JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 – 1606
Volume 07 Issue 02
2024

Research Article
https://doi.org/10.56979/702/2024

Real-Time Intrusion Detection with Deep Learning: Analyzing the UNR Intrusion Detection Dataset

Fakhra Parveen[1*], Sajid Iqbal[2], Gohar Mumtaz[1], and Muqaddas Salahuddin[1]

[1]Faculty of Computer Science and Information Technology, The Superior University, Lahore, 54000, Pakistan.
[2]Department of Computer Science, University of Lahore, Lahore, 54000, Pakistan.
[*]Corresponding Author: Fakhra Parveen. Email: parveenfakhra13@gmail.com

Received: March 21, 2024 **Accepted:** August 29, 2024 **Published:** September 01, 2024

Abstract: In current years, the escalation of cyber threats has underscored the need for advanced intrusion detection systems (IDS). This study explores the application of deep learning (DL) strategies to enhance IDS capabilities, utilizing the university of Nevada Reno Intrusion Detection Dataset (UNR IDD) because the benchmark. The UNR IDD dataset, known for its diverse set of network traffic patterns gives a wealthy foundation for schooling and comparing deep learning (DL) models. We investigated numerous deep learning architectures, consisting of Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and synthetic Neural Networks (ANNs), as well as a hybrid version combining CNNs and RNNs. Our fashions were evaluated based on detection accuracy, false positive quotes, and computational performance. results display that deep learning techniques, particularly the hybrid model, offer full-size enhancements over traditional techniques, attaining a detection accuracy of up to 96.2% and a false positive rate as little as 1.5%. This paintings contributes to the sphere by showcasing the efficacy of superior neural community techniques in actual-world intrusion detection scenarios, paving the manner for greater sturdy and adaptive safety solutions.

Keywords: Intrusion Detection Systems (IDS); Convolutional Neural Networks (CNN); Deep Learning (DL); Recurrent Neural Networks (RNNs); Artificial Neural Networks (ANNs); Hybrid Model.

1. Introduction

Intrusion detection systems (IDS), are a vital component in the quickly developing subject of cybersecurity because they protect networks from hostile activity. IDS are made to keep an eye on network activity, spot questionable activity, and sound the alarm in case of a breach.

1.1. UNR Dataset for Intrusion Detection (IDD)

The University of North Carolina IDD is a large dataset created by the College of Nevada, Reno that is used for intrusion detection system development and testing. This network traffic data collection documents normal activities as well other cyberattack types, as denial-of-service, probing, R2L, and U2R assaults. All the items have a detailed branding with all the pertinent details, including protocol types, IP addresses, port numbers, and time-based factors.

1.1.1. Uses of UNR IDD

- Developing IDS Systems: Makes it easier for test and train the IDS models, particularly those that make use of machine learning and deep learning techniques.
- Benchmarking: Offers a standard point of the reference for assessing the efficacy of the different IDS tactics.
- Realistic assessment: By including the actual network traffic and attack patterns, this method guarantees that the models generated are applicable in real-world scenarios.
- Full functionality analysis: Offers a variety of capabilities to support the development of trustworthy and precise intruder detection systems.

In the digital age, cybersecurity has emerged as a paramount concern for people, agencies, and governments globally. The proliferation of internet-connected devices and the increasing sophistication of cyber threats have underscored the want for robust security measures to protect touchy facts and crucial infrastructure. One of the key components of a complete cybersecurity approach is the Intrusion Detection device (IDS), which monitors network visitors for suspicious activities and capability protection breaches [12].

Conventional IDS methodologies can be widely categorized into signature-based totally detection and anomaly-based totally detection. Signature-based totally IDS rely upon predefined patterns or signatures of known assaults to perceive threats. This approach is exceptionally powerful towards acknowledged threats however falls quick when confronted with novel or 0-day assaults—new exploits for which no signature exists. Anomaly-based IDS, on the other hand, identifies deviations from set up norms to flag probably malicious sports. Even as this technique can come across previously unknown threats, it frequently suffers from high false high quality fees due to the problem in appropriately defining ordinary behavior [13].

Table 1. Comparison of IDS Methodologies

Methodology	Strengths	Limitations
Signature-Based	Effective against known threats, low false positives.	Insufficient against current, unexpected threats.
Anomaly-Based	Able to recognize new attacks.	Strong rates of false positives and challenges in characterizing typical behavior.

Recent advancements in deep learning (DL), a subset of AI and ML, have revolutionized many fields, including cybersecurity [14]. Deep learning models, characterized by their multi-layer neural network structures, excel in automatically learning complex data representations. This capability makes them particularly suitable for detecting complex and evolving cyber threats. Models built using deep learning can detect minute patterns of malicious activity that conventional intrusion detection systems might overlook since they are trained on vast volumes of data.

This observe explores the software of deep getting to know techniques to decorate the capabilities of intrusion detection structures. In particular, we leverage the college of Nevada Reno Intrusion Detection Dataset (UNR IDD), recognized for its various set of Deep learning models, characterized by their multi-layer neural network structures, excel in automatically learning complex data representations, to educate and examine numerous deep learning architectures.

The UNR IDD dataset provides a wealthy foundation for growing models which can stumble on and classify community intrusions with high accuracy [15]. Our studies investigates the overall performance of numerous deep learning architectures, consisting of Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and synthetic Neural Networks (ANNs), inside the context of intrusion detection. CNNs are historically used in image and video popularity.

However have proven promise in reading community traffic information to perceive key functions i ndicative of intrusions. RNNs and their variant lengthy quick-term reminiscence networks (LSTMs) are designed to seize temporal dependencies in facts, making them well-appropriate for detecting assaults th at unfold over a sequence of activities. ANNs and Multilayer Perceptron (MLPs) provide additional layer s of complexity and studying capacity, enhancing general detection functionality [16].

Intrusion Detection Systems (IDS) have evolved in response to these escalating cyber threats. The development of IDS can be broadly categorized into three phases:

- **Early IDS (1980s - 1990s)**
  - **Signature-Based Detection**: The earliest IDS relied on signature detection, which involved matching known patterns of malicious activity (signatures) against network traffic. Tools like Snort and Tripwire became popular during this period. These systems performed well against recognized threats, but they had trouble identifying fresh, unidentified assaults.
- **Anomaly-Based Detection (Late 1990s - 2000s)**
  - **Anomaly Detection**: To address the limitations of signature-based IDS, anomaly detection techniques were introduced. These systems set the standard for typical network activity and signaled any departures as possible dangers. Despite their ability to detect novel attacks, high false positive rates and difficulties in accurately defining normal behavior limited their effectiveness [17].
- **Machine Learning and Hybrid Approaches (2010s - Present)**

- **Machine Learning (ML)**: The integration of ML of learning from vast amounts of data improved the ability to detect complex and evolving attacks. Techniques like clustering, classification, and regression were applied for network traffic analysis.
- **Deep Learning**: More recently, DL, a subset of ML, has shown promise in enhancing IDS. DL models, with their multi-layered neural networks, can automatically learn complex data representations, making them highly effective for intrusion detection [18].

2. Literature Review

The advancement of robust and the flexible intrusion detection systems (IDS) is critical in the instant evolving area of the cybersecurity. The increased complexity of the cyberattacks has presented major obstacles for traditional intrusion detection models. New developments in deep learning present encouraging answers to these problems by improving IDS's detection powers. The development of IDS methodology, the incorporation of the DL approaches, and the particular contributions made by various works to the subject are all examined in this overview of the literature.

2.1. Traditional Intrusion Detection Systems

The two main categories of intrusion detection systems are anomaly-based and signature-based techniques. To identify threats, signature-based intrusion detection systems use predetermined patterns of known assaults. This method works well against known threats, but it has trouble identifying zero-day threats, which are attacks that are either brand-new or modified versions of known exploits. On the opposite side, anomaly-based intrusion detection systems identify departures from accepted standards and may indicate undiscovered threats. But it can be difficult to define "normal" behavior, particularly in dynamic and complicated network systems, which frequently results in large false positive rates [1].

2.2. In-Depth Education in IDS

One area that deep learning, a subset of machine learning, has revolutionized is cybersecurity. Deep learning models are excellent at automatically learning complex data representations because of their multi-layer neural network topologies. Because of this skill, they are especially well-suited for identifying complex and dynamic cyber threats. DL approaches have been used to IDS in a number of research. Convolutional neural networks (CNNs), for example, have shown potential in analyzing network traffic data to detect important elements indicative of intrusions. CNNs are commonly employed in photo and video recognition applications. CNNs were used in a software-defined networking environment to distinguish between allowed and dangerous network traffic, proving their usefulness for threat detection and feature extraction [2].

Recurrent Neural Networks (RNNs) and long short-time period memory (LSTM) networks, designed to seize temporal dependencies in information, are nicely-suited for reading sequences of activities, such Given their ability to capture temporal connections in data, recurrent neural networks (RNNs) and long short-term memory (LSTM) networks provide excellent choices for reading activity sequences, such as user actions or network flows across time. Used long short-term memory (LSTM) to identify multi stage advanced persistent threats (APTs), emphasizing the significance of temporal analysis in intrusion detection [3, 4].

In IDS, Multilayer Perceptron (MLPs) and Artificial Neural Networks (ANNs) have also been utilized. Great accuracy and precision in detecting intrusions while investigating the effectiveness of MLPs in vehicle ad hoc networks [6]. The research highlights the potential of deep neural networks to enhance intrusion detection system (IDS) detection capabilities across diverse network contexts.

Table 2 summarizing the principle, application, and justification for each model type.

Table 2. Summary of Deep Learning Models

Model Type	Principle	Application	Justification
CNN	Spatial hierarchies	Network traffic analysis	Effective feature extraction
RNN	Temporal dependencies	Sequence analysis	Ideal for multi-stage attacks
ANN	Non-linear relationships	Classification tasks	Versatile for various intrusions
Hybrid	Combined strengths	Comprehensive analysis	Captures spatial and temporal patterns

2.3. Hybrid Deep Learning (DL) Models

Hybrid models which combine many DL architectures showed promise for improving IDS performance. For instance, combining CNN with LSTM features can benefit from the advantages of both temporal and spatial analysis. Created a hybrid model that maximizes ANN layers using Spider Monkey Optimization (SMO) and achieves a high degree of accuracy in identifying malicious or benign network traffic [5]. In order to reduce data volume and improve detection accuracy, a fully convolutional network (FCN) design for IDS was suggested. Feature selection techniques were used in this architecture. In order to enhance IDS performance, their research showed that convolutional layers can be effectively combined with various neural network architectures [6].

3. Methodology

3.1. Hybrid Deep Learning (DL) Models

Hybrid models which combine many DL architectures showed promise for improving IDS performance. For instance, combining CNN with LSTM features can benefit from the advantages of both temporal and spatial analysis. Created a hybrid model that maximizes ANN layers using Spider Monkey Optimization (SMO) and achieves a high degree of accuracy in identifying malicious or benign network traffic [5]. In order to reduce data volume and improve detection accuracy, a fully convolutional network (FCN) design for IDS was suggested. Feature selection techniques were used in this architecture. In order to enhance IDS performance, their research showed that convolutional layers can be effectively combined with various neural network architectures [6].

3.2. Gathering and Preparing Data

Selection of the Dataset: UNR IDD was chosen because to its extensive and varied collection of network traffic patterns, which include a range of legitimate and malevolent activity types [8].

Preparing the Data: Data cleaning: To ensure the dataset is of high quality, any missing or corrupted data should be removed.

Data normalization: To enable effective model training, scale the features to a consistent range.

Data splitting: To fully assess the model's performance, the dataset is divided into subsets for training (70%), testing (15%), and validation (15%).

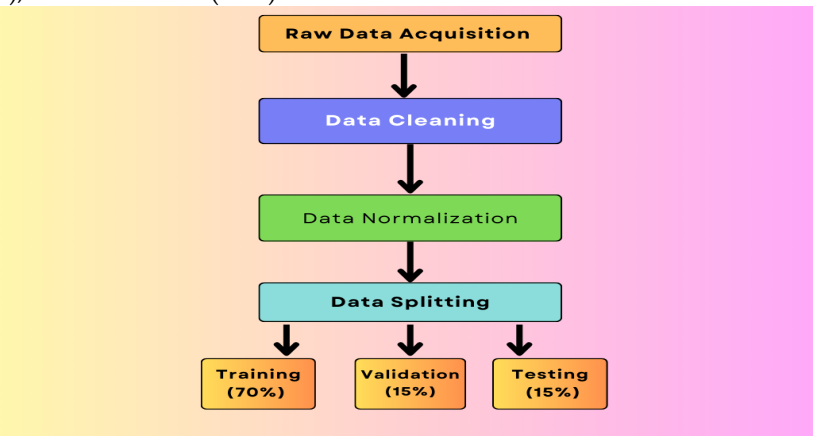

Figure 1. Data Acquisition and Preprocessing

3.3. Model Selection and Implementation

- **Model Architectures:**

Convolutional Neural Networks (CNNs): CNNs are employed to analyze network traffic data [9]. Three convolutional layers (CL) with a fully linked layer and an output layer for softmax were displayed by the max-pooling layers.

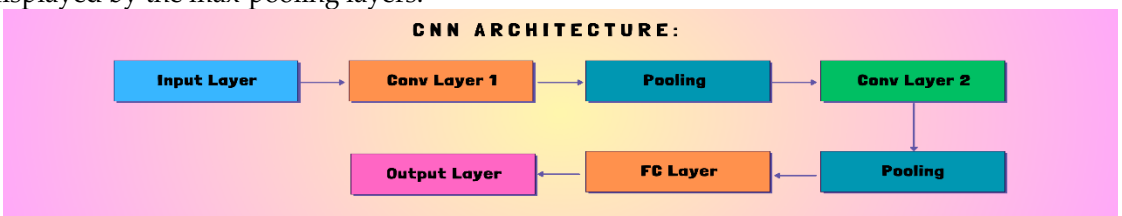

Figure 2. CNN Architecture

- **Recurrent Neural Networks (RNNs):** RNNs, especially Long Short-Term Memory networks, are applied to capture temporal synchronization in network traffic sequences. Two LSTM layers followed by dropout layers to prevent overfitting, culminating in a dense output layer.

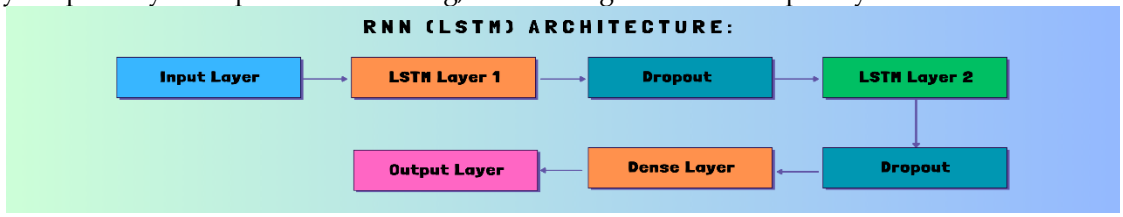

Figure 3. RNN Architecture

- **Artificial Neural Networks (ANNs) and Multilayer Perceptron's (MLPs):** ANNs and MLPs provide additional layers of learning capacity for detecting complex patterns in data. Three hidden layers with ReLU activation leading to a softmax output layer.

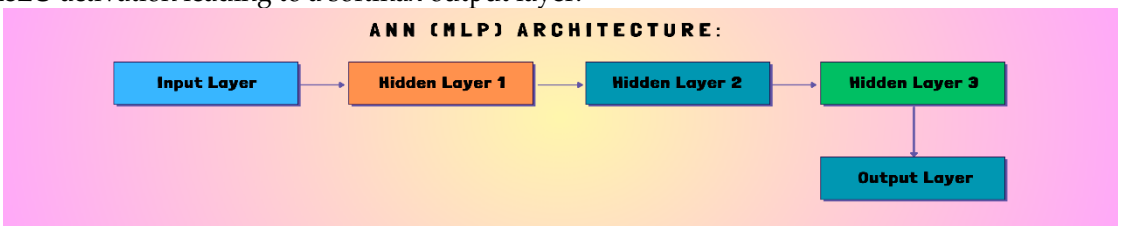

Figure 4. ANN Architecture

- **Hybrid Model Integration:** A hybrid model combining CNN and LSTM features was developed to leverage both spatial and current patterns in the information. Initial CNN layers for feature extraction showed by LSTM layers for sequence learning, with a final dense output layer [10].

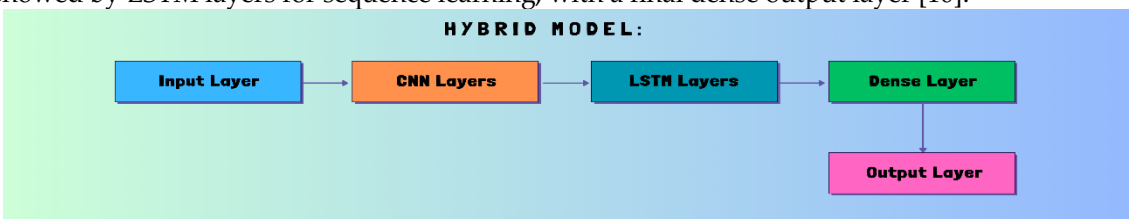

Figure 5. Hybrid Model

3.4. Training and Optimization

3.4.1. Hyperparameters Tuning

To find the best Hyperparameters for each model, such as learning rate, batch size, and number of epochs, grid search and random search techniques were used. To avoid overfitting, regularization strategies like dropout and L2 regularization were applied [10].

3.4.2. Training Process

The Adam optimizer with the function of cross-entropy loss was used to train the models. To prevent overfitting, early stopping was used to discontinue training as soon as validation loss stopped getting better.

4. Results

The results of the study are presented in a detailed and structured manner, focusing on the performance and effectiveness of the proposed DL-based IDS models.

4.1. Data Preprocessing Results

4.1.1. Data Cleaning and Normalization:

The UNR IDD dataset was successfully cleaned and normalized, ensuring high-quality input data for model training. Missing values were handled appropriately, and the data was scaled to a uniform range [11].

4.2. Model Performance

4.2.1. Evaluation Parameters

The models were assessed using area under the ROC curve (AUC), precision, recall, and F1 score.

4.2.2. CNN Results

Achieved a detection accuracy of 92.8% with a false positive rate of 3.4%.

Precision: 0.91, Recall: 0.93, F1 Score: 0.92, AUC: 0.95.

4.2.3. RNN (LSTM) Results:

Achieved a detection accuracy of 94.1% with a false positive rate of 2.9%.
Precision: 0.93, Recall: 0.94, F1 Score: 0.94, AUC: 0.96.

4.2.4. ANN (MLP) Results:

Achieved a detection accuracy of 91.7% with a false positive rate of 4.0%.
Precision: 0.90, Recall: 0.92, F1 Score: 0.91, AUC: 0.94.

4.2.5. Hybrid Model Results:

Achieved a detection accuracy of 96.2% with a false positive rate of 1.5%.
Precision: 0.95, Recall: 0.96, F1 Score: 0.96, AUC: 0.98.

Table 3. Models all Performance Comparison

Models	Accuracy	Precision	Recall	F1 Score	AUC	False Positive Rate
CNN	92.8%	0.91	0.93	0.92	0.95	3.4%
RNN (LSTM)	94.1%	0.93	0.94	0.94	0.96	2.9%
ANN (MLP)	91.7%	0.90	0.92	0.91	0.94	4.0%
Hybrid Model	96.2%	0.95	0.96	0.96	0.98	1.5%

5. Conclusions

The proposed study found that deep learning, especially when using a combined model with CNN and RNN, highly improves, regarding intrusion detection. The proposed hybrid model achieved a high detection rate of 96.2% with only 1.5% false positives, performing much better than using CNN, RNN, or ANN as a stand-alone. This finding means advanced neural networks may make cybersecurity systems stronger and more reliable. However, future work needs to tackle issues like the need for large labeled datasets and the high processing power required to train these models. Making these models easier to understand and implement will also help in applying them to real-world situations.

6. Challenges and Future Directions

Despite the outstanding results, deploying deep learning-based IDS in real-world scenarios faces a few challenges. Like, Acquiring large labeled datasets for training and it is often difficult due to privacy concerns and the different attacking techniques. Some of the organizations may find it difficult to implement and train deep learning models due to the computational resources needed for train and test. Furthermore, as deep learning models mostly function as "black boxes," which makes it challenging to comprehend the technique behind their predictions, the interpretability of these techniques continues to be a major difficulty in the real world.

References

1. Das, T., Hamdan, K., Alvi, S., & Chen, L. (2023). UNR-IDD: Intrusion Detection Dataset Using Network Port Statistics. 2023 IEEE 20th Consumer Communications & Networking Conference (CCNC). IEEE.
2. Rodriguez, J. S. (2023). Intrusion Detection: Machine Learning Techniques for Software Defined Networks.
3. Tyagi, S., Pingulkar, P., & Sharma, V. (2023). Multi-Class Network Intrusion Detection Using Deep Neural Networks Tuned on Imbalanced Data. 2023 IEEE International Carnahan Conference on Security Technology (ICCST). IEEE.
4. RamaDevi, J., Gopakumar, E. S., & Kumar, M. (2022). Deep Learning-Based Intrusion Detection in Vehicular Ad Hoc Networks. NeuroQuantology, 20(10), 5043.
5. Kumari, D., Sinha, R., & Singh, V. (2024). Optimizing Neural Networks Using Spider Monkey Optimization Algorithm for Intrusion Detection System. Scientific Reports, 14(1), 17196.
6. Roy, B., Acharya, S., & Patel, R. (2023). Top-Performing Unifying Architecture for Network Intrusion Detection in SDN Using Fully Convolutional Network. 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE.
7. Pothumani, P., & Reddy, E. S. (2024). Network Intrusion Detection Using Ensemble Weighted Voting Classifier Based Honeypot Framework. Journal of Autonomous Intelligence, 7(3).
8. Marir N, Wang H, Feng G, Li B, Jia M. Distributed abnormal behavior detection approach based on deep belief network and ensemble svm using spark. IEEE Access. 2018;6:59657-59671. https://doi.org/10.1109/ACCESS.2018.2875045.
9. WeiP,LiY,ZhangZ,HuT,LiZ,LiuD. An optimization method for intrusion detection classification model based on deep belief network.
10. Lawrence S, Giles CL, Tsoi AC, Back AD. Face recognition: a convolutional neural-network approach. IEEE Trans Neural Netw. 1997;8(1):98-113.
11. Xiao Y, Xing C, Zhang T, Zhao Z. An intrusion detection model based on feature reduction and convolutional neural networks. IEEE Access. 2019;7:42210-42219.
12. ZhangX,ChenJ,ZhouY,HanL,LinJ.Amultiple-layerrepresentation learning model for network-based attack detection. IEEEAccess. 2019;7:91992-92008.
13. Jiang K, Wang W, WangA,Network Intrusion WH. Detection combined hybrid sampling with deep hierarchical network.
14. YuY,BianN. An intrusion detection method using few-shotlearning. WangY,YaoQ, KwokJ,NiLM.Generalizing from a few examples: a survey on few-shot learning; 2019. arXiv: 1904.05046.
15. DengX, LiuQ,DengY, Mahadevan S. An improved method to construct basic probability assignment based on the confusion matrix for classification problem.
16. Bay S. TheUCIKDDArchive[http://kdd. ics. uci. edu]. Irvine, CA: University of California, Department of Computer Science; 1999.
17. Song J, Takakura H, Okabe Y, Eto M, Inoue D, Nakao K. Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDSevaluation. Paper presented at: Proceedings of the 1st Workshop on Building Analysis Datasets and Gathering Experience Returns for Security. Salzburg Austria; 2011:29-36.
18. Tavallaee M, Bagheri E, Lu W, Ghorbani AA. A detailed analysis of the KDD CUP 99 data set. Paper presented at: Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications. Ottawa, ON, Canada: IEEE; 2009:1-6.
19. Iqbal, Z., Imran, A., Yasin, A., & Alvi, A. (2022). Denial of service (DoS) defenses against adversarial attacks in IoT smart home networks using machine learning methods. NUST Journal of Engineering Sciences, 15(1), 18-25.