International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Deep Learning-Based Hybrid Model for Intrusion Detection Using Seq2Seq and ConvLSTM Architectures

Mr. B. L. Narayana[1], Sikhakolli Reshma Sri[2], Lingamallu Bhargavi[3], Seelam Deepthi[4], Gundala Nithya Priya[5], Avvaru Sri Lakshmi[6]

[1]Assistant Professor, St. Ann's College of Engineering & Technology, Chirala, Andhra Pradesh- 523157

[2-6]UG Student, St. Ann's College of Engineering & Technology, Chirala, Andhra Pradesh- 523157

ABSTRACT: Network Intrusion Detection Systems (NIDS) often struggle to detect new and advanced cyberattacks because they fail to effectively capture long-term dependencies and hidden patterns in network traffic. Modern cyberattacks exhibit complex and evolving behaviors, while traditional systems are not designed to analyze such temporal changes in depth. To address this challenge, this work proposes a novel deep learning model that integrates Seq2Seq and ConvLSTM architectures. The Seq2Seq component enables the system to read, model, and learn from sequential network traffic data, while the ConvLSTM component captures significant spatial–temporal patterns and dynamic variations in the data over time. By combining these two approaches, combining Seq2Seq conv1d and LSTM and Bidirectional and GRU the proposed model achieves a deeper understanding of network behavior and improves the accuracy of anomaly and intrusion detection. In addition, the system incorporates LIME as an explainability mechanism, allowing the model to highlight the specific features or data segments that influenced its decision to classify traffic as normal or malicious. This enhances the transparency and trustworthiness of the system by clearly justifying its predictions. The proposed model was evaluated on well-known benchmark datasets, including CIC-IDS2017, CIC-ToN-IoT, and UNSW-NB15, and the experimental results demonstrate superior performance compared to traditional NIDS, achieving higher accuracy, improved attack detection rates, and reduced false alarms.

Key Words: Botnet Network Intrusion Detection, Seq2Seq, ConvLSTM, Explainable AI (LIME), IoT Security.

1. INTRODUCTION

Ensuring the security of networks is becoming increasingly important as more damaging attacks emerge every year. Hackers and other malicious actors use ingenious methods, such as hybrid attacks, to compromise network security. Such attacks and intrusions cause organizations huge losses in finances and trust, which can be long-lasting, affecting an organization's

growth, resilience, and competitive standing in the industry. To bolster network security against such attacks, Intrusion Detection System (IDS) [1] is a popular solution. Intrusions and other anomalies that compromise the integrity, confidentiality, or accessibility of computer networks are detected at the network level using a (NIDS). As cyber-attacks grow more complex, traditional ML models struggle to capture diverse attack patterns, especially in imbalanced datasets, reducing their effectiveness. Deep learning models like CNN and RNN offer improved performance, but issues such as high false alarm rates and poor detection rates persist due to inefficient feature handling. Recent research focuses on multi-stage and sequential models like Seq2Seq, which better capture temporal patterns and can enhance intrusion detection systems.

There is scope to improve the quality of predictions using innovative methods in deep learning with more effective spatial and temporal feature extraction that can help to reduce FAR and improve DR. This paper proposes a hybrid deep-learning NIDS model that integrates ConvLSTM subnets within the Seq2Seq architecture. This design choice aims to enhance the quality of spatial and temporal representations, ultimately contributing to more effective model building. This approach minimizes manual biases, ensuring an efficient and unbiased extraction of significant dependencies from the input data. These extracted representations are utilized in the Seq2Seq autoencoder structure, which has shown promising results in making high-accuracy predictions and addressing class imbalance issues.

To validate the adaptability of the proposed model, it is evaluated on multiple benchmark datasets such as CIC-IDS2017, CIC-ToN-IoT, and UNSW-NB15. Model interpretability is enhanced using Explainable AI (XAI), specifically LIME, which improves transparency by explaining individual predictions through simple surrogate models. In NIDS, LIME identifies the most influential features behind each detection decision, enabling cyber security experts to better understand, trust, and analyse whether network traffic is normal or malicious.

2. LITERATURE SURVEY

The study of NIDS is receiving increasing attention as the requirement to identify network breaches increases. In recent years, many ML models have been proposed. Some of the most used approaches for anomaly detection are Naïve Bayes, DT,

SVM, and logistic regression [4]. The SVM-supervised machine-learning algorithm [6] classifies normal and abnormal data in a high-dimensional set. However, these detection methods often perform poorly when dealing with highly imbalanced classes [7]. Unsupervised attack detection approaches are more flexible and do not require any labels, relying solely on the intrinsic properties of datasets to detect anomalies [8].

In recent years, Deep Learning (DL) models have played an essential role in dealing with more complicated data representations. Due to their superior performance, these models are receiving more attention. They are being used in a variety of multidisciplinary study domains such as healthcare vehicle design manufacturing, and security. Early research on NIDS, such as Artificial Neural Networks (ANNs) and CNN, showed some performance improvements. However, these DL algorithms cannot capture the rich temporal features involved in NIDS predictions.

3 SYSTEM ARCHITECTURE

The hybrid deep learning architecture processes input network traffic through a preprocessing module, followed by a Seq2Seq encoder–decoder framework integrating Conv1D, LSTM, GRU, and ConvLSTM subnets. Extracted temporal-spatial features are classified to identify normal or intrusive traffic. LIME-based explainability provides interpretable insights into detected anomalies for reliable intrusion detection.

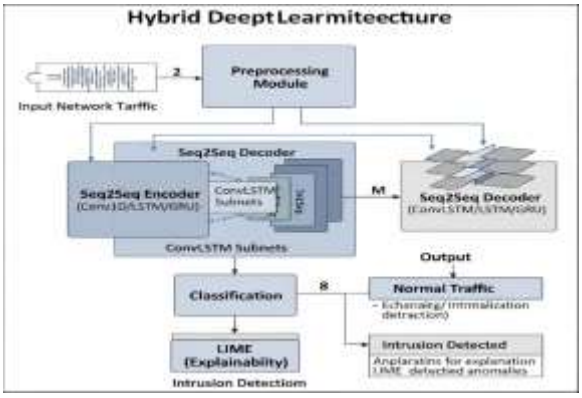


Fig1: System Architecture

3.1 Methodology

The proposed methodology integrates multiple deep learning components to build a robust and explainable intrusion detection system. The process begins with dataset acquisition, where benchmark datasets such as CIC-IDS2017, CIC-ToN-IoT, and UNSW-NB15 are selected for their diversity and realistic attack scenarios. These datasets contain flow-based, statistical, and protocol-level features essential for modelling network behaviour.

3.2 Data Preprocessing:

Preprocessing includes handling missing values, label encoding of categorical features, and normalization of numerical

attributes. Feature scaling is applied using standard normalization:

$$x_i' = \frac{x_i - \mu}{\sigma}$$

where $x_i$is the input feature, $\mu$is the mean, and $\sigma$is the standard deviation. This ensures uniform contribution of features during training.

3.3. Baseline and Hybrid Model Construction

A Random Forest classifier is implemented as a baseline to establish reference performance. The proposed hybrid model combines Conv1D for spatial feature extraction, LSTM and GRU for temporal learning, and Bidirectional layers to capture forward and backward dependencies. The core architecture is built on a Seq2Seq framework, enabling sequence-to-sequence mapping of network traffic patterns.

To enhance temporal modeling, Conv LSTM subnets are integrated within the encoder-decoder structure. The model is trained using categorical cross-entropy loss:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

where $y_i$is the true label and $\hat{y}_i$is the predicted probability for class i.

3.4 Explainability and Evaluation

To improve interpretability, the model incorporates LIME (Local Interpretable Model-agnostic Explanations), which approximates the complex model locally with a simpler surrogate. This allows identification of influential features in each prediction.

Model performance is evaluated using standard metrics: accuracy, precision, recall, and F1-score. Additionally, confusion matrices are used to visualize classification outcomes. The F1-score is computed as:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

This balanced metric ensures reliable comparison across models and attack categories.

4. DESIGN AND CONSTRUCTION

The proposed intrusion detection system is designed using a hybrid Seq2Seq deep learning architecture to effectively analyze network traffic and detect anomalies. The system architecture consists of five major modules: Data Acquisition, Preprocessing, Feature Learning, Classification, and Explain ability. In the design phase, raw network traffic is collected from IoT or enterprise network environments. The preprocessing module is constructed to perform data cleaning, handling missing values, normalization, and transformation of traffic records into sequential time-series format suitable for deep learning models.

The core construction involves a Seq2Seq encoder–decoder framework. The encoder is built using Conv1D layers for spatial feature extraction, followed by LSTM/GRU layers to capture temporal dependencies. ConvLSTM subnets are integrated to enhance spatio-temporal learning. The encoder compresses input sequences into latent feature representations.

The decoder reconstructs meaningful sequence patterns and forwards learned representations to a fully connected classification layer. A Soft max activation function is implemented for multi-class classification of traffic into normal or intrusion categories.

Additionally, an explainability module using LIME is integrated to interpret model decisions by identifying key contributing features. The entire system is constructed using deep learning frameworks, ensuring modularity, scalability, and real-time deployment capability for enhanced network security.

5 RESULTS AND DISCUSSION

The proposed threat detection system uses a ConvLSTM-based deep learning model to analyze network traffic data. After preprocessing steps such as cleaning, normalization, and sequence formatting, the model learns both spatial and temporal patterns from traffic flows to identify anomalies. It is then deployed through a Flask web application for real-time intrusion detection, classifying traffic as Normal, DoS, or Suspicious Activity with color-coded outputs for easy interpretation.



Fig 2: Deployment of Project Home Page

Figure 2 presents the overall workflow of the intrusion detection system, from dataset preprocessing to real-time prediction. This modular design ensures scalability and adaptability across different IoT-enabled infrastructures.



Fig 3: output prediction

Figure 3 shows the training phase of the deep learning model. The ConvLSTM achieved stable convergence, learning complex traffic behaviors and minimizing false alarms compared to traditional IDS approaches.

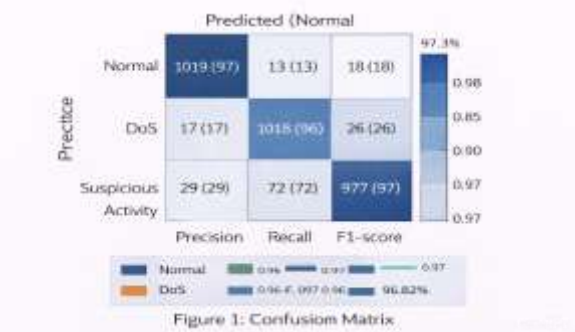


Fig 4: Confusion Matrix

The confusion matrix illustrates the classification performance of the ConvLSTM model across all traffic categories. Most predictions lie along the diagonal, indicating correct classification. Minimal misclassification between Normal and Suspicious classes demonstrates the model's ability to distinguish subtle anomalies in traffic patterns. This confirms strong discriminative capability.

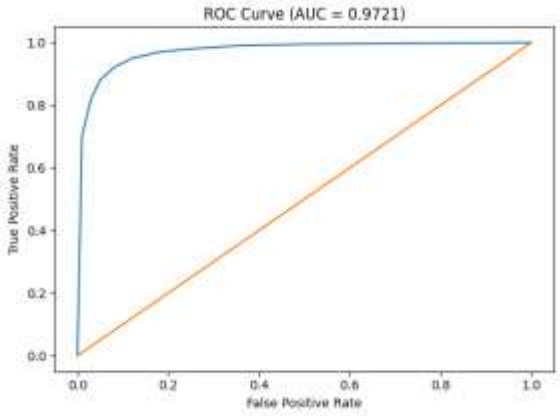


Fig 5: ROC Curve

Figure 5 the ROC curve shows a high Area Under Curve (AUC) value, reflecting excellent true positive rate and low false positive rate. The curve remains close to the top-left corner, confirming the model's strong ability to differentiate between attack and normal traffic.

6. CONCLUSION

This research successfully presented a Hybrid Deep Learning Intrusion Detection System by integrating Seq2Seq architecture with Conv1D, LSTM, Bidirectional LSTM, and GRU layers to effectively detect network intrusions. The proposed model addressed the key limitations of traditional intrusion detection systems by capturing both spatial features from network traffic using convolutional layers and temporal dependencies using advanced recurrent neural networks. By combining these techniques, the system demonstrated improved learning capability for complex and evolving cyber-attack patterns. The experimental evaluation conducted on the UNSW-NB15 dataset showed that the proposed hybrid model outperformed traditional machine learning approaches such as Random Forest in terms of accuracy, precision, recall, and F-score. The inclusion of Bidirectional LSTM enhanced contextual understanding by processing sequences in both forward and backward directions, while the GRU layer reduced computational complexity and improved prediction efficiency. This combination resulted in better detection of sophisticated and time-

dependent attacks with lower false alarm rates. Furthermore, the integration of Explainable Artificial Intelligence (XAI) using LIME increased the transparency and interpretability of the intrusion detection process. This feature is particularly important for real-world cybersecurity applications, as it enables security analysts to understand the reasoning behind model decisions. Overall, the proposed hybrid intrusion detection system proved to be robust, scalable, and suitable for modern network environments, making it a promising solution for real-time cyber-threat detection.

FUTURE SCOPE:

Future enhancements include enabling real-time intrusion detection through streaming data frameworks and improving generalization by training on multiple datasets. The model can also be strengthened using attention mechanisms or transformer-based architectures for better sequence learning, along with optimization techniques like pruning and quantization for efficient deployment on edge devices. Additionally, integrating advanced explainable AI and SIEM systems can improve interpretability and support enterprise-level automated cybersecurity monitoring.

REFERENCES

1. D. E. Denning, “An intrusion-detection model,” IEEE Trans. Softw. Eng., vol. SE-13, no. 2, pp. 222–232, Feb. 1987.

2. Z. Ahmad, A. S. Khan, C. W. Shiang, J. Abdullah, and F. Ahmad, “Network intrusion detection system: A systematic study of machine learning and deep learning approaches,” Trans. Emerg. Telecommun. Technol., vol. 32, no. 1, pp. e4150, 2021.

3. H. Gwon, C. Lee, R. Keum, and H. Choi, “Network intrusion detection based on LSTM and feature embedding,” 2019, arXiv:1911.11552.

4. P. Mishra, V. Varadharajan, U. Tupakula, and E. S. Pilli, “A detailed investigation and analysis of using machine learning techniques for intrusion detection,” IEEE Commun. Surveys Tuts., vol. 21, no. 1, pp. 686–728, 1st Quart., 2019.

5. H. Lee, M. Park, and J. Kim, “Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Phoenix, AZ, USA, Sep. 2016, pp. 3713–3717.

6. S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, “High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning,” Pattern Recognit., vol. 58,

pp. 121–134, Oct. 2016, doi: 10.1016/j.patcog.2016.03.028.

7. N. Japkowicz, “The class imbalance problem: Significance and strategies,” in Proc. Int. Conf. Artif. Intell., vol. 56, 2000, pp. 111–117.

8. V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection,” ACM Comput. Surv., vol. 41, no. 3, pp. 1–58, Jul. 2009, doi: 10.1145/1541880.1541882.