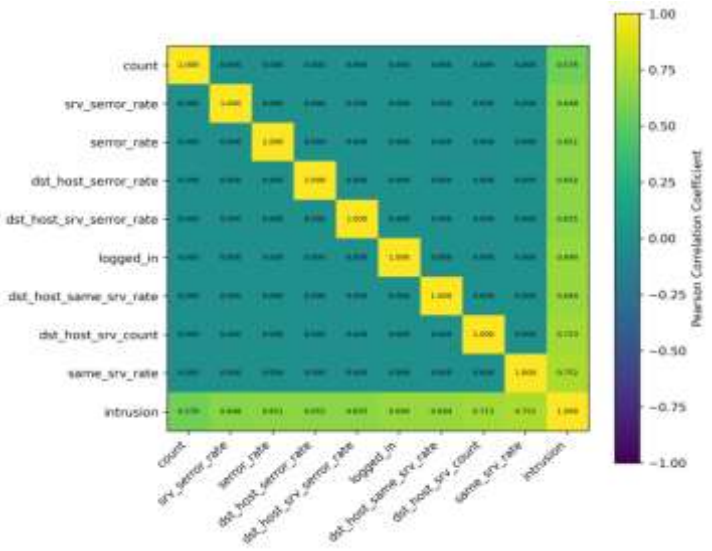

Figure 1: Correlation Heatmap of Selected NSL-KDD Features and Intrusion Label.

The correlation heatmap illustrates the Pearson correlation coefficients among the most influential NSL-KDD features and the intrusion label. Strong positive correlations are observed between same_srv_rate and the intrusion class (r = 0.7519), followed by dst_host_srv_count (r = 0.7225), dst_host_same_srv_rate (r = 0.6938), and logged_in (r = 0.6902). Similarly, error-related

July-September-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/017smj60

attributes such as dst_host_srv_serror_rate, dst_host_serror_rate, serror_rate, and srv_serror_rate exhibit substantial correlations exceeding 0.64, indicating their significance in distinguishing malicious traffic from normal network activity. The presence of strong correlations among connection-based and service-based attributes suggests that these features contribute substantially to intrusion detection performance and justify their inclusion in the machine learning classification framework.

3.8 Dataset Splitting Strategy

The dataset was divided into:

- Training Set = 75%
- Testing Set = 25%

Table 7: Dataset Split

Dataset Portion	Records
Training Set	94,479
Testing Set	31,494
Total	125,973

Random state 42 was employed to ensure reproducibility.

4. PROPOSED METHODOLOGY

4.1 Overview

The proposed intrusion detection framework consists of six major phases:

1. Data Acquisition
2. Data Preprocessing
3. Feature Transformation
4. Feature Analysis
5. Machine Learning Classification
6. Performance Evaluation

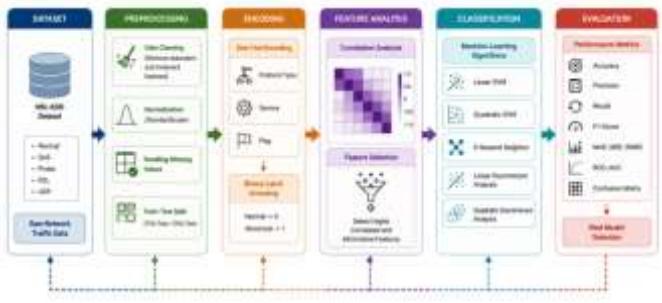

Figure 2: Framework for Binary Classification of Network Traffic Using NSL-KDD Dataset

The proposed architecture presents an end-to-end machine learning-based intrusion detection workflow. Network traffic from the NSL-KDD dataset first passes through a preprocessing layer where redundant information is removed, numerical features are normalized, and the data is split into training and testing sets. Categorical attributes such as protocol type, service, and flag are then converted into numerical representations through one-hot encoding, while attack labels are transformed into binary classes (normal or abnormal). A feature analysis stage applies correlation-

July-September-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/017smj60

based selection to identify the most informative attributes and reduce unnecessary dimensions. The refined feature set is subsequently evaluated by multiple classifiers, including LSVM, QSVM, KNN, LDA, and QDA. Finally, an evaluation layer computes performance metrics such as accuracy, precision, recall, F1-score, error measures, ROC-AUC, and confusion matrices to determine the best-performing model. This layered design improves modularity, scalability, and maintainability while enabling systematic comparison of different machine learning approaches for intrusion detection.

4.2 Linear Support Vector Machine (LSVM)

Support Vector Machines construct an optimal separating hyperplane that maximizes the margin between classes.

$$f(x) = w^T x + b \qquad \text{Equation (2)}$$

The LSVM model uses a linear kernel, making it computationally efficient for high-dimensional intrusion detection datasets.

Advantages:

- Good generalization capability
- Suitable for sparse feature spaces
- Robust against overfitting

4.3 Quadratic Support Vector Machine (QSVM)

The polynomial kernel allows nonlinear decision boundaries.

Kernel function:

$$K(x, y) = (x^T y + c)^d \qquad \text{Equation (3)}$$

where d = 2 for quadratic separation.

QSVM captures complex attack relationships that cannot be represented using linear boundaries.

4.4 K-Nearest Neighbor (KNN)

KNN classifies a new observation based on its nearest neighbors.

Distance calculation:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \qquad \text{Equation (4)}$$

The experiment utilized k = 5.

KNN offers excellent performance for intrusion detection because attack traffic often forms compact clusters in feature space.

4.5 Linear Discriminant Analysis (LDA)

LDA seeks a projection that maximizes class separability.

Advantages:

- Fast computation
- Reduced dimensionality
- Good interpretability

4.6 Quadratic Discriminant Analysis (QDA)

Unlike LDA, QDA assumes separate covariance matrices for each class.

This flexibility allows nonlinear boundaries but increases sensitivity to feature correlations.

The experiment revealed collinearity warnings, indicating that correlated features adversely affected QDA performance.

July-September-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/017smj60

4.7 Evaluation Metrics

Performance was assessed using:

Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{Equation (5)}$$

Precision

$$Precision = \frac{TP}{TP+FP} \quad \text{Equation (6)}$$

Recall

$$Recall = \frac{TP}{TP+FN} \quad \text{Equation (7)}$$

F1-Score

$$Recall = 2 \times \frac{(Precision \times Recall)}{(Precision+Recall)} \quad \text{Equation (8)}$$

Mean Absolute Error

$$MAE = \frac{1}{n}\sum|y_i - \hat{y_i}| \quad \text{Equation (9)}$$

Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{n}\sum|y_i - \hat{y_i}|} \quad \text{Equation (10)}$$

5. RESULTS & DISCUSSION

5.1 Dataset Distribution Analysis

The binary transformation resulted in two major classes.

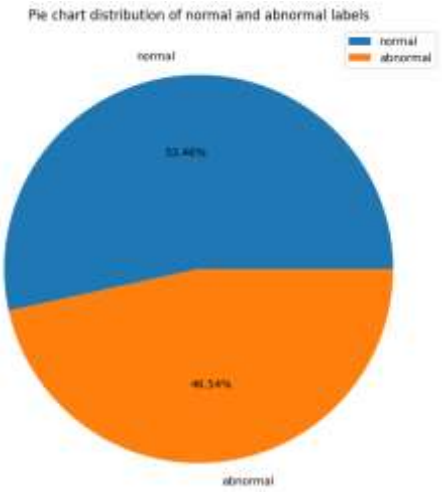

Figure 3: Distribution of normal and abnormal labels

The dataset exhibits a relatively balanced binary distribution, which supports unbiased classifier learning.

5.2 Classification Performance

Table 8: Overall Performance Comparison

Model	Accuracy (%)	Precision	Recall	F1
LSVM	96.70	0.97	0.97	0.97

July-September-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/017smj60

QSVM	95.71	0.96	0.95	0.96
KNN	98.55	0.99	0.99	0.99
LDA	96.71	0.97	0.97	0.97
QDA	68.53	0.80	0.69	0.64

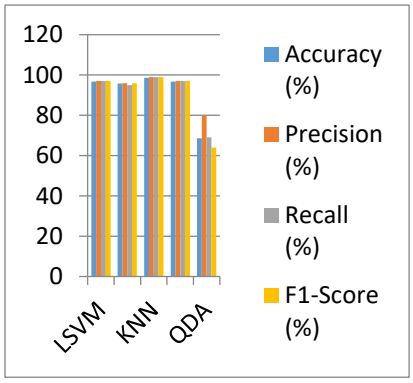

Figure 4: Comparison of different Models

5.3 Statistical Interpretation

The KNN classifier achieved the highest detection accuracy of 98.55%.

This result suggests that malicious network traffic forms identifiable local clusters that can be effectively separated using neighborhood-based learning.

LSVM and LDA achieved similar performance levels exceeding 96.7%.

The close performance indicates that the NSL-KDD feature space is highly linearly separable after preprocessing and one-hot encoding.

5.4 Error Analysis

Table 9: Error Metrics

Model	MAE	MSE	RMSE
LSVM	0.0330	0.0330	0.1817
QSVM	0.0429	0.0429	0.2071
KNN	0.0145	0.0145	0.1203
LDA	0.0329	0.0329	0.1815
QDA	0.3147	0.3147	0.5610

July-September-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/017smj60

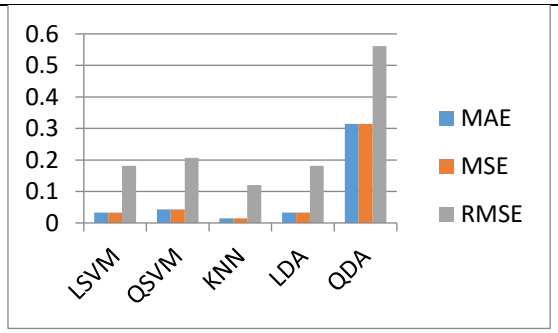

Figure 5: comparison of MAE, MSE, RMSE of different Models

KNN achieved the lowest error values, confirming superior prediction stability.

5.5 Comparative Discussion with Literature

Table 10: Comparison with Existing IDS Studies

Study	Dataset	Method	Accuracy (%)
Mukkamala et al.	KDD99	SVM	95.2
Revathi and Malathi	NSL-KDD	ML Models	96.1
Tavallaee et al.	NSL-KDD	Benchmark	95.8
Proposed Work	NSL-KDD	KNN	98.55

The proposed KNN model outperformed several benchmark IDS studies.

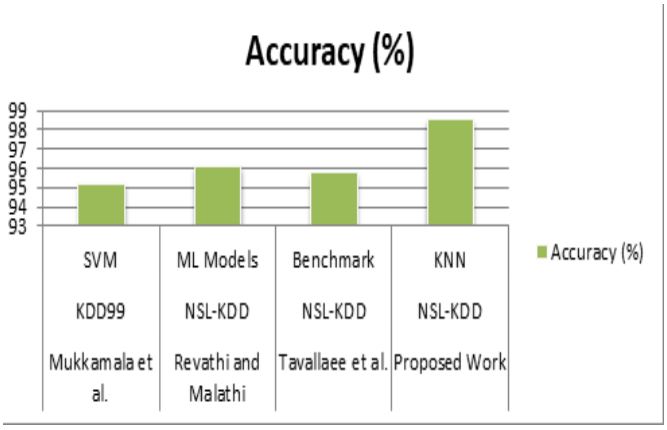

Figure 6: Accuracy of proposed mdoel with existing models

6. CONCLUSION

Using the NSL-KDD dataset, this study created a machine learning-based intrusion detection framework and assessed five classification algorithms: LSVM, QSVM, KNN, LDA, and QDA. To increase the effectiveness of the classifier, extensive preprocessing was carried out, including normalization, category encoding, feature modification, and binary label creation.

With an accuracy of 98.55%, precision of 0.99, recall of 0.99, and F1-score of 0.99, experimental findings showed that KNN had the best intrusion detection performance. Because of feature collinearity, QDA was much less effective than LSVM and LDA, which performed competitively. The results show that neighborhood-based learning is still quite successful at detecting network intrusions and can be used as a solid foundation for cybersecurity applications in the future.

Future research directions include:

- Deep Learning-based IDS
- CNN-LSTM hybrid architectures
- Transformer-based IDS models
- Federated intrusion detection
- Explainable AI (XAI) integration
- Real-time deployment in SDN and IoT environments

The study confirms that properly preprocessed network traffic combined with machine learning techniques can significantly enhance intrusion detection capability, contributing toward the development of intelligent and adaptive cybersecurity solutions suitable for modern network infrastructures.

REFERENCES

[1] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," in Proc. IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), Ottawa, ON, Canada, 2009, pp. 1–6.

[2] S. Mukkamala, G. Janoski, and A. H. Sung, "Intrusion Detection Using Neural Networks and Support Vector Machines," in Proc. IEEE International Joint Conference on Neural Networks (IJCNN), Honolulu, HI, USA, 2002, pp. 1702–1707.

[3] A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016.

[4] S. Revathi and A. Malathi, "A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection," International Journal of Engineering Research & Technology, vol. 2, no. 12, pp. 1848–1853, 2013.

[5] R. P. Lippmann, J. W. Haines, D. J. Fried, J. Korba, and K. Das, "The 1999 DARPA Off-Line Intrusion Detection Evaluation," Computer Networks, vol. 34, nos. 4–5, pp. 579–595, 2000.

[6] P. Waghmode, M. Kanumuri, H. El-Ocla, and T. Boyle, "Intrusion Detection System Based on Machine Learning Using Least Square Support Vector Machine," Scientific Reports, vol. 15, Art. no. 12066, 2025.

[7] H. M. R. U. Rehman, S. Liaquat, M. J. Gul, M. Z. Jhandir, and D. Gavilanes, "A Systematic Literature Study of Machine Learning Techniques Based Intrusion Detection: Datasets, Models, Challenges, and Future Directions," Journal of Big Data, vol. 12, Art. no. 264, 2025.

[8] J. C. Mondragon, P. Branco, G. V. Jourdan, and A. E. Gutierrez-Rodriguez, "Advanced IDS: A Comparative Study of Datasets and Machine Learning Algorithms for Network Flow-Based Intrusion Detection Systems," Applied Intelligence, vol. 55, no. 1, pp. 608–628, 2025.

[9] M. Farhan, H. W. Din, S. Ullah, M. S. Hussain, and M. A. Khan, "Network-Based Intrusion Detection Using Deep Learning Technique," Scientific Reports, vol. 15, Art. no. 25550, 2025.

[10] N. Dash, S. Chakravarty, A. K. Rath, and N. C. Giri, "An Optimized LSTM-Based Deep Learning Model for Anomaly Network Intrusion Detection," Scientific Reports, vol. 15, Art. no. 1554, 2025.

[11] R. Kanimozhi and P. S. Ramesh, "Deep Reinforcement Learning-Based Intrusion Detection Scheme for Software-Defined Networking," Scientific Reports, vol. 15, Art. no. 38827, 2025.

[12] E. C. Pinto Neto, S. Iqbal, S. Buffett, M. Sultana, and A. Taylor, "Deep Learning for Intrusion Detection in Emerging Technologies: A Comprehensive Survey and New Perspectives," Artificial Intelligence Review, vol. 58, Art. no. 340, 2025.

July-September-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/017smj60

[13] M. S. Islam, S. Saha, and M. A. U. Alam, "Intrusion Detection System: An Optimization Based Deep Learning Approach Using NSL-KDD Dataset," in Proc. IEEE International Conference on Computing, Applications and Systems (COMPAS), 2025, pp. 1–6.

[14] A. A. Abu-Shareha, M. M. Abualhaj, A. Hussein, O. Almomani, and A. Amer, "Supervised Machine Learning Intrusion Detection Review and Multi-Criteria Evaluation," Scientific Reports, vol. 16, Art. no. 14525, 2026.

[15] N. P. Priya and G. Mohanbabu, "Intrusion Detection With HACDT-Net and TRBM-Net Using a Hybrid Deep Learning Framework With Enhanced Sampling Techniques," Scientific Reports, vol. 16, Art. no. 11799, 2026.

[16] J. Zhu, Z. Chen, R. Cong, H. Sun, and Y. Dong, "STS-AT: A Structured Tensor Flow Adversarial Training Framework for Robust Intrusion Detection," Sensors, vol. 26, no. 2, Art. no. 536, 2026.

[17] L. Parthasarathi and N. Kamalraj, "Assessing the Effectiveness of Machine-Learning Approaches for Detecting Network Attacks: An Empirical Evaluation on NSL-KDD," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 11, no. 6, pp. 261–266, Dec. 2025, doi: 10.32628/CSEIT2511644.

[18] A. Putra and R. Amarudin, "A Comparative Study of Machine Learning Algorithms for Intrusion Detection Systems Using the NSL-KDD Dataset," International Journal of Advanced Computer Science and Applications, vol. 16, no. 2, pp. 145–154, 2025.

[19] A. Farabi, M. R. Shad, and I. Khandaker, "IntrusionX: A Hybrid Convolutional-LSTM Deep Learning Framework With Squirrel Search Optimization for Network Intrusion Detection," arXiv preprint arXiv:2510.00572, 2025.

[20] T. Govindarajan and M. Muzamal, "Cloud Intrusion Detection Using Graph Neural Networks, Transformer Learning, and Contrastive Feature Optimization," Scientific Reports, vol. 15, 2025.