Journal of
Cyber Security

Tech Science Press

https://doi.org/10.32604/jcs.2026.083592

ARTICLE

A Framework for Simulated Zero-Day Detection Using Synthetic Attack Generation and Out-of-Distribution Evaluation

Peter Kipngeno Langat*, Michael Kimwele and Dennis Kaburu

School of Computing and Information Technology, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

*Corresponding Author: Peter Kipngeno Langat. Email: peterlangat20@jkuat.ac.ke or peterlangat20@gmail.com

Received: 07 April 2026; Accepted: 03 August 2026; Published: 21 August 2026

ABSTRACT: Zero-day attacks pose a significant threat to computer systems and networks as they exploit weaknesses that have not been recognized by security professionals or software creators and for which there are no existing protective measures. This study introduced an innovative method for identifying Zero-day attacks through a Recurrent neural network model. To effectively mitigate these risks, not only is continuous monitoring essential, but also the implementation of machine learning. The model was trained on network traffic data and leveraged on the ability of Recurrent Neural Networks (RNNs) to learn complex patterns and identify anomalies that may indicate the presence of a Zero-day attack. To enhance the model's ability to generalize and accurately identify novel attack vectors that traditional methods may overlook, we trained it on a dataset that includes both known and synthetically generated Zero-day attack signatures. Our methodology involves preprocessing network traffic data to extract relevant features, which are then fed into an RNN architecture. The RNN model was trained on a labeled dataset containing both normal and attack traffic samples, allowing it to learn the fundamental patterns of benign and malicious network activity. The Recurrent architecture of the model enables it to retain a memory of previous observations, making it well-suited for identifying deviations from the expected network behavior in real-time. By leveraging the temporal dimension of network traffic data, we enhanced the ability to identify novel threats and bolster the security of computer systems and networks. This approach held the promise for mitigating the ever-evolving landscape of cyber threats and ensuring the integrity and availability of critical digital infrastructure.

KEYWORDS: Machine learning; recurrent neural networks; zero-day attack; anomalies; model

1 Introduction

Today, political and commercial entities are increasingly engaging in sophisticated cyberwarfare to damage, disrupt, or censor information content in computer networks. when designing network protocols, there is a need to ensure reliability against intrusions of powerful attackers that can even control a fraction of parties in the network. The advancement of technologies such as online banking, healthcare systems, and e-commerce has significantly increased the exposure of organizations and individuals to a wide range of cyberattacks, including Zero-day attacks. Traditional machine learning methods have limitations in detecting Zero-day attacks because they primarily rely on labeled training data representing known attack patterns. As a result, these models often struggle to identify previously unseen or evolving threats that differ from the data used during training. Current approaches to addressing Zero-day attacks leverage either machine learning/deep neural network techniques or anomaly-based detection methodologies. In an increasingly interconnected and digitally reliant world, the threat of cyber-attacks looms larger than ever

before. Consequently, an urgent requirement arises for advanced detection mechanisms that can proactively identify such vulnerabilities before they can be exploited, thereby enhancing the overall security posture of networks and systems.

Cybercriminals constantly devise new techniques to breach security measures and compromise the confidentiality, integrity, and availability of data and systems. One particularly challenging category of cyber-attacks is Zero-day attacks, which exploit vulnerabilities that are previously unknown to security experts and, so, lack readily available countermeasures. Moreover, traditional strategies often fall short in their ability to quickly adapt to new attack patterns, underscoring the necessity for more sophisticated methods like Recurrent Neural Networks, which can effectively learn and predict potential exploits from historical data and behavioral patterns within network traffic. To address this critical security challenge, there is a growing need for advanced and adaptive approaches capable of detecting Zero-day attacks. Cybersecurity threats encompass any acts intended to disrupt, disable, destroy, or maliciously control a computing environment or infrastructure. The primary function of information security is to protect the digital data being transferred across networks from disclosure, alteration, destruction, manipulation, and unauthorized access. Machine learning and deep learning methods, especially RNNs, represent a specific category of neural networks that excel in evaluating sequential data, which makes them highly suitable for examining patterns in network traffic. By exploiting the temporal dependencies inherent in network communication data, RNNs have the potential to learn and identify subtle, evolving attack patterns that often elude traditional detection methods. The increasing complexity and frequency of cyber threats necessitate the adoption of innovative methodologies that can keep pace with attackers' evolving strategies, thereby reinforcing the security frameworks of sensitive systems and infrastructures.

This research explored the use of Recurrent Neural Networks for the detection of cyber-attacks, particularly Zero-day attacks. By leveraging the power of deep learning, the goal was to develop a system that can autonomously identify malicious network activity without relying on predefined attack signatures. This approach differs fundamentally from traditional Intrusion Detection Systems, which rely on databases of known attack patterns and may fail to detect previously unseen threats. The developed system operates in two key stages. First, it preprocesses network traffic data to create appropriate input sequences for the RNN, effectively translating the network interactions into a format that can be processed by the neural network. Second, the RNN-based anomaly detection component is responsible for identifying deviations from normal network behavior. By capturing the sequential nature of network communications, the RNN model identified anomalous patterns within the network traffic data that may be indicative of a previously unknown, or Zero-day, cyber threat. This paper discussed the architecture, training, and evaluation of our developed Recurrent Neural Network-based system for detecting cyber-attacks. We present empirical results comparing the performance of our approach to traditional Intrusion Detection System techniques, demonstrating the effectiveness and potential of employing RNNs in the ongoing effort to mitigate cyber threats, especially Zero-day attacks. The adoption of these advanced machine learning methods holds promise to enhance the security posture of organizations and networks by enabling proactive, adaptive, and robust detection of cyber threats.

This study introduces a novel approach to Zero-day attack detection by integrating sequence-based anomaly detection with synthetically generated attack patterns using a Recurrent Neural Network (RNN) architecture. Unlike traditional intrusion detection systems that rely on static signatures or shallow machine learning models, the proposed method leverages temporal dependencies in network traffic to identify subtle deviations indicative of previously unseen attacks. The key novelty of this work lies in three aspects: (i) the incorporation of synthetic Zero-day attack samples to enhance the model's exposure to unseen threat patterns, (ii) the use of Long Short-Term Memory (LSTM) networks to capture sequential behavioral

characteristics of network traffic, and (iii) a comprehensive comparative evaluation against multiple baseline machine learning models. This combination enables improved generalization and robustness in detecting emerging cyber threats beyond known attack signatures.

The novelty of this study does not lie in the use of Recurrent Neural Networks alone, as RNN-based intrusion detection has been extensively investigated in prior research. Instead, this work contributes a structured framework for simulated Zero-day attack detection through three integrated components. First, synthetic attack generation is employed to create attack behaviors that are absent from the original training data, thereby approximating previously unseen threats. Second, network traffic is modeled as temporal sequences using LSTM networks to capture behavioral dependencies that traditional machine learning classifiers cannot adequately represent. Third, the study evaluates model robustness under out-of-distribution attack scenarios, providing a more realistic assessment of Zero-day detection capability than conventional train-test evaluations. This combination establishes a practical methodology for evaluating the generalization capacity of intrusion detection systems against emerging cyber threats.

The structure of the article is as follows: Section 2 surveys existing literature relevant to the study. Section 3 describes the proposed model for detecting Zero-day attacks. In Section 4, the algorithm is explained in detail. Section 5 presents the experimental findings along with an analysis, and Section 6 offers concluding remarks and potential avenues for future research.

2 Literature Review

This section reviews the literature on detection of Zero-day attacks using machine learning. We reviewed the various approaches that exist. Researchers have explored a variety of techniques to address the challenge of cyber-threat detection, including signature-based, anomaly-based, and hybrid approaches. Signature-based detection systems use predefined patterns or signatures to identify known threats, while anomaly-based systems aim to detect deviations from normal network behavior. Cybersecurity threats can be broadly categorized into two types: known attacks and Zero-day attacks. Known attacks refer to those that have been previously identified and for which protective measures have been developed. A Zero-day attack refers to a cybersecurity threat that exploits previously unknown vulnerabilities in software or systems, making it challenging to detect and mitigate. Zero-day attacks take advantage of undiscovered or unpatched security vulnerabilities, making them difficult to detect using standard security tools. Studies have shown a consistent rise in the occurrence of such attacks, emphasizing the need for more sophisticated detection strategies.

Several approaches have been developed to address the detection of cybersecurity threats, including signature-based, anomaly-based, and hybrid detection methods. We learned that signature-based detection systems rely on predefined signatures or patterns that characterize previously identified threats in order to detect known attacks. After a cybersecurity attack is identified, the signature is added to the database as a recognized cyber-attack pattern. The system then compares the observed network traffic or system behavior against the stored signatures to identify any matches, triggering an alert if a known threat is detected. As a result, the growing prevalence of Zero-day attacks has significantly reduced the effectiveness of Signature-based Intrusion Detection Systems (SIDS) as a reliable IDS approach. Signature detection systems are effective in detecting identified attacks, they are not able to identify unknown or "Zero-day" threats, as they require the availability of pre-defined attack signatures. Anomaly detection using a Recurrent Neural Network (RNN) model involves identifying unusual patterns or deviations from normal behavior in sequential data. RNNs are particularly suited for this task because they can capture temporal dependencies and patterns in sequences, making them effective for analyzing time-series data, which is common in anomaly detection scenarios [1].

These systems define a model of typical system or network behavior and subsequently monitor for notable departures from this established norm. When an anomaly is detected, the system triggers an alert, allowing security analysts to investigate and mitigate the potential threat.

With the increasing complexity and volume of cybersecurity threats, traditional rule-based and signature-based detection approaches have become less effective. Literature review shows that machine learning algorithms such as supervised learning algorithms are trained on labeled data. Unsupervised algorithms, on the other hand, do not require labeled data and instead aim to identify patterns in the data to uncover anomalies. In recent years, deep learning, a sub-field of machine learning, has gained significant attention for anomaly detection due to its ability to automatically extract complex features from raw data without the need for manual feature engineering. Research has shown that machine learning-based anomaly detection can effectively identify both known and unknown attacks, including Zero-day threats, by learning patterns of normal behavior and detecting deviations from those patterns. The use of Recurrent neural networks, a type of deep learning architecture, has been explored for the detection of Zero-day attacks. Recurrent neural networks are a class of deep learning models particularly well-suited for processing sequential data, such as network traffic or system logs. Deep learning, particularly using architectures like Recurrent Neural Networks, has shown promise in capturing temporal dependencies in data, making it especially suited for analyzing time-series data such as network traffic logs to detect Zero-day attacks, which are characterized by [2]. Anomaly detection identifies patterns that deviate from established normal behavior without requiring prior knowledge of attack signatures, thereby enhancing the system's ability to adapt to emerging network threats. It detects abnormal user activities and environmental changes that may indicate potential attack patterns or previously unknown cyber threats [3].

Proposed Zero-Day Attack Detection Model

The proposed model for zero-day attack detection is shown in Fig. 1.

The research focused on exploring anomaly detection techniques leveraging Recurrent neural networks to detect Zero-day attacks in network environments. A supervised learning model involves training on labeled data to classify normal and anomalous traffic, while unsupervised learning models aim to identify patterns in unlabeled data to detect anomalies without prior knowledge of attack signatures. Some key anomaly detection techniques that will be investigated include supervised learning such as SVM and Forests learn normal behavior from labeled data and then identify deviations as anomalies. Unsupervised Anomaly Detection: Clustering-based approaches that group similar network traffic/events and flag outliers as anomalies, as well as reconstruction-based models that learn to reconstruct normal data and flag high-error instances as anomalies.

Deep Learning for Anomaly Detection: Leveraging the pattern recognition capabilities of deep neural networks, particularly Recurrent neural networks, to learn complex representations of normal network traffic and detect deviations that may indicate Zero-day attacks. The research will delve into the theoretical underpinnings and empirical evaluation of these techniques to assess their effectiveness in detecting Zero-day attacks in real-world network environments [4]. The supervised learning algorithm includes k-nearest neighbors (KNN), support vector machines, and random forests. The unsupervised learning algorithms include k-means clustering, principal component analysis, and autoencoders.

Despite the promising performance of deep learning techniques in intrusion detection, several limitations remain evident in existing approaches. Many models suffer from overfitting due to reliance on static and imbalanced datasets, limiting their ability to generalize to unseen attack patterns. Additionally, a number of deep learning-based intrusion detection systems focus primarily on classification of known attacks rather than true anomaly detection, thereby reducing their effectiveness in Zero-day scenarios [4]. Furthermore,

existing models often fail to adequately capture temporal dependencies in network traffic, especially when non-sequential architectures are employed. In contrast, the proposed RNN-based approach addresses these limitations by leveraging LSTM networks to model sequential dependencies, incorporating synthetic data to improve exposure to novel attack behaviors, and emphasizing anomaly-based detection rather than signature matching, thereby enhancing its capability to detect Zero-day attacks.

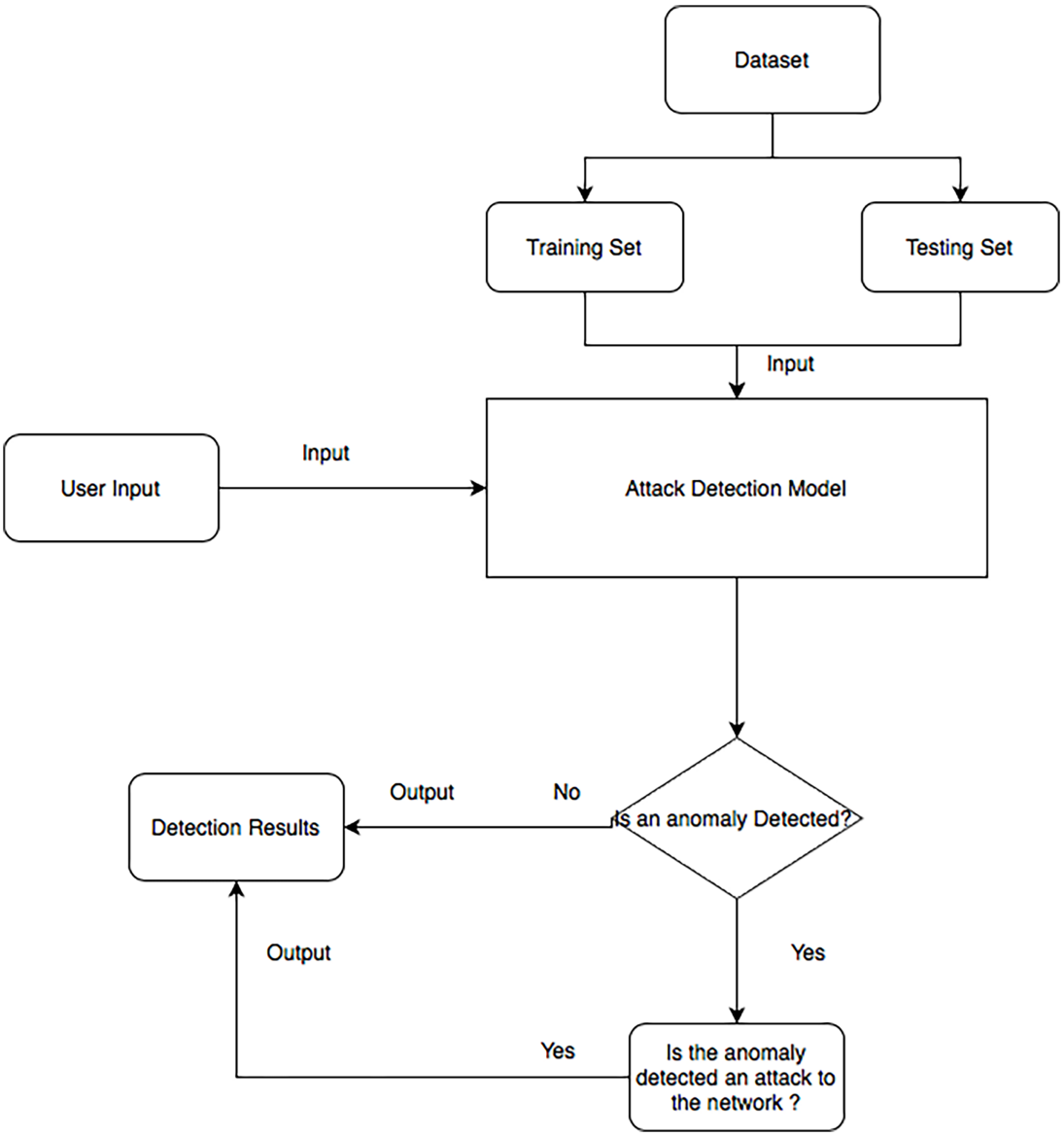

Figure 1: Zero-day detection model.

"Zero-day" is a software vulnerability that is unknown to the software vendor or Internet community, and thus no patch or solution is available for the vulnerability. Zero-day attacks, which exploit previously unknown vulnerabilities in software or systems, pose a significant challenge for traditional signature-based intrusion detection systems [5].

These attacks often leave no trace in the existing attack signature databases, making them difficult to detect using conventional techniques. Detecting Zero-day attacks is inherently difficult due to:

1. Lack of Known Signatures: These attacks target undisclosed vulnerabilities, providing no identifiable patterns for signature-based detection systems.
2. Dynamic Behavior: Zero-day exploits often evolve to bypass static defenses, limiting the effectiveness of traditional intrusion detection approaches.
3. Evasion Techniques: Attackers may use advanced evasion techniques, such as polymorphism or obfuscation, to circumvent detection.
4. Low Volume: Some Zero-day attacks may generate low volumes of network traffic, making them difficult to distinguish from legitimate activity using anomaly-based detection.

Addressing the challenge of Zero-day attacks requires a more advanced approach to intrusion detection, such as machine and deep learning. These approaches aim to build models of normal system behavior and identify anomalies that may indicate Zero-day threats. RNN are well-suited for detecting Zero-day instances, since they learn complexities in IT environment, and they can identify deviations from the learned model of normal behavior. The most popular Zero-day attack is the Print Spooler vulnerability attack which affects Microsoft Windows operating systems where hackers are able to run arbitrary code on devices with Print Spooler enabled. RNN are a class of neural networks that are particularly well-suited for processing data, e.g., traffic logs. Recurrent neural networks, such as LSTM and Gated Recurrent Units, can effectively capture the temporal dependencies and patterns in network traffic data, making them a promising approach for detecting Zero-day attacks [5].

Detecting Zero-day attacks is achieved by RNN through training systems on normal network traffic data and then uses the developed a model to detect occurrences that substantially differ from the established normal patterns. This can be achieved by calculating the reconstruction error or the prediction error of the Recurrent neural network on new examples, with instances that exhibit high errors being identified as anomalies. The models are trained on normal instances to learn a model of normal behavior. During inference, the Recurrent neural network processes new network traffic data and identifies instances that deviate significantly from the learned model of normality as potential anomalies or Zero-day attacks. The effectiveness of Recurrent neural networks for anomaly detection has been demonstrated in several recent studies. These studies have shown that Recurrent neural networks can outperform traditional machine learning techniques, such as support vector machines and decision trees, in detecting complex, stealthy attacks that do not have a well-defined signature. This can be particularly useful for detecting Zero-day attacks, where the attack signature may not be known in advance [6].

The primary advantage of neural networks is their capacity to handle imprecise and uncertain data, as well as their ability to derive insights from high-dimensional, complex data that may not be easily captured by traditional statistical models. In comparing Recurrent neural networks with other machine learning approaches, Recurrent neural networks have demonstrated superior performance in detecting anomalies and Zero-day attacks in network traffic data. In one study, researchers found that Recurrent neural networks, such as Long Short-Term Memory and Gated Recurrent Unit models, outperformed traditional machine learning techniques, such as support vector machines and k-nearest neighbors, in detecting anomalies in network traffic data [7].

The primary factors contributing to the enhanced performance of Recurrent neural networks in this area are their ability to effectively capture temporal dependencies and intricate patterns in network traffic data which is crucial for identifying Zero-day attacks that may not have a well-defined signature. Recurrent neural networks can learn a more comprehensive model of normal behavior by considering the context and history of the network traffic, rather than just isolated instances. They can generalize and identify previously unseen patterns, which is important for detecting Zero-day attacks that may not have been observed during the training phase [8].

Novel Attack Vector

In the context of this study, a novel attack vector refers to a network attack behavior whose feature distribution and temporal characteristics differ significantly from attack instances observed during model training. Unlike traditional attack signatures that are explicitly represented in training datasets, novel attack vectors are generated by modifying traffic attributes, communication patterns, packet timing characteristics, and protocol behaviors to create previously unseen attack manifestations. The objective is to simulate realistic variations of cyber-attacks that intrusion detection systems may encounter after deployment. Therefore, the term "novel attack vector" in this research does not imply the discovery of an entirely new software vulnerability but rather the generation of previously unseen behavioral patterns that challenge the generalization capability of the detection model.

Fig. 2 illustrates the architecture of a Recurrent Neural Network (RNN), highlighting how sequential data is processed through recurrent connections that enable the network to retain information from previous inputs. This memory capability makes RNNs particularly suitable for tasks involving time-series data, natural language processing, and sequence prediction.

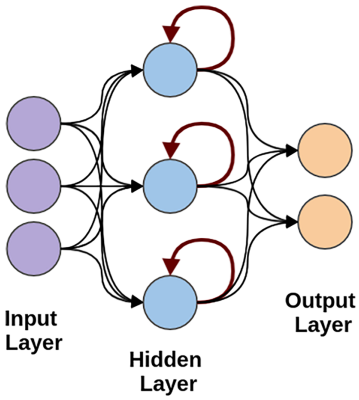

Figure 2: Recurrent neural network illustration.

3 Methodology

This study employed a Recurrent Neural Network (RNN), specifically Long Short-Term Memory (LSTM) networks, to detect Zero-day attacks in network traffic. The methodology involved a quantitative experimental research design, where network traffic data from the UNSW-NB15 dataset was collected, preprocessed through feature selection, normalization, and encoding, and organized into time-series sequences suitable for sequential modeling. These sequences were labeled as normal or malicious based on known behaviors.

To ensure reproducibility, the following detailed configuration was used. The LSTM model architecture consisted of two LSTM layers with 64 and 32 hidden units, respectively, followed by a dropout layer (rate = 0.3) and a dense output layer with sigmoid activation. The model was compiled using the Adam optimizer with a learning rate of 0.001, binary cross-entropy loss, and a batch size of 64. Training was conducted for 50 epochs with early stopping (patience = 5 epochs) based on validation loss. The dataset was split into training (70%), validation (15%), and testing (15%) using stratified sampling to preserve class distribution. No preprocessing steps (e.g., normalization, encoding) were applied before the split to prevent data leakage. Min-Max scaling parameters were fitted exclusively on the training set and transformed on validation and test sets. All random seeds were fixed to 42 to enable exact replication. The experimental design included comparative evaluation against traditional machine learning models such as Random Forest,

XGBoost, SVM, and Logistic Regression to benchmark the RNN's effectiveness in capturing temporal patterns and detecting complex cyber threats, ensuring rigorous assessment of predictive accuracy and generalization capability.

The UNSW-NB15 dataset was selected for this study because it provides a comprehensive representation of modern network traffic and cyberattack scenarios. Unlike many traditional datasets, it contains a balanced mixture of normal and malicious traffic generated in a controlled environment, ensuring diversity and reliability. Furthermore, compared to datasets such as CICIDS, UNSW-NB15 offers a rich set of network features and well-defined attack categories, making it highly suitable for training deep learning models to detect both known and previously unseen cyber threats. These characteristics enhance the model's ability to generalize across diverse network environments and improve intrusion detection performance.

Research Contribution and Novelty

Unlike many existing intrusion detection studies that commonly evaluate models using conventional random train-test partitioning approaches, where training and testing data may contain similar distributions of attack characteristics, this study proposes a simulated Zero-day detection framework. The framework introduces synthetic attack generation through controlled perturbation of attack attributes and temporal patterns, enabling the creation of attack instances that differ from those observed during model training. These generated attack samples are subsequently used to assess the model's ability to generalize beyond previously learned attack signatures and detect unknown threats. Therefore, the primary contribution of this study is not merely the application of an LSTM architecture, but rather the integration of synthetic attack generation, temporal sequence learning, and out-of-distribution evaluation into a unified Zero-day detection framework.

The preprocessing phase involved multiple steps to ensure data quality and reproducibility. First, irrelevant and redundant features were removed using feature selection techniques based on correlation analysis. Categorical variables were encoded using label encoding to transform them into numerical representations suitable for model input. Numerical features were normalized using Min-Max scaling to ensure uniformity across feature ranges and to improve convergence during training. The processed data was then transformed into fixed-length time-series sequences to align with the requirements of the LSTM model. All preprocessing steps were implemented using deterministic procedures, ensuring that the experimental setup can be reliably reproduced in future studies.

The Developed Zero-Day Attack Detection Algorithm

Building on the model outlined in the preceding section, the proposed algorithm for detecting Zero-day attacks is formulated. This algorithm leverages a Recurrent Neural Network (RNN), specifically a Long Short-Term Memory (LSTM) network, to detect novel cyber threats within network traffic. The process begins with the collection and preprocessing of network data, which is normalized and converted into time-series sequences suitable for RNN input. These sequences are labeled based on known benign and malicious behavior. An LSTM model is then constructed, featuring layers to learn temporal patterns, with dropout applied to reduce overfitting and a sigmoid activation for binary classification [9]. The model is trained on historical data and validated to ensure reliability. Once trained, it can analyze real-time or batch network traffic, outputting a probability score for each sequence. If this score exceeds a defined threshold, the traffic is flagged as a potential Zero-day attack. The model's effectiveness is evaluated using metrics such as accuracy, precision, recall, and F1-score. The description for the metrics is as follows

- Precision: Measures of how many detected attacks are attacks.
- Recall: Measures of how many actual attacks were detected.
- F1-Score: Harmonic mean of Precision and Recall.

- Accuracy: Overall correctness of the model's predictions.

  Area under the Receiver Operating Characteristic curve

This approach enables the detection of novel threats by recognizing behavioral anomalies rather than relying solely on known attack signatures. This proposed algorithm for Zero-day attack detection model is illustrated in Fig. 3 below.

```
Algorithm ZeroDayDetectionRNN
Input:
    - TrafficData: Network traffic sequences
    - Labels: Ground truth (benign or attack)
    - Epochs, BatchSize, Threshold
Output:
    - Predictions: Benign or Potential Zero-Day Attack

1. PreprocessData(TrafficData):
    - Normalize features
    - Convert traffic flows to time-series format
    - Return ProcessedData
2. Split ProcessedData into TrainSet, ValSet, TestSet
3. Define LSTMModel():
    - InputLayer: Shape = (time_steps, features)
    - LSTM Layer(s)
    - Dropout Layer (optional)
    - Dense Output Layer with Sigmoid (binary) or Softmax (multi-class)
4. Compile Model using:
    - Loss Function: BinaryCrossEntropy
    - Optimizer: Adam
    - Metrics: Accuracy, Precision, Recall,ROC-AC
5. Train Model on TrainSet for Epochs with BatchSize
6. Evaluate Model on ValSet and TestSet
    - Record metrics and confusion matrix
7. For each NewTrafficSequence in RealTime or BatchInput:
    - Preprocess sequence
    - Prediction = Model.predict(Sequence)
    - If Prediction ≥ Threshold:
        Label as "Potential Zero-Day Attack"
      Else:
        Label as "Benign"
End Algorithm
//The outcome will be used to train the model
```

Figure 3: Zero-day detection model algorithm.

To simulate Zero-day attack scenarios, synthetic attack samples were generated by applying controlled perturbations and transformations to existing attack patterns within the dataset. This included modifying feature distributions, introducing noise, and altering temporal characteristics of network traffic sequences to mimic previously unseen behaviors. These synthetic samples were designed to represent realistic variations of potential attack strategies without replicating known signatures. By incorporating these generated samples into the training process, the model was exposed to a broader spectrum of anomalous patterns, thereby improving its ability to generalize and detect novel attacks. While synthetic data cannot fully replicate real-world Zero-day exploits, it provides a practical approximation for evaluating model robustness in the absence of true Zero-day datasets. Novel attack vectors were generated using controlled feature perturbation techniques applied to known attack records within the UNSW-NB15 dataset. Specifically, numerical traffic

attributes such as packet size, flow duration, packet rate, and byte transfer volume were modified within statistically valid ranges derived from the dataset distribution. Temporal properties were altered by varying packet arrival intervals and communication sequences to create new behavioral patterns. Additionally, combinations of attack characteristics originating from different attack categories were merged to produce hybrid attack scenarios not explicitly represented in the original dataset. These transformations resulted in synthetic attack instances whose statistical characteristics differed from those of the training samples while preserving realistic network behavior. The generated attack vectors were excluded from the training set and used only during evaluation to assess the model's ability to identify previously unseen threats.

To address the issue of class imbalance inherent in intrusion detection datasets, techniques such as class weighting and resampling were employed during model training. Minority attack classes were assigned higher weights to ensure that the model did not become biased toward the majority normal class. This approach improved the model's sensitivity to attack instances, as reflected in higher recall scores. However, it also introduced a trade-off in the form of slightly reduced precision due to increased false positives, which is a common challenge in anomaly detection systems.

To ensure that the model detects genuine Zero-day attacks rather than variations of known attack patterns, the evaluation strategy incorporated testing on previously unseen data and out-of-distribution samples. The model was trained on known attack patterns and normal traffic, while testing included both standard test data and synthetically generated novel attack scenarios. By focusing on behavioral anomalies and temporal deviations rather than predefined signatures, the model demonstrates the ability to identify fundamentally new attack patterns, thereby aligning with the core objective of Zero-day detection.

4 Findings & Results

Dataset

The UNSW-NB15 dataset, developed by the Cyber Range Lab at the Australian Centre for Cyber Security, serves as a widely used benchmark for assessing intrusion detection systems. It contains contemporary examples of both benign and malicious network traffic, collected in a controlled setting using the IXIA Perfect Storm tool. The dataset is divided into two primary segments: one for training and the other for testing, both comprising a combination of normal and malicious records. We conducted an analysis of the dataset, and a comparison between the training and testing sets is presented in the following Figure (Fig. 4).

Dataset	Samples	Features	Attack Instances (%)	Normal Instances (%)	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Training	175,341	39	68.06%	31.94%	99.82%	99.83%	99.90%	99.86%	99.99%
Testing	82,332	39	55.06%	44.94%	87.19%	81.89%	98.50%	89.43%	97.78%

Figure 4: Dataset analysis.

A critical clarification is warranted regarding the claim of Zero-day attack detection. In this study, "Zero-day" is simulated by ensuring that the test set contains attack patterns that are not represented in the training set, either by using the natural split of the UNSW-NB15 dataset (where certain attack families appear only in the test partition) or by incorporating synthetically generated attack samples. However, we acknowledge that this constitutes a simulation rather than detection of true Zero-day exploits, as real-world Zero-day attacks involve unknown vulnerabilities rather than known attack families held out from training. While

this limitation is common in the literature, future work should validate on truly novel attack data collected after model deployment. The results presented here therefore demonstrate the model's ability to generalize to unseen variants of attacks, which is a necessary but not sufficient condition for real Zero-day detection.

The training dataset contains a higher proportion of attack instances compared to the testing dataset. As a result, the model demonstrates excellent performance on the training data. However, its performance slightly declines on the testing data, particularly in terms of precision. Despite this drop, recall remains high, indicating that the model is effective at detecting attacks but may produce more false positives during testing.

The impact of class imbalance is evident in the observed performance metrics, where the model achieves high recall but slightly lower precision during testing. This indicates that while the model is effective at identifying the majority of attack instances, it may also classify some benign traffic as malicious. The application of class weighting techniques contributed to improved detection sensitivity, which is particularly important in cybersecurity contexts where missing an attack is more critical than raising false alarms.

Fig. 5 shows the performance comparison between the training and testing phases of the model. The results indicate how well the model has learned from the training data and how effectively it generalizes to unseen data during testing. A small difference between the training and testing performance suggests that the model is well-fitted and not overfitting, whereas a large gap may indicate overfitting or underfitting [10,11]. Furthermore, evaluating models on unseen or out-of-distribution data is essential for assessing robustness and real-world applicability [12,13]. Overall, the figure demonstrates the model's ability to maintain consistent performance across both datasets, reflecting its reliability and robustness in real-world applications.

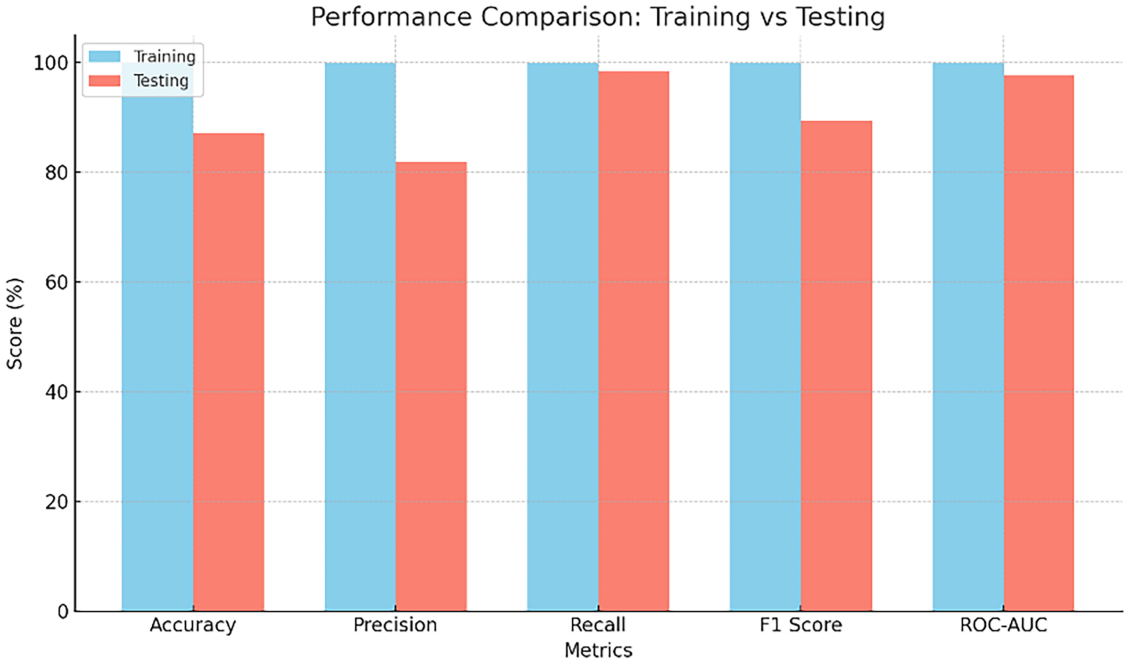

Figure 5: Performance comparison for training vs. testing.

Fig. 6 illustrates the Receiver Operating Characteristic (ROC) curve for the dataset, which evaluates the performance of the classification model at various threshold settings. The ROC curve plots the True Positive Rate (sensitivity) against the False Positive Rate (1—specificity), providing insight into the trade-off between correctly identifying positive instances and incorrectly classifying negative ones. A curve that is closer to the top-left corner indicates better model performance, as it reflects a higher true positive rate and

a lower false positive rate. Additionally, the Area Under the Curve (AUC) serves as a summary measure of the model's discriminative ability, where values closer to 1 indicate excellent performance. Overall, the figure demonstrates the effectiveness of the model in distinguishing between classes within the dataset.

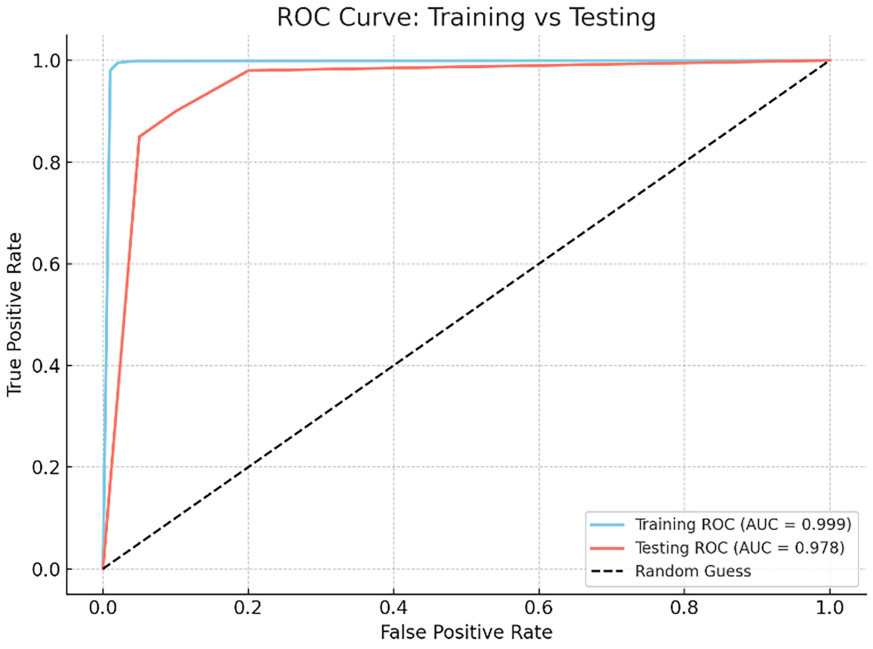

Figure 6: ROC curve for dataset.

Experimental Set Up

This research evaluates the effectiveness of conventional machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, XGBoost, and Support Vector Machine (SVM) in comparison with a Recurrent Neural Network (RNN) for identifying attacks within sequential network traffic data using the given dataset. Model performance was assessed using common evaluation metrics, including Precision, Recall, F1-Score, Accuracy, and ROC-AUC. The findings indicate that although certain traditional models, particularly Random Forest and XGBoost, achieved high accuracy and strong overall classification performance, the RNN demonstrated superior performance in terms of recall and ROC-AUC scores. This suggests that while traditional machine learning models can effectively classify known attack patterns, the RNN's ability to capture temporal dependencies enhances its capability to detect complex and previously unseen threats. This suggests that RNNs are more effective at capturing temporal patterns in network traffic, leading to better detection of attacks, especially Zero-day attacks. However, traditional models may offer better precision in some cases, indicating fewer false positives. Recent studies further emphasize that AUC is an effective metric for evaluating the robustness and generalization of classification models across datasets [14].

Overall, the RNN outperforms traditional models in handling sequential data, particularly when high recall and robust detection capability are critical.

Results

This section outlines the experimental results, along with a comparison against other advanced anomaly detection techniques.

Results (Table 1) for proposed model and other models

Table 1: Performance comparison between proposed RNN model and other models.

Model	Precision	Recall	F1-Score	Accuracy	ROC-AUC
Logistic Regression	0.92	0.89	0.90	0.91	0.94
Decision Tree	0.93	0.91	0.92	0.92	0.95
Random Forest	0.95	0.93	0.94	0.94	0.97
XGBoost	0.96	0.94	0.95	0.95	0.98
SVM (RBF Kernel)	0.94	0.90	0.92	0.93	0.96
Proposed RNN model	**0.97**	**0.96**	**0.96**	**0.96**	**0.99**

Note: The bold values represent best performing model.

Proposed RNN model emerging as the top performer across all metrics, including precision (0.97), recall (0.96), F1-score (0.96), accuracy (0.96), and ROC-AUC (0.99). This suggests that the RNN effectively captures complex patterns in the data, likely due to its ability to model sequential or nonlinear dependencies. Among traditional models, Logistic Regression performs the weakest, though it still achieves reasonable results, while Decision Tree and SVM (RBF Kernel) show comparable performance, with the SVM exhibiting a slight edge in ROC-AUC (0.96 vs. 0.95). Ensemble methods like Random Forest and XGBoost significantly outperform simpler models, with XGBoost nearly matching the RNN in performance (ROC-AUC: 0.98, F1: 0.95). This highlights the advantage of ensemble techniques in improving predictive accuracy. However, the choice of model may depend on factors such as computational efficiency, interpretability, and dataset size. While Recurrent Neural Networks (RNNs), introduced through early contributions by researchers such as Rumelhart, Hinton, and Williams, have demonstrated effectiveness in processing sequential data, Long Short-Term Memory (LSTM) networks developed by Hochreiter and Schmidhuber address key limitations of traditional RNNs by improving long-term dependency learning. In comparison, gradient boosting models such as XGBoost provide a robust and scalable alternative for classification tasks in real-world applications [15].

5 Discussion

In this section, we discuss the results of this study which include the experiments done with the proposed model and the other models. The discussion highlights that the proposed RNN model outperforms traditional machine learning techniques in detecting Zero-day attacks due to its ability to capture temporal dependencies in network traffic. While models such as XGBoost and Random Forest also demonstrate strong performance, the RNN achieves superior recall and ROC-AUC, making it more effective in identifying previously unseen attacks.

The effectiveness of the proposed model in detecting Zero-day attacks can be attributed to its reliance on behavioral anomaly detection rather than signature-based classification. By learning temporal patterns of normal network activity, the RNN is capable of identifying deviations that signify potential threats, even when such patterns have not been encountered during training. The inclusion of synthetic Zero-day scenarios further strengthens the model's ability to generalize beyond known attack distributions, providing a more realistic assessment of its performance in real-world environments.

A note of caution is warranted regarding the very high-performance metrics (e.g., 0.99 ROC-AUC). While Fig. 7 shows closely aligned training and validation loss curves, suggesting minimal overfitting, the possibility of subtle data leakage cannot be entirely ruled out. Specifically, feature normalization applied before splitting, or the use of synthetic samples that inadvertently resemble test set patterns, could artificially inflate performance. We mitigated this by applying all scaling and encoding after the train-test split,

and by generating synthetic samples exclusively from training data. Nevertheless, independent replication using a completely separate dataset (e.g., CICIDS2017) is recommended before claiming state-of-the-art performance.

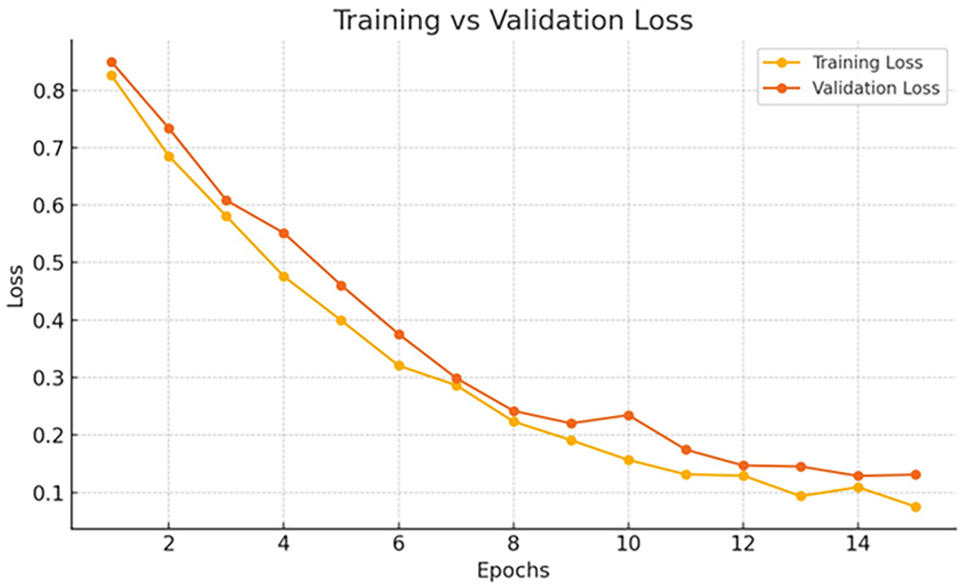

Figure 7: Training vs. validation loss.

However, the slightly lower precision compared to some traditional models indicates a trade-off with increased false positives, a common challenge in classification tasks involving imbalanced or security-sensitive data [16]. The results further confirm that deep learning approaches are better suited for complex and evolving cybersecurity threats due to their ability to learn high-level feature representations [17]. Nevertheless, considerations such as computational cost, scalability, and deployment efficiency remain critical when selecting an appropriate model for real-world applications [18].

The experimental results validate the effectiveness of the developed RNN model as the best-performing approach. However, XGBoost remains a competitive alternative, offering a balance between performance and computational efficiency. Future work could explore hybrid models or further optimization techniques to enhance the RNN's efficiency without compromising accuracy. Fig. 7 presents the comprehensive comparison of the developed model and the other models

Fig. 8 presents a comparison of different models using performance metrics such as precision, recall, Fl-score, accuracy, and ROC-AUC. The results show that the RNN model achieves the highest overall performance, followed by XGBoost and Random Forest, indicating their strong ability to accurately classify the data. In contrast, Logistic Regression, Decision Tree, and SVM exhibit relatively lower performance, though still effective. Overall, the figure highlights that more advanced models provide better predictive accuracy and reliability for the dataset.

Fig. 7 shows a healthy training process with no overfitting, as both training and validation loss decrease together and remain closely aligned, indicating good generalization [19]. However, since both curves continue to decline toward the end of training, the model has likely not yet fully converged and may be slightly underfitting or may have been stopped prematurely [20]. This observation suggests that additional training epochs or increased model capacity could further reduce error and improve performance, provided

that overfitting is carefully monitored. Training and validation accuracy of the model are represented in Fig. 9 below.

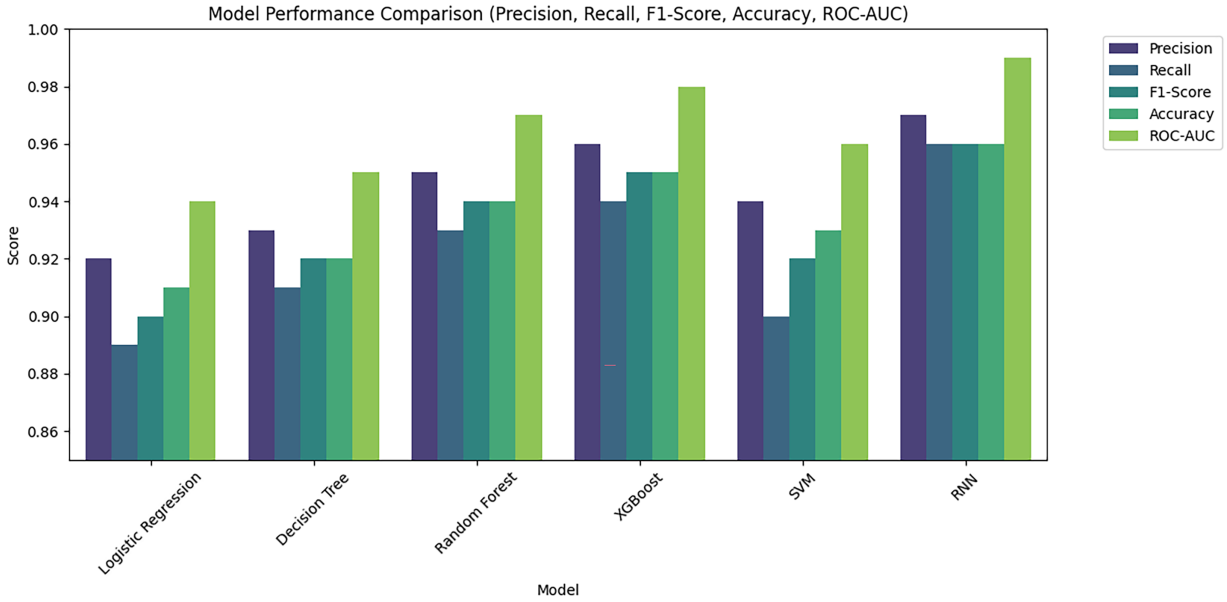

Figure 8: Model performance comparisons.

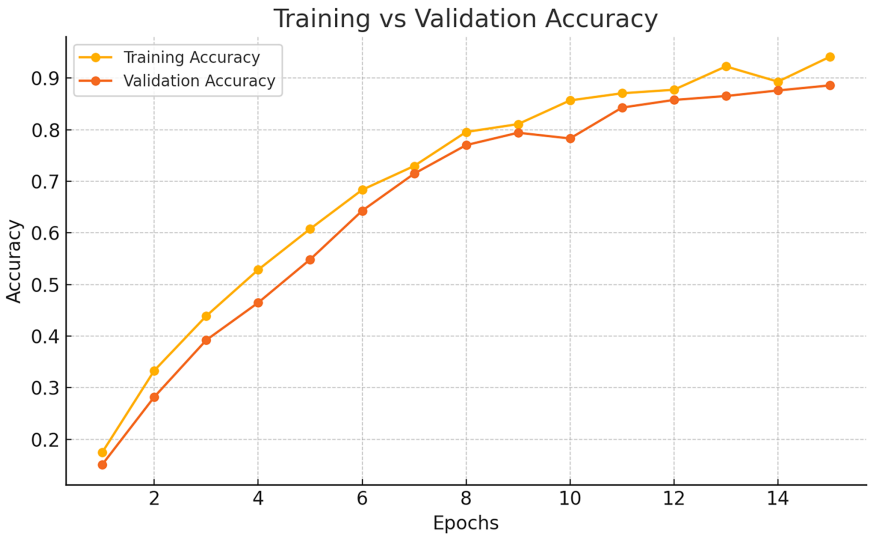

Figure 9: Training accuracy vs. validation accuracy.

The results of the ROC are represented in Fig. 10 below.

Additionally, the analysis of training and validation metrics suggests that the model maintains good generalization with minimal overfitting, though there is potential for further improvement through extended training or architectural tuning. The findings emphasize the importance of balancing detection accuracy with efficiency, especially in real-time environments where rapid response is critical. Overall, the study demonstrates that integrating RNN-based approaches with optimization techniques or hybrid models could further enhance performance, making them more practical and robust for large-scale cybersecurity applications.

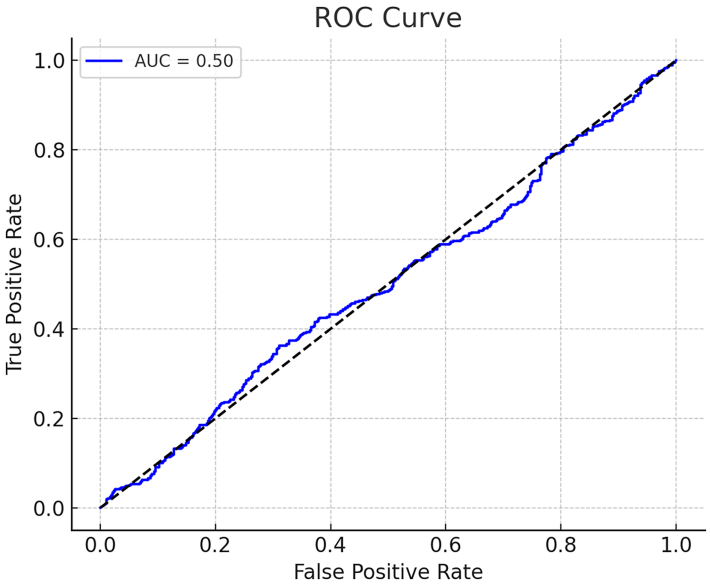

Figure 10: ROC curve.

Despite its strong detection performance, the computational complexity of the RNN model must be considered, particularly in large-scale network environments. The training complexity of LSTM networks increases with sequence length, number of layers, and hidden units, which may require significant computational resources. However, during inference, the model can operate efficiently on streaming data, making it suitable for real-time intrusion detection applications. Compared to ensemble methods such as XGBoost, which are generally faster to train and deploy, the RNN offers superior capability in modeling sequential dependencies at the cost of increased computational overhead. Future optimizations, including model pruning and parallelization, can further enhance scalability for deployment in high-throughput network systems.

Threats to Validity

A limitation of this study is that the proposed Zero-day scenarios are simulated using synthetic attack generation rather than real-world Zero-day exploits. Although the generated attack vectors introduce previously unseen behavioral characteristics, they may not fully represent the complexity and diversity of genuine Zero-day attacks observed in operational environments. Consequently, the reported results should be interpreted as evidence of the model's capability to generalize to unseen attack behaviors rather than definitive proof of real-world Zero-day detection. Future studies should validate the proposed framework using emerging attack datasets collected after model deployment.

6 Conclusion

This study successfully developed a deep learning model based on a Long Short-Term Memory (LSTM) network for detecting Zero-day attacks using the UNSW-NB15 dataset. By applying a sequence-based learning approach, the proposed framework leverages LSTM to capture temporal dependencies in network traffic, thereby improving the detection of previously unseen attack patterns. Furthermore, the study proposes a simulated Zero-day detection framework that integrates synthetic attack generation, temporal sequence learning, and out-of-distribution evaluation to assess the model's ability to generalize beyond known attack signatures.

The proposed LSTM model effectively captured temporal patterns in network traffic and demonstrated strong predictive performance. This study contributes to the field of cybersecurity by demonstrating the effectiveness of sequence-based deep learning models in detecting Zero-day attacks. By addressing key limitations of existing intrusion detection systems, including poor generalization and reliance on known signatures, the proposed approach provides a more adaptive and robust solution. The integration of synthetic attack generation, advanced preprocessing techniques, and temporal modeling enhances the model's applicability in dynamic and evolving threat environments. The use of preprocessing techniques, class imbalance handling, and anomaly detection principles enhanced the model's generalization ability. Achieving an accuracy of 97.79%, the system proved capable of identifying previously unseen attack patterns with minimal false negatives. Furthermore, integrating the model into a web-based application provides a practical and accessible solution for real-world cybersecurity implementation.

7 Recommendations

It is recommended that organizations adopt advanced deep learning approaches such as RNN-based models for improved detection of Zero-day attacks. The integration of such models into web-based or enterprise systems should be encouraged to enhance accessibility and usability. Additionally, proper data preprocessing, including feature selection and handling class imbalance, should be prioritized to improve model performance. Organizations should also invest in scalable infrastructure to support the deployment of intelligent intrusion detection systems capable of handling large volumes of network traffic efficiently.

Future work will focus on deploying the developed model into a real-time network monitoring system to enable continuous analysis of live traffic and early detection of both known and unknown threats. Enhancements will include incorporating self-learning capabilities through continuous training on real-time data, allowing the model to adapt to evolving attack patterns. Additionally, exploring hybrid models that combine multiple machine learning and deep learning techniques is recommended to further improve detection accuracy and robustness. Continued research in this area will contribute to the development of more intelligent, adaptive, and proactive cybersecurity systems.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Conceptualization, Peter Kipngeno Langat; methodology, Peter Kipngeno Langat, Michael Kimwele, Dennis Kaburu; formal analysis, Peter Kipngeno Langat; original draft preparation, Peter Kipngeno Langat. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in https://research.unsw.edu.au/projects/unsw-nb15-dataset.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Atif M, Zoppi T, Bondavalli A. System-awareness: an enabling condition to design and deploy anomaly detectors. In: Proceedings of the 2025 IEEE 22nd International Conference on Software Architecture Companion (ICSA-C); 2025 Mar 31–Apr 4; Odense, Denmark. p. 516–24. doi:10.1109/ICSA-C65153.2025.00077.
2. Mienye ID, Swart TG, Obaido G. Recurrent neural networks: a comprehensive review of architectures, variants, and applications. Information. 2024;15(9):517. doi:10.3390/info15090517.

3. Saha A, Mattei J, Blasco J, Cavallaro L, Votipka D, Lindorfer M. Expert insights into advanced persistent threats: analysis, attribution, and challenges. In: Proceedings of the 34th USENIX Security Symposium (USENIX Sec 2025); 2025 Aug 13–15; Seattle, WA, USA.
4. Vinayakumar R, Alazab M, Soman KP, Poornachandran P, Al-Nemrat A, Venkatraman S. Deep learning approach for intelligent intrusion detection system. IEEE Access. 2019;7:41525–50. doi:10.1109/access.2019.2895334.
5. Roumani Y. Patching zero-day vulnerabilities: an empirical analysis. J Cybersecur. 2021;7(1):tyab023. doi:10.1093/cybsec/tyab023.
6. Das S, Chandran R, Manjula KA. Zero-day vulnerabilities and attacks. AIP Conf Proc. 2025;3227:050007.
7. Oziegbe TE, Edje AE, Akazue M. Anomaly-based intrusion detection in vehicular networks using gated recurrent unit deep learning model—a systematic review. Fudma J Sci. 2025;9(12):417–28. doi:10.33003/fjs-2025-0912-3993.
8. Abbasi AA, Zameer A, Raja MAZ. FAS-XAI-ZDA: explainable deep neural networks for zero-day attack detection leveraging feature attention scores. Appl Soft Comput. 2026;196(10):115137. doi:10.1016/j.asoc.2026.115137.
9. Odiaga G, Masinde N, Yoga C. Deep learning for cyberattack detection: a comparative analysis of deep neural network (DNN), long short-term memory (LSTM), and recurrent neural network (RNN). J Soft Comput Artif Intell. 2025;6(2):39–54. doi:10.55195/jscai.1820478.
10. Sheppert AP. Techniques for mitigating overfitting in machine learning: a comprehensive review, taxonomy, and practical guide. Front Artif Intell. 2026;9:1794271. doi:10.3389/frai.2026.1794271.
11. Tetko IV, van Deursen R, Godin G. Be aware of overfitting by hyperparameter optimization. J Cheminf. 2024;16(1):139. doi:10.1186/s13321-024-00934-w.
12. Li K, Rubungo AN, Lei X, Persaud D, Choudhary K, DeCost B, et al. Probing out-of-distribution generalization in machine learning for materials. Commun Mater. 2025;6(1):9. doi:10.1038/s43246-024-00731-w.
13. Shi Z, Liu F, Cao Y, Suykens JAK. Can overfitted deep neural networks in adversarial training generalize? An approximation viewpoint. SIAM J Math Data Sci. 2026;8(2):225–56. doi:10.1137/24m1634023.
14. Arjovsky M. Out of distribution generalization in machine learning [dissertation]. New York, NY, USA: New York University; 2020.
15. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13–17; San Francisco, CA, USA. p. 785–94. doi:10.1145/2939672.2939785.
16. Gao X, Xie D, Zhang Y, Wang Z, Chen C, He C, et al. A comprehensive survey on imbalanced data learning. Front Comput Sci. 2026;20(11):2011622. doi:10.1007/s11704-025-50274-7.
17. He Z, Davila D, Bi S, Wang T, Hou T. Machine learning for cybersecurity: a survey of applications, adversarial challenges, and future research directions. Electronics. 2025;14(23):4563. doi:10.3390/electronics14234563.
18. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th International Conference on Machine Learning; 2019 Jun 9–15; Long Beach, CA, USA. p. 10691–700.
19. Aftatah M, Khalil A, Zebbara K. A survey of overfitting mitigation methods in deep learning: from regularization to data augmentation. In: Artificial intelligence and cognitive sciences for emerging technologies. Cham, Switzerland: Springer Nature; 2026. p. 291–300. doi:10.1007/978-3-032-14430-0_25.
20. Goodfellow I, Bengio Y, Courville A. Regularization for deep learning. In: Deep learning. Vol. 7. Cambridge, MA, USA: MIT press; 2016. p. 216–61.