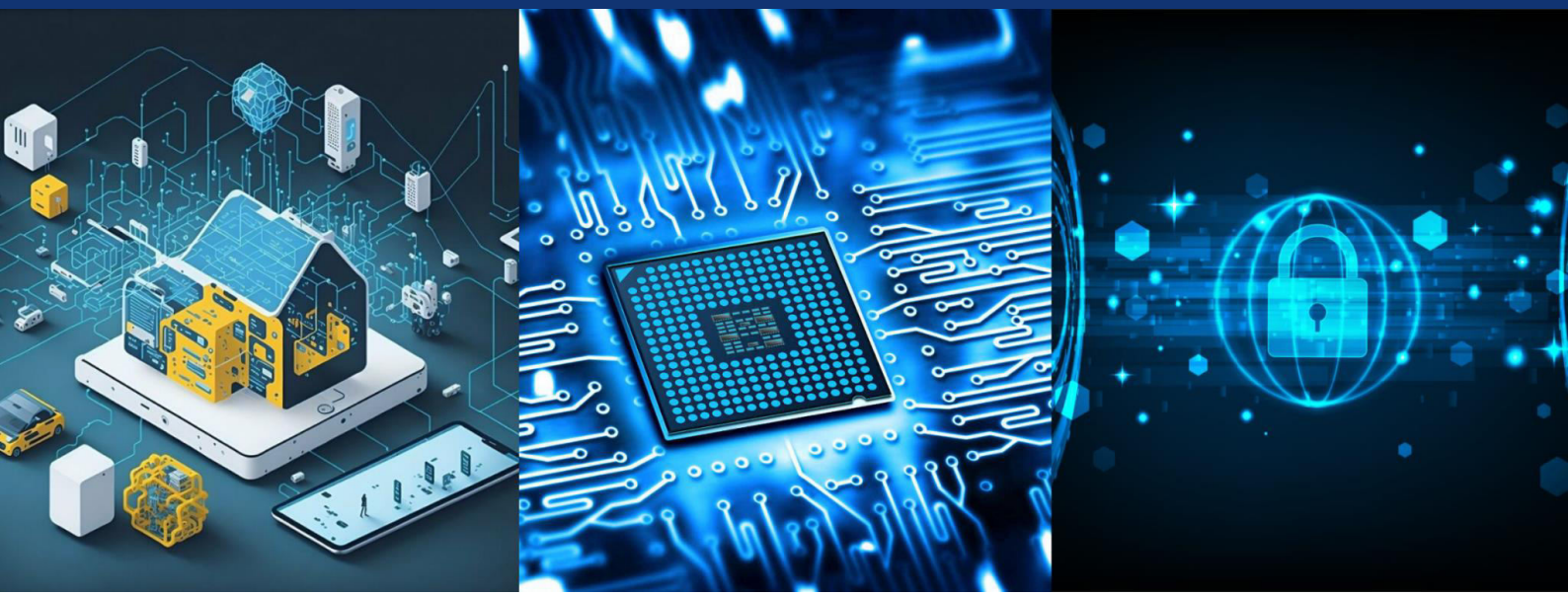
IJIRCCE

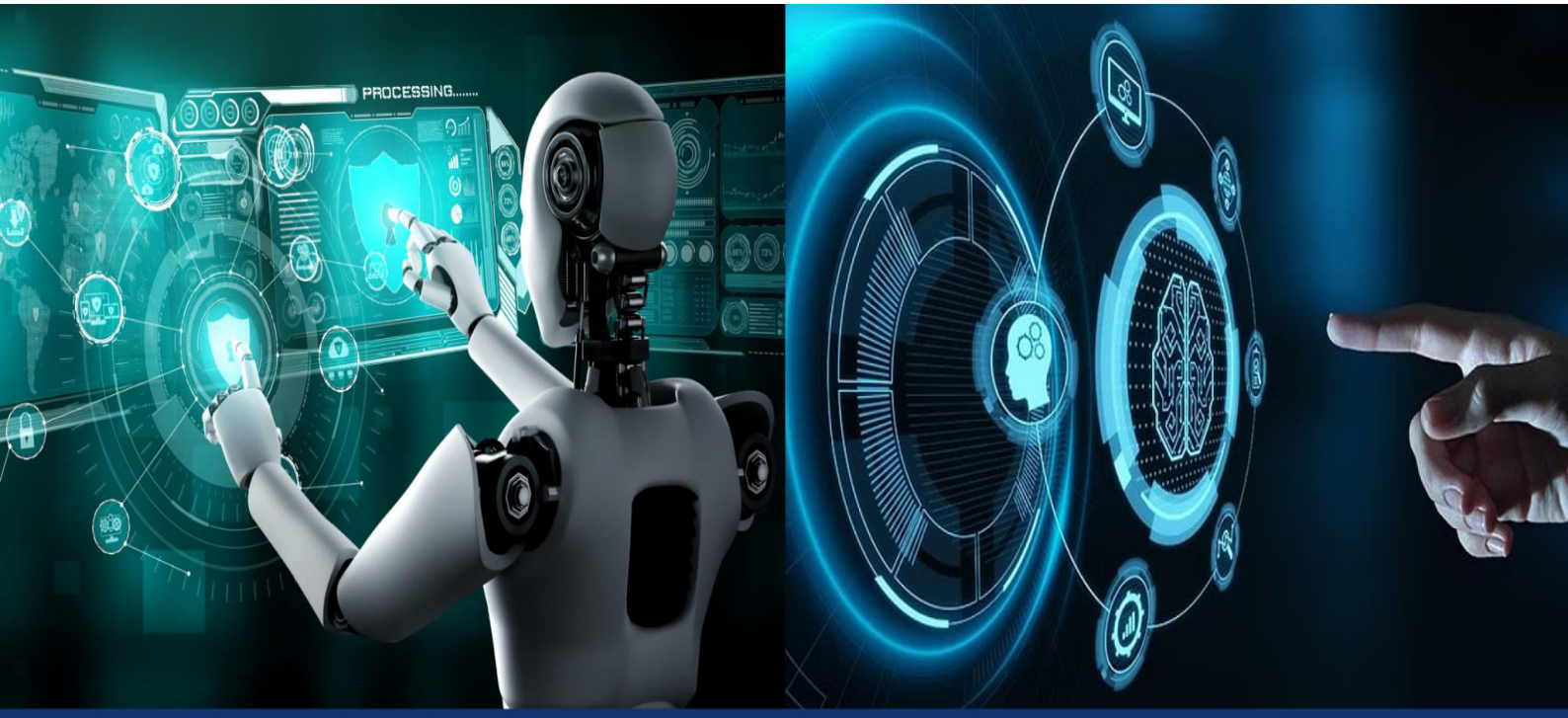
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.927

Volume 14, Issue 7, July 2026

© 2026 IJIRCCE | Volume 14, Issue 7, July 2026 DOI: 10.15680/IJIRCCE.2026.1407029

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A Machine Learning - Based Intrusion Detection and Forecasting System

Gannabathula Sai Kanth Pushkar, Dr. V. Uma Rani

Post-Graduate Student, Department of Computer Science Engineering, Computer Networks and Information Security, Jawaharlal Nehru Technological University, Hyderabad, India

Professor, Department of Computer Science Engineering, Jawaharlal Nehru Technological University, Hyderabad, India

ABSTRACT: The increasing sophistication of cyberattacks demands intelligent systems capable of both detecting ongoing intrusions and anticipating future threats. This paper presents a Machine Learning-Based Intrusion Detection and Forecasting System that integrates intrusion detection with attack trend prediction using the UNSW-NB15 dataset. The proposed framework applies data preprocessing techniques, including data cleaning, label encoding, normalization, and train-test splitting, before training Random Forest, Gradient Boosting, and Deep Neural Network (DNN) models for intrusion detection. To enable proactive cybersecurity, Linear Regression and Long Short-Term Memory (LSTM) models are employed to forecast future attack trends based on historical attack data. Experimental results demonstrate that the proposed detection models achieve high classification performance, while the forecasting models accurately predict attack patterns with low MAE and RMSE values. An interactive Streamlit dashboard is developed to visualize detection results, model performance, and forecasted cyber-attack trends, providing an effective decision-support tool for network security administrators.

KEYWORDS: Intrusion Detection System, Machine Learning, Deep Neural Network, Random Forest, LSTM, Cybersecurity, Attack Forecasting, UNSW-NB15, Network Security, Streamlit.

I. INTRODUCTION

The rapid growth of internet-based services has significantly increased the volume and sophistication of cyberattacks, posing serious threats to network security and data integrity. Intrusion Detection Systems (IDS) play a vital role in identifying malicious activities by monitoring network traffic and detecting abnormal behavior. However, conventional signature-based and rule-based detection techniques are often ineffective against zero-day attacks and evolving cyber threats due to their dependence on predefined attack patterns.

Recent advancements in Machine Learning (ML) and Deep Learning (DL) have improved intrusion detection by enabling systems to learn complex attack characteristics from historical network traffic. Despite these improvements, most existing approaches focus only on attack detection and lack the capability to predict future attack trends, limiting proactive security planning.

This paper proposes a Machine Learning-Based Intrusion Detection and Forecasting System using the UNSW-NB15 dataset. Random Forest, Gradient Boosting, and Deep Neural Network (DNN) models are employed for intrusion detection, while Linear Regression and Long Short-Term Memory (LSTM) models are utilized to forecast future attack trends. An interactive Streamlit dashboard is developed to visualize detection results, forecasting outputs, and model performance, providing an effective decision-support framework for intelligent network security management.

II. RELATED WORK

Machine learning and deep learning techniques have significantly improved the performance of Intrusion Detection Systems (IDS) by enabling automated analysis of network traffic and accurate identification of cyber threats. Venkata Siva et al. proposed a cyber attack detection and prediction framework that demonstrated the effectiveness of machine

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

learning algorithms in identifying malicious network activities. However, the study primarily focused on attack detection with limited emphasis on long-term attack forecasting.

Buczak and Guven presented a comprehensive survey of data mining and machine learning approaches for cybersecurity intrusion detection. Their work highlighted the advantages of supervised learning techniques in improving detection accuracy while emphasizing challenges such as class imbalance, feature selection, and evolving attack patterns. Similarly, Ferrag et al. reviewed artificial intelligence-based cybersecurity solutions and reported that deep learning models are capable of identifying complex attack behaviors more effectively than conventional techniques.

Alom et al. demonstrated the applicability of Recurrent Neural Networks (RNNs) for intrusion detection by exploiting temporal relationships in network traffic. Liu et al. further surveyed convolutional and recurrent neural network architectures for cybersecurity applications, concluding that deep learning models provide superior performance for large-scale and complex datasets.

Although these studies achieved promising intrusion detection results, most existing approaches concentrate only on identifying ongoing attacks. Limited attention has been given to forecasting future attack trends, which is essential for proactive cybersecurity planning. To address this limitation, the proposed work integrates machine learning-based intrusion detection using Random Forest, Gradient Boosting, and Deep Neural Network (DNN) models with attack forecasting using Linear Regression and Long Short-Term Memory (LSTM). The developed Streamlit dashboard further enhances the system by providing interactive visualization of detection results, model performance, and future attack predictions within a unified framework.

III. PROPOSED ALGORITHM

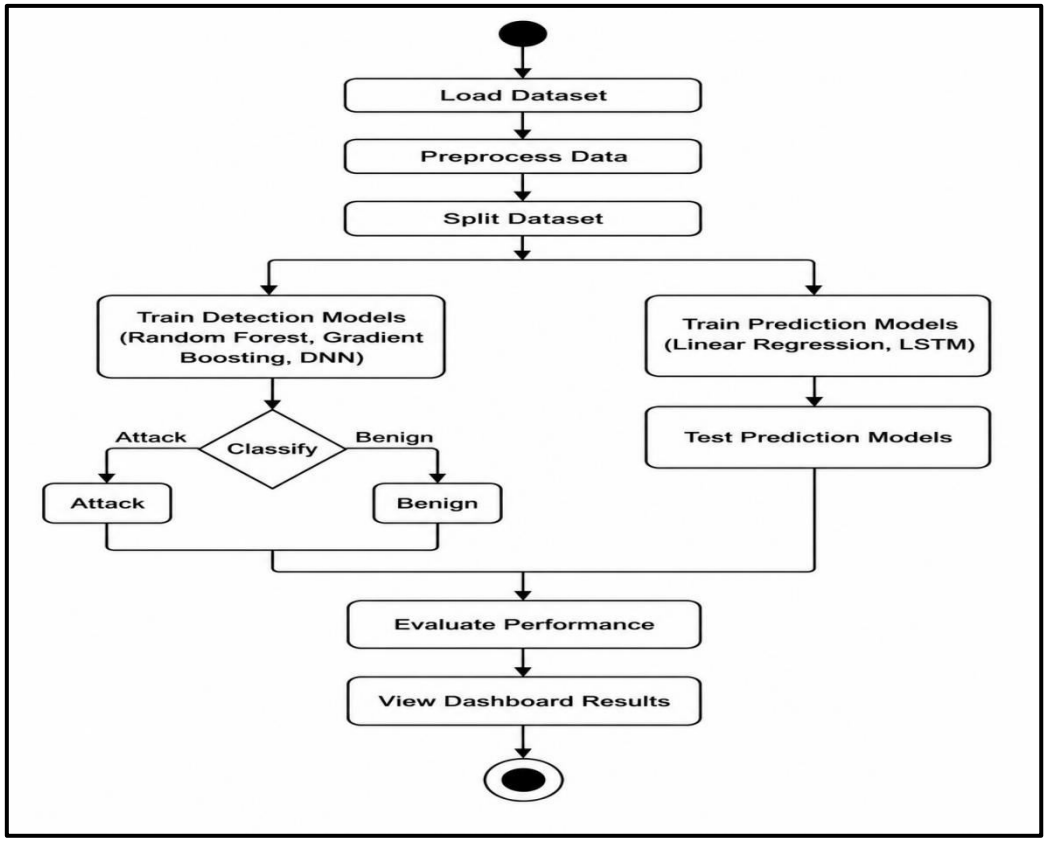


Figure 1: Proposed Architecture

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A. ***Design Considerations:***

- The proposed system utilizes the UNSW-NB15 dataset containing 257,673 network traffic records, including 93,000 normal and 164,673 attack samples.
- Data preprocessing involves handling missing values, removing duplicate records, encoding categorical features using Label Encoding, and normalizing numerical attributes using Min-Max Scaling.
- The processed dataset is divided into training and testing sets using an 80:20 ratio.
- Random Forest, Gradient Boosting, and Deep Neural Network (DNN) models are employed for intrusion detection.
- Linear Regression and Long Short-Term Memory (LSTM) models are utilized to forecast future cyber attack trends based on historical attack data.
- Detection models are evaluated using Accuracy, Precision, Recall, F1-Score, and Confusion Matrix, while forecasting models are evaluated using MAE and RMSE.
- An interactive Streamlit dashboard is developed to visualize intrusion detection results, forecasting outputs, and comparative model performance.

B. ***Description of the Proposed Algorithm:***

The objective of the proposed algorithm is to accurately detect malicious network traffic and forecast future cyber attack trends by integrating machine learning, deep learning, and time-series forecasting techniques. The proposed methodology consists of four major steps.

Step 1: Data Preprocessing

The UNSW-NB15 dataset is loaded and preprocessed by removing duplicate records, handling missing values, encoding categorical features using Label Encoding, and normalizing numerical attributes using Min-Max Scaling. The target class is represented as:

Classification = {1, Attack; 0, Normal} ..... (1)

Feature normalization is performed using Min-Max Scaling:

$$X' = (X - X_{min}) / (X_{max} - X_{min}) \quad \text{..... (2)}$$

where:

X = Original feature value

Xmin = Minimum feature value

Xmax = Maximum feature value

Step 2: Intrusion Detection Model Training

The preprocessed dataset is divided into training and testing datasets using an 80:20 ratio. Three classification models are trained:

- Random Forest
- Gradient Boosting
- Deep Neural Network (DNN)

The prediction function is represented as:

$$\hat{y} = f(X) \quad \text{..... (3)}$$

where:

X = Input feature vector

$\hat{y}$ = Predicted class label

f(X) = Trained intrusion detection model

Step 3: Attack Forecasting

Historical attack counts are converted into time-series data to forecast future attack occurrences. Linear Regression and LSTM models are trained to predict attack trends for the next 30 days.

The forecasting function is represented as:

$$\hat{y}_{t+1} = g(X_t) \quad \text{..... (4)}$$

where:

X_t = Historical attack data

$\hat{y}_{t+1}$ = Predicted future attack count

$g(X_t)$ = Forecasting model

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Step 4: Performance Evaluation

The intrusion detection models are evaluated using standard classification metrics, while forecasting models are assessed using regression metrics.

Accuracy = (TP + TN) / (TP + TN + FP + FN) ..... (5)

Precision = TP / (TP + FP) ..... (6)

Recall = TP / (TP + FN) ..... (7)

F1-Score = 2 × (Precision × Recall) / (Precision + Recall) ..... (8)

MAE = (1/n) Σ |Actual − Predicted| ..... (9)

RMSE = √[(1/n) Σ (Actual − Predicted)²] ..... (10)

where:

TP = True Positives

TN = True Negatives

FP = False Positives

FN = False Negatives

The model with the highest classification performance and the lowest forecasting error is selected as the final model. Experimental results demonstrate that the proposed framework effectively detects network intrusions while accurately forecasting future cyber attack trends, providing a comprehensive solution for proactive network security.

IV. PSEUDO CODE

Step 1: Load the UNSW-NB15 dataset.

Step 2: Pre-process the dataset.
- a) Remove duplicate records and handle missing values.
- b) Apply Label Encoding to categorical features.
- c) Normalize numerical features using Min-Max Scaling.

Step 3: Split the dataset into Training Set (80%) and Testing Set (20%).

Step 4: Train the Intrusion Detection Models.
- a) Initialize the Random Forest classifier.
- b) Train the model using the training dataset.
- c) Predict intrusion labels for the test dataset.
- d) Initialize the Gradient Boosting classifier.
- e) Train the model and predict intrusion labels.
- f) Initialize the Deep Neural Network (DNN).
- g) Train the DNN model and predict intrusion labels.

Step 5: Evaluate the Intrusion Detection Models.
- a) Calculate Accuracy.
- b) Calculate Precision.
- c) Calculate Recall.
- d) Calculate F1-Score.
- e) Generate the Confusion Matrix.

Step 6: Prepare historical attack data for forecasting.
- a) Extract attack category counts.
- b) Convert attack records into time-series format.

Step 7: Train the Attack Forecasting Models.
- a) Initialize the Linear Regression model.
- b) Train the model and forecast future attack counts.
- c) Initialize the Long Short-Term Memory (LSTM) model.
- d) Train the model and forecast future attack counts.

© 2026 IJIRCCE | Volume 14, Issue 7, July 2026 DOI: 10.15680/IJIRCCE.2026.1407029

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Step 8: Evaluate the Forecasting Models.
a) Calculate Mean Absolute Error (MAE).
b) Calculate Root Mean Square Error (RMSE).
c) Compare forecasting performance.

Step 9: Select the best-performing intrusion detection and forecasting models based on evaluation metrics.

Step 10: Display intrusion detection results, attack forecasts, and performance metrics using the Streamlit dashboard.

Step 11: End.

V. SIMULATION RESULTS

The proposed Machine Learning-Based Intrusion Detection and Forecasting System was evaluated using the **UNSW-NB15** dataset containing **257,673** network traffic records, comprising **93,000** normal traffic samples and **164,673** attack samples. The framework was implemented using **Python, Scikit-learn, TensorFlow/Keras,** and **Streamlit.** During preprocessing, missing values and duplicate records were removed, categorical features were transformed using Label Encoding, and numerical attributes were normalized using Min-Max Scaling. The processed dataset was divided into training and testing sets using an **80:20** ratio for model development and evaluation.

The intrusion detection module was implemented using **Random Forest, Gradient Boosting, and Deep Neural Network (DNN)** models, while the attack forecasting module employed **Linear Regression and Long Short-Term Memory (LSTM)** models. Classification performance was evaluated using **Accuracy, Precision, Recall, F1-Score, and Confusion Matrix**, whereas forecasting performance was assessed using **Mean Absolute Error (MAE) and Root Mean Square Error (RMSE)**. Experimental results demonstrate that the proposed framework accurately detects malicious network traffic and effectively forecasts future cyber attack trends. The integrated Streamlit dashboard further enhances the system by providing interactive visualization of detection results, attack forecasts, and comparative model performance.

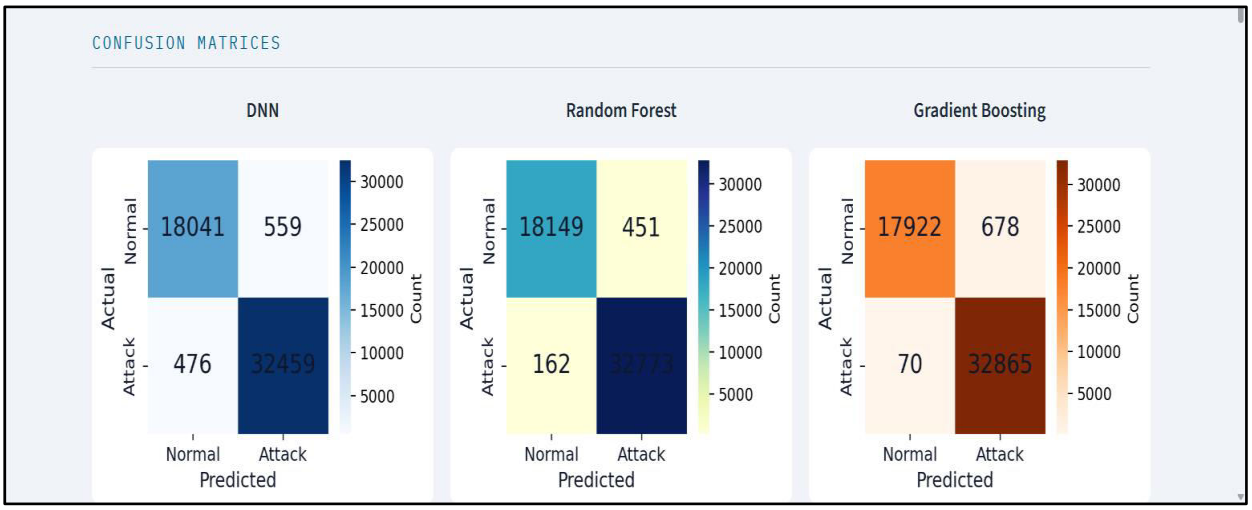


Figure 2: Confusion Matrix Comparison of Intrusion Detection Models (Random Forest, Gradient Boosting, and Deep Neural Network)

The confusion matrices illustrate the classification performance of the intrusion detection models by comparing correctly and incorrectly classified network traffic samples. The results indicate that all three models successfully distinguish benign and malicious traffic with low false classification rates, demonstrating their effectiveness in network intrusion detection.

© 2026 IJIRCCE | Volume 14, Issue 7, July 2026 DOI: 10.15680/IJIRCCE.2026.1407029

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

```
--- Final Model Comparison ---

--- Final Model Comparison (Updated) ---
                   Accuracy  Precision (W-Avg)  Recall (W-Avg)  \
DNN                  0.9799             0.9799          0.9799
Random Forest        0.9881             0.9881          0.9881
Gradient Boosting    0.9855             0.9857          0.9855

                   F1-Score (W-Avg)
DNN                          0.9799
Random Forest                0.9881
Gradient Boosting            0.9854
```

Figure 3: Performance Comparison of Intrusion Detection Models.

The performance comparison graph presents the Accuracy, Precision, Recall, and F1-Score achieved by Random Forest, Gradient Boosting, and Deep Neural Network models. The results show that **Random Forest** achieved the highest overall classification accuracy, while Gradient Boosting and DNN also demonstrated reliable detection performance.

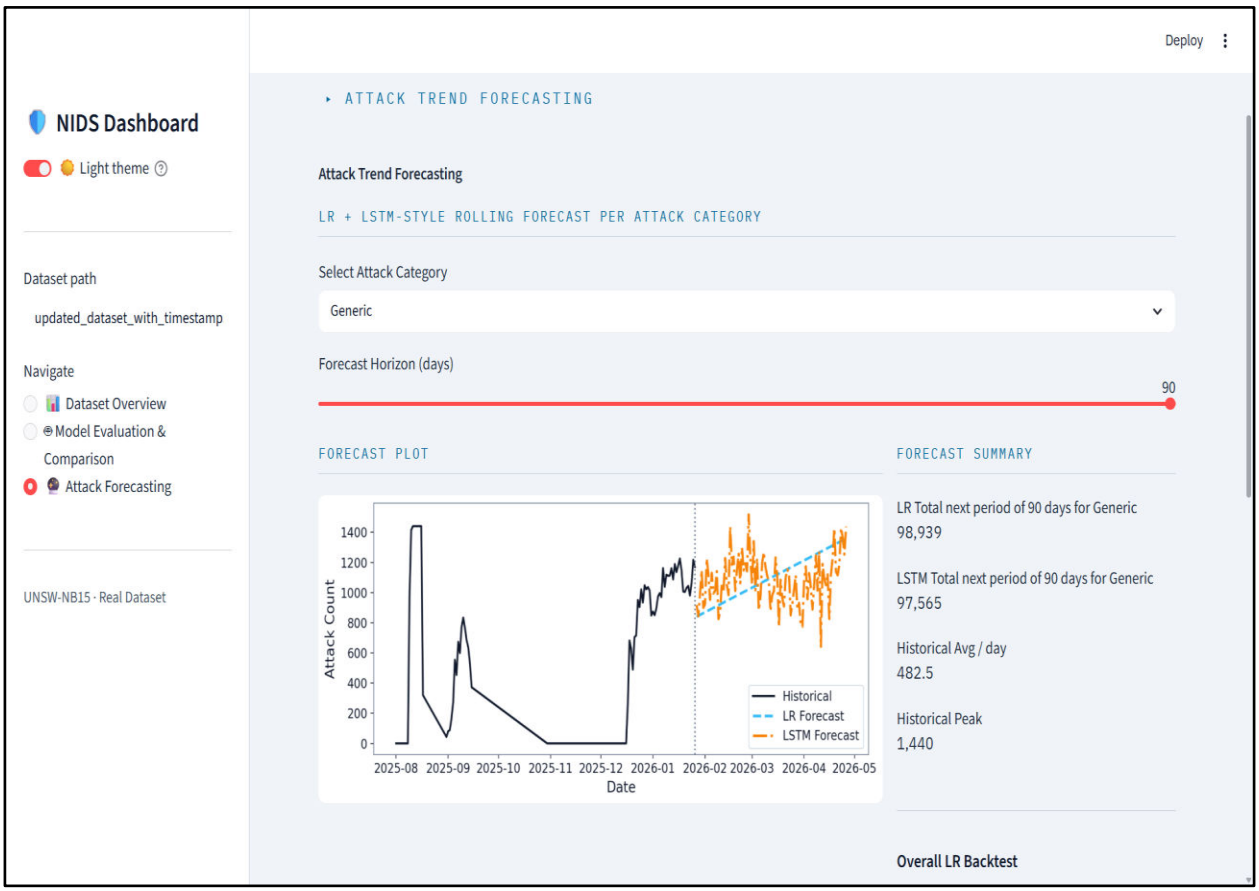


Figure 4: Attack Forecasting Results Using Linear Regression and LSTM Models.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The forecasting graph illustrates the predicted attack trends generated using Linear Regression and LSTM models based on historical attack data. The results indicate that the LSTM model effectively captures temporal attack patterns and provides reliable predictions of future cyber attack occurrences, supporting proactive network security planning.

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	***0.9881***	***0.9881***	***0.9881***	***0.9881***
Gradient Boosting	*0.9855*	*0.9857*	*0.9855*	*0.9854*
Deep Neural Network	*0.9799*	*0.9799*	*0.9799*	*0.9799*

Table 1: Performance Comparison of Intrusion Detection Models

The table summarizes the performance metrics of the intrusion detection models. The comparative analysis demonstrates that all implemented models achieved high classification performance, with Random Forest providing the best overall accuracy, while Gradient Boosting and Deep Neural Network also produced competitive results. The forecasting models further achieved low MAE and RMSE values, confirming the effectiveness of the proposed framework for both intrusion detection and cyber attack forecasting.

VI. CONCLUSION AND FUTURE WORK

The proposed Machine Learning-Based Intrusion Detection and Forecasting System provides an effective framework for detecting malicious network traffic and predicting future cyber attack trends. By integrating Random Forest, Gradient Boosting, and Deep Neural Network (DNN) models with Linear Regression and LSTM forecasting techniques, the system achieves accurate intrusion detection and reliable attack prediction. The Streamlit dashboard further enhances usability through interactive visualization of detection results and forecasting outputs. Experimental results demonstrate the effectiveness of the proposed framework in improving network security. The combination of machine learning and deep learning techniques enhances detection accuracy while enabling proactive cybersecurity through attack trend prediction. Experimental results demonstrate that the proposed framework achieves high intrusion detection performance with low forecasting errors, confirming its effectiveness for intelligent network security. Future work includes integrating real-time network traffic analysis, advanced deep learning architectures, automated threat response mechanisms, and cloud-based deployment to further improve scalability, forecasting accuracy, and practical applicability in modern cybersecurity environments.

REFERENCES

1. Venkata Siva, O., Neeraja, K., Kalyan, D., & Siva Naga, K. (2024). “Cyber Attack Detection and Prediction System”. In 2024 International Conference on Advances in Computer Science and Engineering. IEEE.
2. Buczak, A. L., & Guven, E. (2016). “A survey of data mining and machine learning methods for cyber security intrusion detection”. IEEE Communications Surveys & Tutorials, 18(2), 1153–1176.
3. Ferrag, M. A., Shu, L., & Friha, O. (2020). “Cyber-security for intelligent systems: A survey”. IEEE Access, 8, 14810–14828.
4. Alom, M. Z., Islam, M. M., & Khan, I. (2018). “Recurrent neural network (RNN) based intrusion detection systems”. In 2018 International Conference on Computing, Electronic and Electrical Engineering (ICE-CEE) (pp. 1–6). IEEE.
5. Liu, H., Lang, B., Liu, M., & Yan, H. (2019). “CNN and RNN based deep learning methods for cybersecurity applications: A survey”. IEEE Access, 7, 105–112.
6. N. Moustafa and J. Slay, “The UNSW-NB15 Dataset for Network Intrusion Detection Systems,” in 2015 Military Communications and Information Systems Conference (MilCIS), IEEE, pp. 1–6, 2015.
7. L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
8. J. H. Friedman, “Greedy Function Approximation: A Gradient Boosting Machine,” The Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
9. F. Chollet, Deep Learning with Python, 2nd ed., Shelter Island, NY, USA: Manning Publications, 2021.
10. S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997..

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details