International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V4I4P113
Eureka Vision Publication | Volume 4, Issue 4, 118-128, 2023

Original Article

Intelligent Network Traffic Identification Based on Advanced Machine Learning Approaches

Venkata Deepak Namburi[1], Aniruddha Arjun Singh Singh[2], Vaibhav Maniar[3], Vetrivelan Tamilmani[4], Rami Reddy Kothamaram[5], Dinesh Rajendran[6]

[1]University of Central Missouri, Department of Computer Science.
[2]ADP, Sr. Implementation Project Manager, aniruddha.
[3]Oklahoma City University, MBA / Product Management.
[4]Principal Service Architect, SAP America.
[5]California University of management and science, MS in Computer Information systems.
[6]Coimbatore Institute of Technology, MSC. Software Engineering.

Abstract - Network Traffic Analysis (NTA) refers to the process whereby the traffic of the network is logged and analyzed, to identify security risks or performance issues. There are, however, instances when machine learning (ML) is applied in the mechanization of NTA. Network information may be categorized, anomalies detected, and malicious actions identified using machine learning. They could also be employed in foreseeing the same traffic patterns in the future, which can be exploited and utilized to improve network performance. The article gives a recurrent neural network (RNN)-based intrusion detection system (IDS) on the NSL-KDD dataset. The methodology consists of a lengthy preprocessing stage, which involves treating missing values, performing dimensionality reduction, applying one-hot encoding, normalizing the data, selecting features, as well as dividing the training and test sets. The RNN is then trained to identify sequential dependencies in network traffic, enabling it to distinguish between malicious and legitimate activities. These experimental findings show that the suggested model has the following values: F1-score of 99.97%, accuracy of 99.98%, precision of 99.98%, and recall of 99.99% when compared with more traditional ML models like Naive Bayes (NB) and Support Vector Machine (SVM). These findings prove that RNN is efficient and applicable to skewed classes and multidimensional time series patterns in network intrusion detection. The study also identifies the possibility that deep learning (DL) solutions can be used to scale up IDS and enhance its accuracy in real-world network systems.

Keywords - Cybersecurity, Network Intrusion Detection, RNN, Deep Learning, Machine Learning.

1. Introduction

The fast development of information technology has rendered computer networks to become the key to industry, business and many other spheres of ordinary human life. With the continuous increase in the use of digital communication, the maintenance of reliable and secure networks has become an important task for IT administrators[1][2]. Nonetheless, modern networks have become more complex and open, creating a significant challenge in securing their availability, integrity, and confidentiality. Malicious conduct is among the most damaging and disruptive types of cyber challenges (i.e. denial of service (DoS), ransomware, identity theft and data theft). Such attacks take advantage of system vulnerabilities and, in most cases, they are able to circumvent the normal defense mechanisms like antivirus software and firewalls. To combat this increasing threat, network traffic classification has emerged as a critical option to identify and classify network operations into normal and malicious categories [3][4]. Effective classification enables administrators to monitor traffic patterns by detecting intrusions and mitigating potential damage in real-time. Intelligent classification systems not only identify ongoing attacks, but they also issue warnings of abnormal or suspicious activities that are not the normal activities in the network [5]. This is necessary in other scenarios where it is used to handle sophisticated threats like zero-day exploits that are specifically designed to evade the conventional signature systems.

The concept of intelligent network traffic classification combines the domain information, statistical information and computational intelligence; in order to separate legitimate and malicious activities in a high volume and heterogeneous traffic[6][7]. This is done by examining the nature of the traffic that signifies the fingerprints of an attack, mainly based on the attributes and correlation, and anomaly analysis[8][9]. Nonetheless, such an issue as a very uneven distribution of attack types is among the most acute ones since common attacks, such as DoS, are widespread, and less common attacks, such as remote-to-local (R2L) or user-to-root (U2R), are hard to find. This imbalance creates the need to have more learning interventions that can accommodate both the majority and the minority classes. The highest level of machine learning (ML) is centered on the advanced stage, where conventional shallow learning systems are based on manual feature engineering and can handle only substantial and intricate data. The new systems are focused on automation, scalability and increased precision[10][11]. Adaptive learning mechanism Artificial Intelligence (AI) algorithms offer adaptive learning mechanisms, ML models offer excellent classification using predictive analytics, and Complex temporal and nonlinear relationships are learnt using traffic data in RNNs, hybrid models, and deep learning (DL) long short-term memory (LSTM)[12][13]. All these together constitute the foundation of creating intelligent, robust, and scalable intrusion detection systems (IDSs), which are capable of accommodating the threat arising with the emergence of new cyber threats.

1.1. Motivation and Contribution

The emergence of the digital communication business and heightened sophistication of cyberattacks, electronic IDS are not differentiating normal and malicious traffic accordingly. Classical ML approaches frequently neglect both sequential and time-based relationships in network data, leading to reduced accuracy and higher false alarms. Moreover, widely used benchmark datasets such as KDDCup99 are both redundant and imbalanced, making the validity of experimental results lower. The NSL-KDD data set can be used to overcome these limitations, and it is an ideal source of data to develop better approaches to DL. The logic of this research is grounded in the fact that RNNs can model temporal dependencies, which is why there is a need to develop an effective IDS that can increase its detection probability, reduce false positives, and accommodate diverse types of attacks in evolving network environments. This input to the employee turnover forecasting in HR planning research is valuable in several ways:

- Development of an intrusion detection framework that is specifically designed to suit the NSL-KDD data, which makes use of sequential modeling to enhance classification.
- Implementation of a comprehensive preprocessing pipeline including missing value handling, dimensionality reduction, one-hot encoding, normalization, and feature selection to optimize model performance.
- Comparison that demonstrates the strength of the RNN against the challenge of unbalanced attack types and sophisticated time series in network traffic.
- Measured model performance with such high-level metrics as recall, accuracy, precision, F1-score, and AUC.

1.2. Novelty and Justification

The reason behind this work is that the current IDS still endures various problems, such as difficulty in handling high false alarm rates, poor performance in generalizing to attacks that have not been identified, and the inability to support sequential dependencies among the network traffic. Conventional ML methods, which include SVM and NB, assume that network records are independent samples, thus not able to capture temporal relationships, which are of paramount importance in identifying advanced and dynamic patterns of attacks. To address these shortcomings, this study introduces an RNN-based model that leverages its inherent capability to model sequential dependencies, enabling more accurate differentiation between normal and malicious traffic. This work stands out because it uses the RNN architecture in conjunction with extensive preprocessing to improve detection performance on the NSL-KDD dataset. As compared to the conventional methods, it produces noticeably higher F1-scores, recalls, accuracy, and precision. This demonstrates that RNNs have the ability to become an advanced and scalable means of future network intrusion detection systems.

1.3. Structure of the Paper

The paper is structured as follows: Section II presents the Literature review. Section III explains the Methodology, including data preprocessing. Section IV covers the results. Finally, Section V concludes the research and emphasizes areas for future research.

2. Literature Review

An extensive review and analysis of existing research on NIDS with ML and DL techniques has been carried out to provide a strong foundation for this study. Liu et al. (2020) IDS becomes the leading security solution rationally. The main tool for defending networks against different types of hostile activity is anomaly-based network intrusion detection. Applying a variety of ML methods, this study successfully identifies irregularities on the IoT Network Intrusion Dataset. The outcomes demonstrate promise as achieved great efficiency and 99%–100% accuracy[14]. Szostak, Walkowiak and Wlodarczyk (2020) propose Linear Discriminant Analysis (LDA) is an ML technique used to anticipate short-term traffic levels. Since much of the earlier research in this area used time series to simulate the traffic, the traffic prediction issue is thus presented as a classification problem. Additionally, provide numerical figures to demonstrate the efficacy of the suggested methodology. Look at both the real traffic that Seattle Exchange Point collects and the traffic that is produced using real data, which includes genuine traffic dependencies. For actual data flow, were able to make up to 93% of accurate forecasts[15].

Singh and Mathai (2019) To ascertain how well both algorithms worked, 60% of 40% of the four lakh entries in the NSL-KDD dataset for training purposes were utilized and the remaining four lakh for testing. In this experiment, the researcher compared the processing speeds of the new SPELM method with those of the current DBN methodology, accuracy, precision, and recall. When comparing SPELM's accuracy of 93.20% against the computing time of 90.8 seconds, compared to 102 seconds for DBN, the accuracy of 69.492 compared to 66.836 for DBN, and the DBN algorithm's accuracy of 52.8%, the results demonstrate SPELM's superior performance[16]. Taher, Mohammed Yasin Jisan and Rahman (2019) A brand-new supervised ML system is designed to detect legitimate and malicious network traffic. A combination of feature selection and supervised learning algorithms has been used to determine which model has the best detection success rate. Using network traffic as an example, this study shows that ANN may be helpful for ML classification and that the wrapper strategy, which uses features instead of support vector machines (SVMs), is more effective. The NSL-KDD dataset is utilized to assess the efficacy of supervised ML algorithms, such as SVM and ANN, for the purpose of classifying network traffic[17].

Troia et al. (2018) proactively optimize the optical backbone network's resource distribution by concentrating on the traffic matrix prediction technique made possible by DLs. Although RNNs were initially developed for use in sequence prediction tasks, they have recently shown exceptional performance in other domains, including voice recognition, handwriting identification, and time series data prediction. They studied the Gated Recurrent Units (GRU), a specific kind of RNN that can achieve high accuracy (< 7.4% mean absolute error). It can calculate a 66.3% reduction in the capacity of the network by contrasting the numerical results of dynamic allocation with those of static allocation based on forecasts, which enables us to manage unexpected traffic spikes[18]. Zaman and Lung (2018) The majority of IDS in this field and those that are sold commercially are signature-based. ML classification techniques are the foundation of the current anomaly detection trend. Assess the effectiveness of seven distinct Methods for ML that compute information entropy using the Kyoto 2006+ dataset. The results demonstrate that the majority of ML approaches offer precision, recall, and accuracy for this particular data set of more than 90%[19].

The Table I provides a summary of recent studies on Network Traffic classification using ML, methodology, datasets utilized, key findings, and the limitations and future work.

Table 1. Recententrecent Studies on Network Traffic using machine learning

Author	Proposed Work	Dataset	Key Findings	Limitations & Future Work
Liu et al. (2020)	Applied ML algorithms for anomaly-based IDS in IoT networks	IoT Network Intrusion Dataset	Achieved 99–100% accuracy with high efficiency	Need for testing on diverse and larger real-world IoT datasets for generalization
Szostak, Walkowiak & Wlodarczyk (2020)	ML process with LDA classification for short-term traffic volume forecasting	Real traffic (Seattle Exchange Point) + generated traffic	Predictions for actual data flow can be up to 93% accurate	Limited to fixed bitrate levels; future work could extend to dynamic/multi-level prediction
Singh & Mathai (2019)	Compared DBN vs. proposed SPELM algorithm for intrusion detection	NSL-KDD dataset (40% training, 60% testing)	SPELM achieved 93.20% accuracy vs. 52.8% for DBN; lower computational time	Study restricted to NSL-KDD; testing on modern datasets required
Taher, Jisan & Rahman (2019)	Supervised ML for malicious vs. benign traffic classification; compared ANN and SVM with feature selection	NSL-KDD dataset	ANN with wrapper feature selection outperformed SVM	Focused only on two algorithms; other deep learning models could be explored
Troia et al. (2018)	Used GRU-based RNNs for traffic matrix prediction in optical networks	Real optical backbone network traffic data	Achieved <7.4 MAE; enabled dynamic allocation saving 66.3% network capacity	Limited to GRU; future work could explore hybrid deep learning or real-time deployment
Zaman & Lung (2018)	Applied 7 ML techniques with entropy-based features for anomaly detection	Kyoto 2006+ dataset	Most ML techniques achieved >90% accuracy, precision, and recall	Still dependent on dataset quality; need for robust models against evolving attacks

3. Research Methodology

The methodology for the proposed Network traffic Detection begins with the use of this dataset as the experimental benchmark: NSL-KDD, as shown in Figure. 1. The dataset undergoes pre-processing that includes handling missing values, applying dimensionality reduction, transforming categorical attributes through one-hot encoding, and performing normalization using Min-Max scaling. Following this, feature selection is conducted to retain the most illuminating characteristics for successful model training. Partitioning the data into subgroups when processing is done, with 70% going into training and 30% into testing. The RNN model is trained on the training set and subsequently searches for temporal correlations in the network traffic data. The model is subsequently evaluated using common evaluation measures that were evaluated using the test data, including F1-score, AUC, accuracy, precision, and recall. This demonstrates the robustness of the suggested technique.

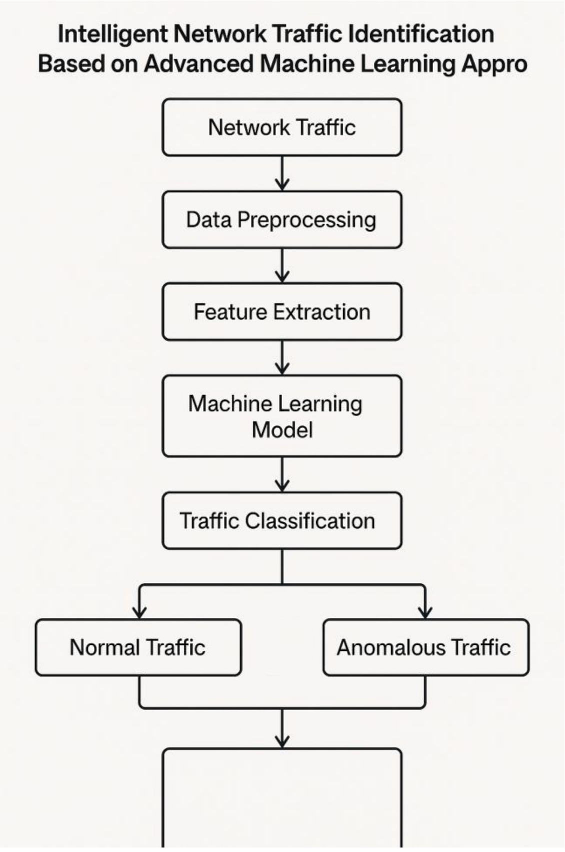


Figure 1. Flowchart for Network Traffic Detection

The following section presents a detailed explanation of each step below.

3.1. Data Gathering and Analysis

The main issue with the KDDcup99 dataset was addressed with the creation of this dataset. Four types of attacks were showcased at the KDDcup99: DoS, Probe, R2L, and U2R. Both files are utilized in the NSL-KDD training and testing processes. In the training session, there are 126,620 incidents and 21 attacks. There are 22,850 occurrences of the 37 different assaults in this tested collection. In order to better understand the dataset, used exploratory data analysis (EDA) to draw several graphs, as it is presented below:

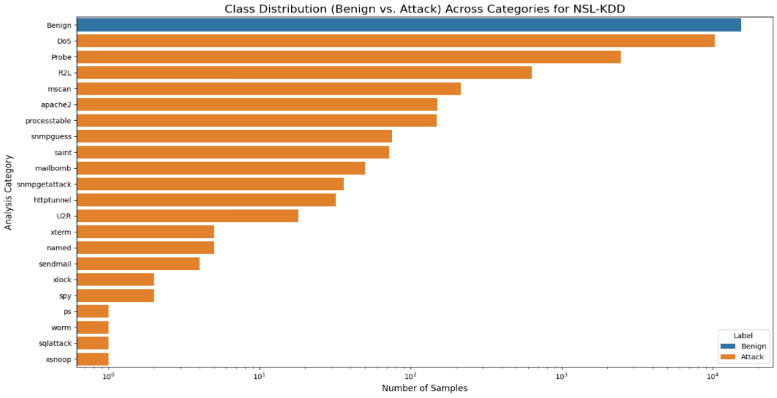


Figure 2. Class Distribution of Benign and Attack Categories in the NSL-KDD Datase

Figure 2 displays the NSL-KDD dataset's attack and benign class distribution. The data is heavily skewed, and the benign group has the most measurements. The attack types include DoS, Probe, and R2L, which occupy the majority of the malicious samples. The other types of attacks that are less common, such as snmpgetattack, sqlattack, and xlock, have relatively lower samples. This distribution demonstrates the existing class imbalance of the data, which is a major consideration to be made when training a model to detect intrusion effectively.

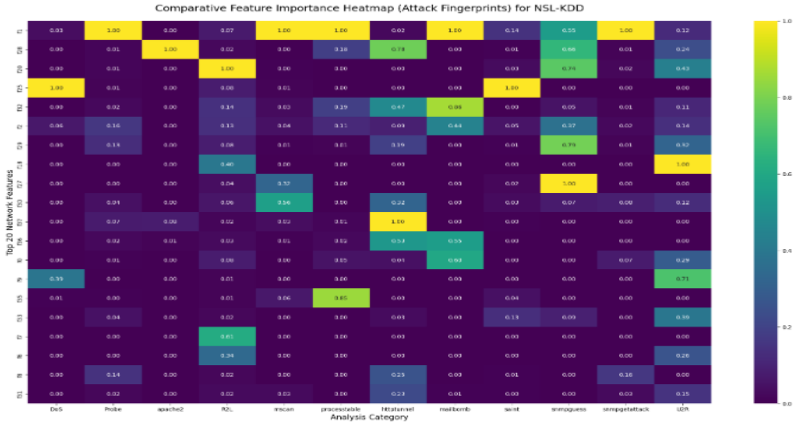


Figure 3. Correlations of Attributes

The NSL-KDD dataset's heatmap showing the relative feature relevance of attack fingerprints is displayed in Figure 3. The heat map identifies the comparative level of contribution of features to various assault types, including R2L, U2R, DoS, and Probe. The values of the intensity provide higher values with greater relevance of particular features in recognizing the relevant types of attacks, and lower values display diminished relevance. The given visualization helps to obtain an understanding of the discriminatory power of features, which can be useful in intrusion detection and attack classification.

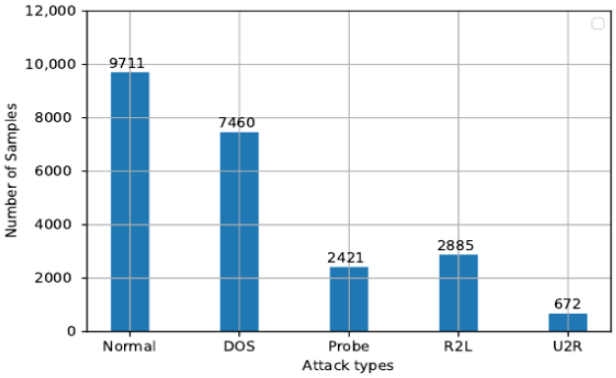


Figure 4. Distribution of Samples by Attack Types

In Figure 4, the bar chart illustrates the distribution of Samples by Attack types, showing the frequencies of various categories of network traffic within a dataset. The sum of all samples is displayed on the y-axis, and the different forms of attacks, including regular traffic and four other types of assaults, are displayed on the x-axis. The four terms are: probe, user-to-root, denial-of-service, and remote-to-local. The chart reveals that Normal traffic has the highest count with 9,711 samples, followed by DOS attacks with 7,460 samples. The other attack types are significantly less frequent, with R2L at 2,885, Probe at 2,421, and U2R being the rarest, with only 672 samples. The dataset's imbalance demonstrates that DOS assaults are the most prevalent form of harmful behaviour in this sample.

3.2. Data Pre-processing

The importance of data preparation, which is crucial to machine learning, cannot be overstated. Missing values, data reduction, one-hot encoding, data scaling and feature selection processing are the main elements of pre-processing raw data. The key pre-processing steps are provided below:

- **Missing values:** The initial stage of data cleansing involves dealing with missing values. The term "missing values" refers to the intentional or unintentional omission of data from a record. Although identifying and encoding missing data is the first stage, resolving the missing values is the second.
- **Data reduction:** It is crucial to optimize the feature and, in this example, lower the number of unique values for categorical variables once the data has been cleared of missing values and other potential biases. To group comparable observations together, clustering was performed.
- **One-hot encoding:** In order for ML models to employ categorical variables, one-hot encoding transforms them into binary vectors. A vector for each category has one element labelled 1 and all others labelled 0.
- **Data Scaling using Min-Max Scaler:** To normalize, the min–max method was employed on the records after a simplified dataset with many characteristics was obtained. This allowed the values to be contained within a range of 0

to 1. To minimize the impact of outliers and maximize the effectiveness of the classifiers that were employed, this was done. Equation (1), the following mathematical formula, was used to execute normalization:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Where x stands for the feature's initial value, x' for its normalised value, x_{min} for its minimum value, and x_{max} for its maximum value.

- **Feature selection:** The reliance between two random variables can be measured using linear measures of dependency, such as the linear correlation coefficient, when considering correlations between records of network traffic as linear links. However, taking into account conversations in the actual world, Additionally, there may be a nonlinear relationship between the variables[20]. It appears that the relationship between two nonlinearly dependent variables cannot be shown by a linear measure. Assessing the capacity to examine the relationship between two variables, whether or not they depend on one another linearly or nonlinearly, is therefore important. Because of these factors, the objective of this study is to determine which features from a feature space are most valuable, independent of their connection type.

3.3. Data Splitting

The NSL-KDD dataset contains training sets and testing sets. Using the Sklearn application, the data is divided. The data used for training is 70% while the data used for testing is 30%.

3.4. Proposed RNN Model

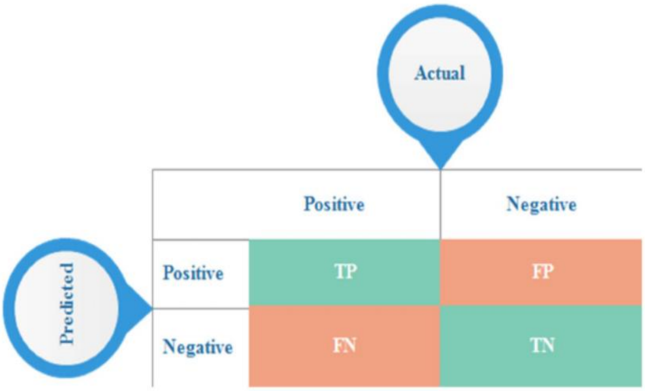
A feed-forward neural network extension called an RNN uses the sequential information[21]. The reason RNNs are referred to as recurrent is that they carry out each sequence piece using the same procedure, based on the outcomes of previous computations. In order to determine the RNN's hidden states, one must use Equation (2):

$$h_t = \sigma(Wx_t + Uh_{t-1} + b_h), \qquad for\ t = T, \ldots .1,$$

In this context, W represents the input to a hidden weight matrix, U stands for the hidden-to-hidden weight matrix, b_hDenotes the bias term, σ denotes the nonlinearity function,x_t denotes the input vector at time T, and h_t embodies the hidden state vector at time t.

3.5. Evaluation Metrics

The foundation of all classifiers is the computation of parameters, including F-score, recall, accuracy, and precision. These metrics were utilised to evaluate the performance of the IDS. The number of samples appropriately identified as positive, using true positive (TP) and true negative (TN), and the number incorrectly identified as positive, using false positive (FP) and false negative (FN), are provided by the formula. The accuracy, which is the number of samples correctly classified as a percentage of all samples, is displayed. Figure 5: Confusion matrices are used to calculate all the aforementioned variables.



	Positive	Negative
Positive	TP	FP
Negative	FN	TN

Figure 5. Confusion Matrix

3.5.1. Accuracy

Accuracy is determined by correctly predicting both positive and negative occurrences it is given as Equation (3):

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+Fp+TN+FN}} \qquad \text{□□□}$$

3.5.2. Precision(P)

The precision calculation yields the probability of an accurate affirmative forecast, which is Equation (4):

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \qquad \text{□□□}$$

3.5.3. Recall(R)

The recall calculates the percentage of accurate classifications to missing entries in mathematical form it is given as Equation (5):

$$Recall = \frac{\mathrm{TP}}{TP + FN}$$

3.5.4. F1 score

The resulting effectiveness metric, recognized as the F-measures is calculated analytically by calculating the accuracy and recall harmonic mean, it is given as Equation (6):

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

3.5.5. AUC

A model's overall performance is shown by the area under the ROC (AUROC), and the ROC curve displays the rates of TP and FP. The more successfully the model outperforms other methods, the higher the AUROC. The ROC curve and AUROC are helpful metrics for evaluating how well the model detects and mitigates attack traffic while reducing false alarms and accurately identifying normal traffic when it comes to detecting DDoS assaults in SDN systems.

4. Results and Discussion

The following platforms for software and hardware were used to conduct the experiments Hardware: 2.2 GHz, 12-core Intel Xeon E5-2650 v4, NVIDIA Quadro P400 2GB, 30MB L3 cache, and 16 GB of RAM. Software: 64-bit WEKA 3.8.2 and Windows 10 Professional. Table II displays the findings from the RNN model's testing on the NSL-KDD dataset for identifying network intrusions. A 99.98% accuracy rate, precision, and F1-score are all achieved by the model. With a fair trade-off between recall and accuracy, these scores demonstrate the precision with which the model identifies intrusions.

Table 2. Experimental Results of Proposed on nsl-kdd dataset for network intrusion detection

Performance matrix	RNN model
Accuracy	99.99
Precision	99.98
Recall	99.99
F1-score	99.97

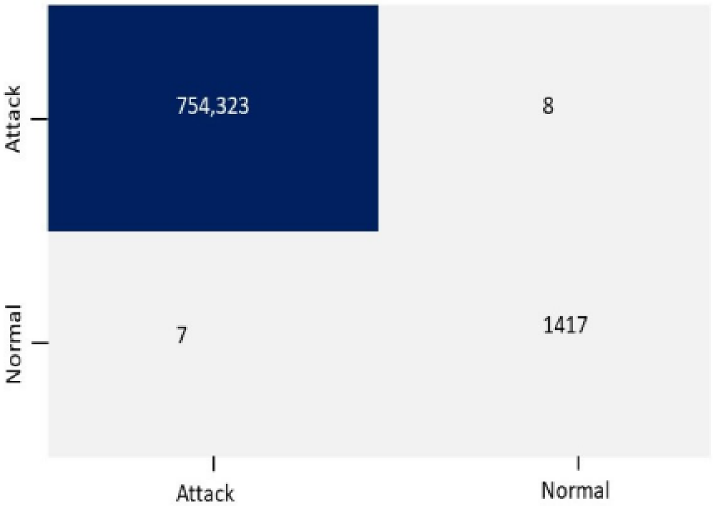


Figure 6. Confusion Matrix for the RNN Model

The RNN model used to categorise the network data produced a confusion matrix, as shown in Figure 6. The model has rightly categorized 754,323, 1, 417, normal and attack cases respectively, had 8 false positives and 7 false negatives, respectively. These results demonstrate effective assault detection with high accuracy and low misclassification.

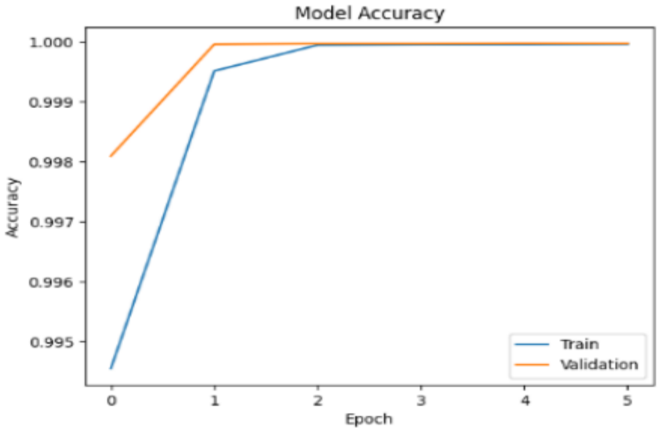


Figure 7. Accuracy Graph for the RNN Model

Figure 7 shows the precision of the RNN model in five epochs. The training and validation curves show how quickly both curves rise during the first and second epochs, and then both curves converge at about 1.000. It proves that the model provides stable and optimal results with no signs of overfitting.

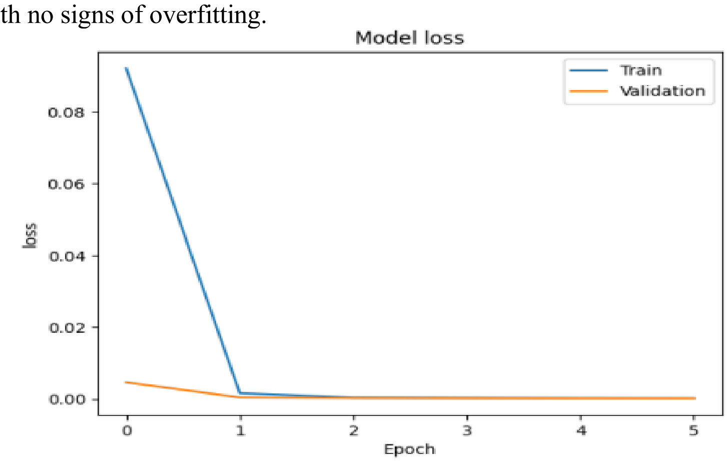


Figure 8. Loss graph for the RNN Model

Figure 8 illustrates the RNN model's performance over five epochs. The training and validation losses drop significantly during the 0th and 1st epochs and then approach a value of 0. The low values of final loss are evidence of successful learning, convergence and no overfitting.

4.1. Comparative Analysis

Table III shows the results of an analysis of the NSL-KDD dataset using ML and DL models to identify potential network threats. With an F1-score of 99.97%, a recall of 99.99%, the proposed RNN model achieves the highest accuracy rate of all the models tested, at 99.98%. On the other hand, SVM exhibits significantly poorer performance with an accuracy of 68%, whereas NB achieves a relatively higher accuracy of 98.02%, although it is still lower than that of the RNN model. These results demonstrate that the RNN is more effective in recognizing intrusions through its improved classification.

Table 3. Performance Comparison of machine and deep learning models for network intrusion detection

Performance matrix	RNN	SVM[22]	NB[23]
Accuracy	99.99	68	98.02
Precision	99.98	68.35	97.19
Recall	99.99	65.15	99.94
F1-score	99.97	67.86	98.55

To leverage the NSL-KDD data, an IDS was built using RNNs. If an RNN wants to distinguish good network traffic patterns from bad ones, it leverages its ability to create sequential dependencies. Compared to the previous ML models, the new ones had better recall, accuracy, and F1 Scores, according to the experiments. As a result, it is evident that RNNs are well-suited for handling complex time-related relationships in intrusion detection tasks.

5. Conclusion and future study

A powerful and essential tool for gathering important information, increasing safety, optimizing performance, and automating network management is the application of ML in network traffic. Every day, businesses create data from their networks, and ML algorithms help them make sense of it. Researchers in this article suggest a model based on an RNN for use in detect network incursions using the NSL-KDD database. Following thorough preparation that included handling missing values, dimensionality reduction, encoding, normalization, and feature selection, the optimized dataset was used to train the models. As it has been demonstrated in the course of the experiments, the proposed RNN turned out to be very successful in sequential dependencies of network traffic, scoring 99.97%, recalling 99.99%, performing precision of 99.98% and achieving an accuracy of 99.99%. It is also apparent in the results of comparative studies that the RNN is better than classical models such as SVM and NB since it is very resistant to identifying both common and uncommon types of attacks, and it has low error rates. Despite the promising outcomes, future work could extend the framework with advanced DL models such as LSTM, GRU, or CNN-RNN hybrids to detect complex attack patterns. Validation on recent datasets that reflect current traffic and emerging threats is also necessary. Additionally, real-time intrusion mitigation using SDN and cloud systems may improve flexibility and scalability.

References

[1] M. S. S. Priya, B. K. Sahu, B. Kumar, and M. Yadav, "Network Intrusion Detection System using XG Boost," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 4070–4073, Oct. 2019, doi: 10.35940/ijeat.A1307.109119.

[2] D. D. Rao, "Multimedia-Based Intelligent Content Networking for Future Internet," in *2009 Third UKSim European Symposium on Computer Modeling and Simulation*, 2009, pp. 55–59. doi: 10.1109/EMS.2009.108.

[3] A. Thakkar and R. Lohiya, "A Review of the Advancement in Intrusion Detection Datasets," *Procedia Comput. Sci.*, vol. 167, pp. 636–645, 2020, doi: https://doi.org/10.1016/j.procs.2020.03.330.

[4] H. P. Kapadia, "Cross-Platform UI/UX Adaptions Engine for Hybrid Mobile Apps," *Int. J. Nov. Res. Dev.*, vol. 5, no. 9, pp. 30–37, 2020.

[5] S. S. S. Neeli, "Serverless Databases: A Cost-Effective and Scalable Solution," *Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci.*, vol. 7, no. 6, p. 7, 2019.

[6] C. Yin, Y. Zhu, J. Fei, and X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," *IEEE Access*, vol. 5, pp. 21954–21961, 2017, doi: 10.1109/ACCESS.2017.2762418.

[7] A. Thapliyal, P. S. Bhagavathi, T. Arunan, and D. D. Rao, "Realizing Zones Using UPnP," in *2009 6th IEEE Consumer Communications and Networking Conference*, 2009, pp. 1–5. doi: 10.1109/CCNC.2009.4784867.

[8] H. Shapoorifard and P. Shamsinejad, "Intrusion Detection using a Novel Hybrid Method Incorporating an Improved KNN," *Int. J. Comput. Appl.*, vol. 173, no. 1, pp. 5–9, Sep. 2017, doi: 10.5120/ijca2017914340.

[9] S. S. S. Neeli, "Real-Time Data Management with In-Memory Databases : A Performance-Centric Approach," *J. Adv. Dev. Res.*, vol. 11, no. 2, p. 49, 2020.

[10] L. Wang, "Big Data in Intrusion Detection Systems and Intrusion Prevention Systems," *J. Comput. Networks*, vol. 4, no. 1, pp. 48–55, Aug. 2017, doi: 10.12691/jcn-4-1-5.

[11] S. S. S. Neeli, "Decentralized Databases Leveraging Blockchain Technology," vol. 8, no. 1, pp. 1–8, 2020.

[12] M.-J. Kang and J.-W. Kang, "Intrusion Detection System Using Deep Neural Network for In-Vehicle Network Security," *PLoS One*, vol. 11, no. 6, p. e0155781, Jun. 2016, doi: 10.1371/journal.pone.0155781.

[13] A. Kushwaha, P. Pathak, and S. Gupta, "Review of optimize load balancing algorithms in cloud," *Int. J. Distrib. Cloud Comput.*, vol. 4, no. 2, pp. 1–9, 2016.

[14] Z. Liu, N. Thapa, A. Shaver, K. Roy, X. Yuan, and S. Khorsandroo, "Anomaly Detection on IoT Network Intrusion Using Machine Learning," in *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, IEEE, Aug. 2020, pp. 1–5. doi: 10.1109/icABCD49160.2020.9183842.

[15] D. Szostak, K. Walkowiak, and A. Wlodarczyk, "Short-Term Traffic Forecasting in Optical Network using Linear Discriminant Analysis Machine Learning Classifier," in *2020 22nd International Conference on Transparent Optical Networks (ICTON)*, IEEE, Jul. 2020, pp. 1–4. doi: 10.1109/ICTON51198.2020.9203040.

[16] K. Singh and K. J. Mathai, "Performance Comparison of Intrusion Detection System Between Deep Belief Network (DBN)Algorithm and State Preserving Extreme Learning Machine (SPELM) Algorithm," in *Proceedings of 2019 3rd IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT 2019*, 2019. doi: 10.1109/ICECCT.2019.8869492.

[17] K. A. Taher, B. M. Y. Jisan, and M. Mahbubur, "Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection," in *2019 International Conference on Robotics,Electrical and Signal Processing Techniques (ICREST)*, IEEE, Jan. 2019, pp. 643–646. doi: 10.1109/ICREST.2019.8644161.

[18] S. Troia, R. Alvizu, Y. Zhou, G. Maier, and A. Pattavina, "Deep Learning-Based Traffic Prediction for Network Optimization," in *International Conference on Transparent Optical Networks*, 2018. doi: 10.1109/ICTON.2018.8473978.

[19] M. Zaman and C.-H. Lung, "Evaluation of machine learning techniques for network intrusion detection," in *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, IEEE, Apr. 2018, pp. 1–5. doi: 10.1109/NOMS.2018.8406212.

[20] S. S. Panwar, P. S. Negi, and Y. P. Raiwani, "Implementation of Machine Learning Algorithms on CICIDS-2017 Dataset

for Intrusion Detection using WEKA," *Int. J. Recent Technol. Eng.*, vol. 8, no. 3, pp. 2195–2207, Sep. 2019, doi: 10.35940/ijrte.C4587.098319.

[21] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, "Deep Recurrent Neural Network for Intrusion Detection in SDN-based Networks," in *2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft)*, IEEE, Jun. 2018, pp. 202–206. doi: 10.1109/NETSOFT.2018.8460090.

[22] M. A. Khan, M. R. Karim, and Y. Kim, "A Scalable and Hybrid Intrusion Detection System Based on the Convolutional-LSTM Network," *Symmetry (Basel).*, vol. 11, no. 4, p. 583, Apr. 2019, doi: 10.3390/sym11040583.

[23] S. Das, M. Ashrafuzzaman, F. T. Sheldon, and S. Shiva, "Network Intrusion Detection using Natural Language Processing and Ensemble Machine Learning," in *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020*, 2020. doi: 10.1109/SSCI47803.2020.9308268.

[24] Polam, R. M., Kamarthapu, B., Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., & Vangala, S. R. (2021). Big Text Data Analysis for Sentiment Classification in Product Reviews Using Advanced Large Language Models. *International Journal of AI, BigData, Computational and Management Studies*, *2*(2), 55-65.

[25] Gangineni, V. N., Tyagadurgam, M. S. V., Chalasani, R., Bhumireddy, J. R., & Penmetsa, M. (2021). Strengthening Cybersecurity Governance: The Impact of Firewalls on Risk Management. *International Journal of AI, BigData, Computational and Management Studies*, *2*, 10-63282.

[26] Pabbineedi, S., Penmetsa, M., Bhumireddy, J. R., Chalasani, R., Tyagadurgam, M. S. V., & Gangineni, V. N. (2021). An Advanced Machine Learning Models Design for Fraud Identification in Healthcare Insurance. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *2*(1), 26-34.

[27] Kamarthapu, B., Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., Vangala, S. R., & Polam, R. M. (2021). Advanced Machine Learning Models for Detecting and Classifying Financial Fraud in Big Data-Driven. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *2*(3), 39-46.

[28] Tyagadurgam, M. S. V., Gangineni, V. N., Pabbineedi, S., Penmetsa, M., Bhumireddy, J. R., & Chalasani, R. (2021). Enhancing IoT (Internet of Things) Security Through Intelligent Intrusion Detection Using ML Models. *International Journal of Emerging Research in Engineering and Technology*, *2*(1), 27-36.

[29] Vangala, S. R., Polam, R. M., Kamarthapu, B., Kakani, A. B., Nandiraju, S. K. K., & Chundru, S. K. (2021). Smart Healthcare: Machine Learning-Based Classification of Epileptic Seizure Disease Using EEG Signal Analysis. *International Journal of Emerging Research in Engineering and Technology*, *2*(3), 61-70.

[30] Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., Vangala, S. R., Polam, R. M., & Kamarthapu, B. (2021). Big Data and Predictive Analytics for Customer Retention: Exploring the Role of Machine Learning in E-Commerce. *International Journal of Emerging Trends in Computer Science and Information Technology*, *2*(2), 26-34.

[31] Penmetsa, M., Bhumireddy, J. R., Chalasani, R., Tyagadurgam, M. S. V., Gangineni, V. N., & Pabbineedi, S. (2021). Next-Generation Cybersecurity: The Role of AI and Quantum Computing in Threat Detection. *International Journal of Emerging Trends in Computer Science and Information Technology*, *2*(4), 54-61.

[32] Polu, A. R., Vattikonda, N., Gupta, A., Patchipulusu, H., Buddula, D. V. K. R., & Narra, B. (2021). Enhancing Marketing Analytics in Online Retailing through Machine Learning Classification Techniques. *Available at SSRN 5297803*.

[33] Polu, A. R., Buddula, D. V. K. R., Narra, B., Gupta, A., Vattikonda, N., & Patchipulusu, H. (2021). Evolution of AI in Software Development and Cybersecurity: Unifying Automation, Innovation, and Protection in the Digital Age. *Available at SSRN 5266517*.

[34] Polu, A. R., Vattikonda, N., Buddula, D. V. K. R., Narra, B., Patchipulusu, H., & Gupta, A. (2021). Integrating AI-Based Sentiment Analysis With Social Media Data For Enhanced Marketing Insights. *Available at SSRN 5266555*.

[35] Buddula, D. V. K. R., Patchipulusu, H. H. S., Polu, A. R., Vattikonda, N., & Gupta, A. K. (2021). INTEGRATING AI-BASED SENTIMENT ANALYSIS WITH SOCIAL MEDIA DATA FOR ENHANCED MARKETING INSIGHTS. *Journal Homepage: http://www. ijesm. co. in*, *10*(2).

[36] Gupta, A. K., Buddula, D. V. K. R., Patchipulusu, H. H. S., Polu, A. R., Narra, B., & Vattikonda, N. (2021). An Analysis of Crime Prediction and Classification Using Data Mining Techniques.

[37] Rajiv, C., Mukund Sai, V. T., Venkataswamy Naidu, G., Sriram, P., & Mitra, P. (2022). Leveraging Big Datasets for Machine Learning-Based Anomaly Detection in Cybersecurity Network Traffic. *J Contemp Edu Theo Artific Intel: JCETAI/102*.

[38] Sandeep Kumar, C., Srikanth Reddy, V., Ram Mohan, P., Bhavana, K., & Ajay Babu, K. (2022). Efficient Machine Learning Approaches for Intrusion Identification of DDoS Attacks in Cloud Networks. *J Contemp Edu Theo Artific Intel: JCETAI/101*.

[39] Bhumireddy, J. R., Chalasani, R., Tyagadurgam, M. S. V., Gangineni, V. N., Pabbineedi, S., & Penmetsa, M. (2020). Big Data-Driven Time Series Forecasting for Financial Market Prediction: Deep Learning Models. *Journal of Artificial Intelligence and Big Data*, *2*(1), 153–164.DOI: 10.31586/jaibd.2022.1341

[40] Nandiraju, S. K. K., Chundru, S. K., Vangala, S. R., Polam, R. M., Kamarthapu, B., & Kakani, A. B. (2022). Advance of AI-Based Predictive Models for Diagnosis of Alzheimer's Disease (AD) in Healthcare. *Journal of Artificial Intelligence and Big Data*, *2*(1), 141–152.DOI: 10.31586/jaibd.2022.1340

[41] Tyagadurgam, M. S. V., Gangineni, V. N., Pabbineedi, S., Penmetsa, M., Bhumireddy, J. R., & Chalasani, R. (2022). Designing an Intelligent Cybersecurity Intrusion Identify Framework Using Advanced Machine Learning Models in Cloud Computing. *Universal Library of Engineering Technology*, (Issue).

[42] Vangala, S. R., Polam, R. M., Kamarthapu, B., Kakani, A. B., Nandiraju, S. K. K., & Chundru, S. K. (2022). Leveraging Artificial Intelligence Algorithms for Risk Prediction in Life Insurance Service Industry. *Available at SSRN 5459694*.

[43] Polam, R. M., Kamarthapu, B., Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., & Vangala, S. R. (2021). Data Security in Cloud Computing: Encryption, Zero Trust, and Homomorphic Encryption. *International Journal of Emerging Trends in Computer Science and Information Technology*, *2*(3), 70-80.

[44] Gangineni, V. N., Pabbineedi, S., Penmetsa, M., Bhumireddy, J. R., Chalasani, R., & Tyagadurgam, M. S. V. Efficient Framework for Forecasting Auto Insurance Claims Utilizing Machine Learning Based Data-Driven Methodologies. *International Research Journal of Economics and Management Studies IRJEMS*, *1*(2).

[45] Vattikonda, N., Gupta, A. K., Polu, A. R., Narra, B., Buddula, D. V. K. R., & Patchipulusu, H. H. S. (2022). Blockchain Technology in Supply Chain and Logistics: A Comprehensive Review of Applications, Challenges, and Innovations. *International Journal of Emerging Research in Engineering and Technology*, *3*(3), 99-107.

Narra, B., Vattikonda, N., Gupta, A. K., Buddula, D. V. K. R., Patchipulusu, H. H. S., & Polu, A. R. (2022). Revolutionizing Marketing Analytics: A Data-Driven Machine Learning Framework for Churn Prediction. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *3*(2), 112-121.

[46] Polu, A. R., Narra, B., Buddula, D. V. K. R., Patchipulusu, H. H. S., Vattikonda, N., & Gupta, A. K. BLOCKCHAIN TECHNOLOGY AS A TOOL FOR CYBERSECURITY: STRENGTHS, WEAKNESSES, AND POTENTIAL APPLICATIONS.

[47] Bhumireddy, J. R., Chalasani, R., Tyagadurgam, M. S. V., Gangineni, V. N., Pabbineedi, S., & Penmetsa, M. (2022). Big Data-Driven Time Series Forecasting for Financial Market Prediction: Deep Learning Models. *Journal of Artificial Intelligence and Big Data*, *2*(1), 153–164.DOI: 10.31586/jaibd.2022.1341

[48] Nandiraju, S. K. K., Chundru, S. K., Vangala, S. R., Polam, R. M., Kamarthapu, B., & Kakani, A. B. (2022). Advance of AI-Based Predictive Models for Diagnosis of Alzheimer's Disease (AD) in Healthcare. *Journal of Artificial Intelligence and Big Data*, *2*(1), 141–152.DOI: 10.31586/jaibd.2022.1340