Development of a Network Intrusion Detection Model using Hybridised Machine Learning Algorithms

Ogundele Oluwafeyisayo Mary [1], Okokpujie Kennedy[2], Maha Ojimaojo King-David[2], Adaora P. Ijeh[3], Imhade P. Okokpujie[4,5]

[1]Department of Network and User Administration, NigeriaUnified Payment Idowu Martins, Victoria Island, Lagos, Nigeria
[2]Department of Electrical and Information Engineering, College of Engineering, Covenant University, Ota, Nigeria
[3]Department of Computer and Information Science, College of Science and Technology, Covenant University, Ota, Nigeria
[4]Department of Mechanical and Mechatronics Engineering, Afe Babalola University, Ado Ekiti, 360001 Nigeria
[5]Department of Mechanical and Industrial Engineering Technology, University of Johannesburg, Johannesburg, 2028, South Africa

Article Info

Article history:

Received Sep 15, 2024
Revised May 26, 2025
Accepted Jun 24, 2025

Keywords:

Hybrid Models,
Intrusion Detection,
Network Security,
Recurrent Neural Networks,
Random Forests

ABSTRACT

Cyber threats continue to grow in this era since the bad actors are attempting to exploit individuals, organisations, and systems. The latest development in artificial intelligence has unleashed strong agents at the fingertips of humanity. As open as it is, it has made more room for possible bad actors. Systems that can successfully counter these threat actors need to be created to rescue humanity. In this research work, RNN and Random Forest classifiers' hybridised models are combined for the development of a Network Intrusion Detection System (NIDS) based on the benchmark dataset (CICIDS 2017) The requirement for an efficient and accurate method to detect network intrusions, both known and zero-day anomalies, is the primary problem considered. This research aims to enhance the accuracy and reliability of intrusion detection systems through a hybrid modelling approach. For evaluating the performance of the proposed model, various measures like accuracy, precision, recall, F1 measure, true positive rate, and true negative rate were employed. The hybrid model showed very good results with testing accuracy of 96.08%, precision of 96.0%, and recall of 96.0%, along with an F1 measure of 96.0%. The result of the experiment indicates that the model is effective and, when implemented, can detect and classify cyberattacks in modern environments.

Corresponding Author:

Okokpujie Kennedy
Department of Electrical and Information Engineering,
College of Engineering, Covenant University,
Ota, Nigeria
kennedy.okokpujie@covenantuniversity.edu.ng

1. INTRODUCTION

As more people started using the Internet, the late 21st century, called the "Digital Age",—marked a critical turning point in worldwide connectedness. Statista's internet trend statistics reveal a remarkable surge in global internet users, with a 13.9% increase in 2015 and an average annual growth rate of 7.4%. 5.18 billion individuals, or 64.6% of the global population, were online as of April 2023. This digital era has also seen the proliferation of devices managed via wireless networks, and the demand for reduced latency in technologies such as self-driving cars, robots, and healthcare has intensified reliance on evolving technologies to enhance job performance and efficiency across various industries. Human lives are increasingly intertwined with the Internet, from food services to financial transactions, entertainment, communication, and health services, particularly in this AI-driven age [1].

However, alongside this ubiquitous connectivity and integration of digital technologies, cyberspace has evolved into a dynamic and complex landscape. Here, the expanding capabilities of information systems coexist with the persistent threat of malicious activities. The vastness of the web has made it an attractive target for criminals, necessitating robust measures to detect and prevent attacks on systems and individuals. Cyberspace has significantly benefited from using intrusion detection systems (IDS), which many researchers use to combat cyber threats. Despite significant advancements, challenges such as novel threats, zero-day attacks, and false positives persist, highlighting the need for continued innovation in cybersecurity [2].

Intrusion Detection Systems (IDS) are essential for cybersecurity because they can detect and stop harmful activity or unauthorized access to computer networks. Traditional signature-based IDS have proven effective against known attacks but struggle with novel threats, false positives, and lengthy training times. This research addresses these limitations by developing a hybrid system that combines a deep learning detector, specifically a Recurrent Neural Network (RNN), with a supervised learning algorithm, the Random Forest (RF) algorithm. Both algorithms have demonstrated high accuracy and robustness in handling previous datasets. This hybrid approach aims to enhance the detection of known and unknown attacks, reduce false positives, and improve overall system performance.

Traditionally, supervised machine learning algorithms such as Random Forest, Decision Trees, Naïve Bayes, and Support Vector Machines, or unsupervised learning algorithms like K-Nearest Neighbor (KNN), Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN) have been used in the use of IDS to address cybersecurity issues. These algorithms are trained on well-known datasets such as the CICIDS2017 and NSL-KDD datasets to identify attacks. Research has shown that Kamil and Mohammed [3] demonstrated that CNN models achieve high accuracy rates and low false positives. Similarly, a study by Ouiazzane et al. [4] highlighted the effectiveness of decision tree algorithms in recognising regular network traffic with high accuracy and minimal false alarms. These findings support the potential of hybrid models to address the limitations of traditional IDS.

The hybrid network intrusion detection system proposed in this paper combines a Random Forest algorithm with an RNN. The RNN is used for feature extraction, leveraging its ability to detect local patterns in network traffic data. The Random Forest algorithm will then classify these features, utilising its ensemble of Decision Trees to enhance accuracy and robustness. This approach aims to capitalise on the strengths of both algorithms, ensuring high detection rates for both known and unknown attacks while minimising false positives. The system is trained on the CSE-CICIDS2017 dataset, which includes numerous attack scenarios, including web attacks, DDoS, Heartbleed, botnets, brute force, and internal network intrusion. Developing a hybrid network intrusion detection system holds significant promise in enhancing cybersecurity measures. The suggested system responds to known and unexpected threats more efficiently by combining deep learning and supervised learning techniques, significantly lowering false positives. With cyber dangers on the rise and digital ecosystems becoming more complex due to technologies like 5G and IoT, innovation like this is essential. The hybrid model's ability to leverage labelled data to detect novel threats and minimize false alarms is a pivotal aspect of this research, addressing common challenges in obtaining extensive labelled datasets. Furthermore, this study is in line with Sustainable Development Goals (SDGs) 16 (Peace, Justice & Strong Institutions) and 9 (Industry, Innovation & Infrastructure), highlighting its broader significance in defending vital infrastructure and thwarting cybercrime.

1.2 Classification of Intrusion Detection Systems

Intrusion Detection Systems (IDS) are classified into deployment method-based IDS and detection method-based IDS. Deployment method-based IDS is further classified into Host-based IDS (HIDS) and Network-based IDS (NIDS). In contrast, detection Method-based IDS is further classified into Signature-based IDS and Anomaly Detection-based IDS. Figure 1 shows the categorisation of Intrusion Detection Systems. One essential security tool is a HIDS, which monitors and assesses the internal conditions of a single host, such as a server or personal computer (PC). HIDS looks at the host's internal workings and data flows instead of network-based intrusion detection systems, monitoring network traffic. HIDS periodically takes a snapshot of the host's file system, which it then compares over time. An essential security equipment is the NIDS, which scans network traffic for indications of hostile activity or policy infractions. NIDS is concerned with the entire network environment, unlike Host Intrusion Detection Systems (HIDS), which concentrate on specific host systems. NIDS examine packets as they traverse the network and analyze them in real-time to detect suspicious patterns or behaviours suggestive of cyber threats. Typically, the system combines anomaly-based detection, which detects departures from typical network behaviour, with signature-based detection, which searches for known attack patterns [5].

Signature-based IDS (SIDS), or misuse or knowledge-based detection, utilizes predefined attack patterns stored in a database. This method efficiently identifies known threats but struggles with new, unidentified attacks and requires significant resources to maintain and compare extensive databases. Conversely, Anomaly-Based IDS (Anomaly-Based AIDS, or behaviour-based detection establishes a profile for regular network activity and flags any deviations as potential threats. This approach is adept at detecting novel attacks due to its ability to identify abnormal behaviour. However, it can have a high false alarm rate (FAR) because distinguishing between normal and abnormal behaviour can be challenging. To maximize the effectiveness of IDS, combining both SIDS and AIDS methods is recommended. This hybrid approach reduces the risk of false positives and negatives, enhancing overall threat detection and response capabilities. Regularly updating signature databases and refining anomaly detection algorithms improve IDS performance, ensuring robust protection against familiar and emerging threats. Figure 1 shows the ategorization of Intrusion Detection Systems [6].

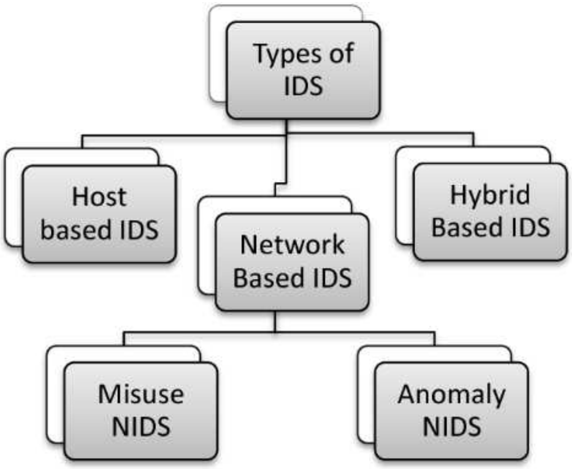


Figure 1. Categorization of Intrusion Detection Systems

2. RELATED WORKS

In recent years, various hybrid network intrusion detection systems (IDSs) have been integrated with multiple machine learning algorithms to get the best possible results with attack detection and categorization.

Du *et al.* [7] suggested a deep learning model for network intrusion detection (DLNID), a traffic anomaly detection model. An attention mechanism and a bidirectional long short-term memory (Bi-LSTM) network are combined in this model. The Bi-LSTM was used to ascertain the network pattern sequence after reassigning weights and extracting the attack features using a pure Convolutional Neural Network (CNN). However, it wasn't used to create an online intrusion detection model using an integrated network capture module. As a result, while the model may be effective at spotting known patterns, it will be less successful at detecting zero-day attacks.

Hussain *et al.* [8] proposed the semi-supervised one-class Support Vector Machine (OC-SVM) and Supervised Random Forest (RF) methods to create a Hybrid Network Intrusion Detection System. It used two stages to operate: the first stage filtered malicious and benign traffic using an OC-SVM. In the following stages, several parallel supervised models and an extra OC-SVM model were employed to distinguish between known and unknown attacks and malicious communications. Although the model was trained on a small dataset, its performance on different types of attacks is unclear; it performed optimally and achieved high accuracy scores of 99.45% and 93.99% on known and zero-day attacks, respectively. Due to the FPR to FNR trade-off, the FNR displayed was high at 7.28%, even if the FPR rate was shallow at 0.44%.

Silivery *et al.* [9], the authors created a dependable intrusion detection system to recognize malicious attempts by implementing a multi-model methodology that included Recurrent Neural Network (RNN), Long-Short Term Memory Recurrent Neural Network (LSTM-RNN), and Deep Neural Network (DNN). As a result, the study work in our proposed model had a solid foundation due to the LSTM-RNN's high accuracy of 98.68% and low FPR of 2.47%.

Hnamte *et al.* [10] intended to identify threats using a hybrid machine-learning approach. The Crow Search Algorithm (CSA) is used to identify critical characteristics, and the Extreme Learning Machine (ELM) is mapped to the decision tree to increase classification accuracy. Additionally, they demonstrated through a comparative examination of test data that the model's accuracy decreases with increased features being used. They obtained an accuracy of 97.89% with their 11 features, 94.23% with their 8 features, and 94.13% with their 4 features, indicating that the number of features in the system affects the detection rate.

Pande *et al.* [11], the NSL-KDD dataset has a precision of 99.96%, a recall of 99.97%, and a precision of 99.79%. However, the UNSW NB15 detection findings using this paper's model have an overall identification accuracy rate of 90.12%, a recall of 95.20%, and a precision of 89.93%. The authors employed a CBL DDQN Model Based on an Improved Double Deep Q Network, a CNN and a BiLSTM hybrid. It performs poorly in classification prediction, such as in random forest, SVM, and MLP. Hybrid Network

Intrusion Detection Systems have been implemented in various scientific research works. Still, the performance of zero-day attacks and the response of the developed models to false alarm rates is very low. Summary of the related works is shown in Table 1. Our proposed model aims to respond more effectively to known and unknown threats, drastically reducing false positives and combating cyberattacks in networks and systems.

Table 1. Summary of Related Works

S/N	References	Dataset	Models	Performance	Limitations
1	[5]	CICIDS2017 dataset	Near-autonomous Hybrid IDS comprising a Deep Neural Network (DNN) and the K-Nearest Neighbors (KNN) algorithm	The Port Scan attacks occur in sliding window 16; the AUC drops from 0.97 to less than 0.50 and then resumes after learning with an AUC of around 0.96.	A sharp drop in the AUC measures each time a new attack is introduced to the system, indicating the failure of the neural network to detect the unseen attacks.
6	[8]	KDDCUP'99 and CIC-MalMem-2022 Datasets	A hybrid strategy that makes use of deep learning (DL) and machine learning (ML) techniques, such as XGBoost for feature selection and SMOTE for data balancing	Accuracy of 99.99% and 100% for KDDCUP'99 and CIC-MalMem-2022	Did not apply the model to novel emerging threats.
7	[19]	NSL-KDD Dataset	An adaptive deep learning algorithm with data pre-processing. A module, a neural network pre-training module, and a classifier module.	After adding the proposed ADL to the Naïve Bayes, the accuracy of R2L is improved from 84.13% to 91.32%; the accuracy rate increased from 91.23% to 96.44%, the accuracy rate of U2L increased from 28.49% to 43.83%, and the accuracy rate increased from 66.88% to 75.02%	The performance of the model was not verified with other datasets, such as the UNSW-NB15 dataset
11	[11]	NSL-KDD and UNSWNB15 data sets	A CBL (a hybrid of CNN and BiLSTM) DDQN Model Based on Improved Double Deep Q Network	UNSW NB15 detection results using this paper's model has an overall identification accuracy rate of 90.12%, recall of 95.20%, and a precision of 89.93%.NSL-KDD dataset with precision at 99.96%, recall at 99.97%, and precision at 99.79%.	It performs poorly as its counterparts – Random Forest, SVM, and MLP in Classification Prediction.
14	[7]	NSL-KDD Dataset	A bidirectional long-short-term memory (Bi-LSTM) network and an attention mechanism combined into a deep learning network intrusion detection (DLNID) model	Accuracy of 90.73% and F1 score of 89.65%	Did not apply to an actual, combined network capture module to implement an online intrusion detection model.
15	[4]	CICIDS2017 dataset	A hybrid NIDS model combining both the use of an ADNIDS for anomaly detection, integrated with an SNIDS to identify known cyber-attacks based on their signatures	The Decision Tree could recognize normal network traffic with up to 99.9% accuracy and a very low false alarm rate.	The work does not address the problem of detecting novel attacks.
16	[16]	KDD CUP99 dataset	Network Intrusion Detection Using Stacked NDAE and SVM Classification in a Non-Symmetric Deep Auto-Encoder Algorithm	The overall Accuracy of 99.65%, a Precision of 99.99%, a Recall of 99.85%, and an F1-score of 99.55%.	There is no mention of how the model performs on novel threats and its response to zero-day attacks.

3. METHODOLOGY

The research framework for developing a hostile traffic detection system is seen in Figure 2. A typical research process is depicted in the conceptual framework diagram for this study. An extensive literature review informs this framework on utilising hybridised models. The stages for the development of a model include data acquisition, data preprocessing, hybridisation, training, validation, and finally testing. Each step will be explained in the preceding sub-section.

3.1 Data Acquisition.

After the problem definition phase, the next step is collating and collecting the required data. Relevant training and testing data are needed to train the machine learning model using hybridised models.

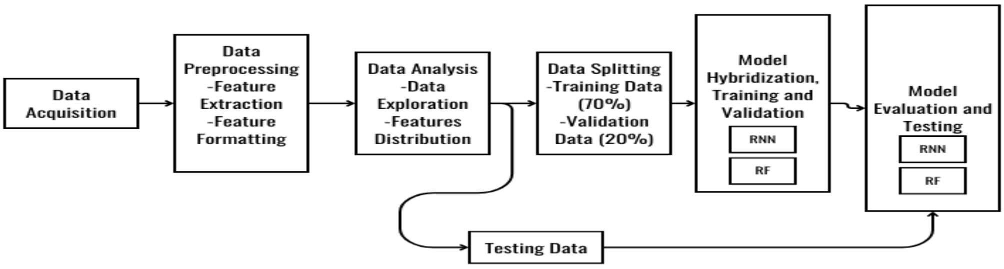

Data Acquisition
Data Preprocessing -Feature Extraction -Feature Formatting
Data Analysis -Data Exploration -Features Distribution
Data Splitting -Training Data (70%) -Validation Data (20%)
Model Hybridization, Training and Validation
RNN
RF
Model Evaluation and Testing
RNN
RF
Testing Data

Figure 2. The Research Conceptual Framework

The dataset was retrieved from the Canadian Institute of Cybersecurity [17], which contains many other datasets, including IoT, DNS, IDS, Malware, Operational Technology, ISCX, and others. The dataset identified in this work is the CICEDS2017 dataset, one of the IDS Datasets. It was created in simulated and flow-based environments and was grouped to contain attacks in the following categories: DoS, DDoS, Web Attacks, Botnet, Brute force, and PortScan attacks [18].

3.2 Data Preprocessing and Feature Extraction.

After obtaining the datasets, pre-processing and feature extraction are the subsequent phases. Data preprocessing makes the data suitable for carrying out model training. The following preprocessing activities will be carried out for the datasets: data cleaning and normalisation. We collated and concatenated the CICEDS2017 data files in the data cleaning stage. The CICEDS2017 dataset has eight CSV files containing various attack scenarios recorded during the simulation. The columns required for the model training were chosen based on performance, and the rows with missing values were removed. The attacks with insufficient samples were dropped, various DoS attack types were grouped into a single "dos" label, and various brute-force attack types were grouped into a single "brute force" label.

The various web attack types were grouped into a single "web_attack" label, and the data was merged along the same axis. The combined data summary is shown in Table 2, and the feature summary is in Table 3. In the subsequent step, the data was normalised by replacing the values 0 and 1 with the Min-Max Scaler normalisation function to avoid lowering the model's performance. After that, the dataset was divided into three sets: 70% of the data for training, 20% for validation, and 10% for testing.

Table 2. An Overview of the 2017 CICIDS Dataset

S/N	Traffic Label	Number of Records
1	Normal	2273097
2	DDoS	128027
3	Web Attack	2180
4	Botnet	1966
5	Brute force	13835
6	Portscan	158930
7	DoS	252661

Table 3. The CICIDS 2017 Dataset's features

S/N	Feature Name
1	Duration
2	Source Port
3	Destination Port
4	Protocol (TCP, UDP, ICMP, IGMP)
5	Packets
6	Bytes
7	Urgent Flag
8	Acknowledge Flag
9	Push Flag
10	Reset Flag
11	Finish Flag
12	Attack Label

3.3 Model Configuration, Training, Validation, and Model Selection.

The two machine learning models used for this research are Recurrent Neural Network (RNN) and Random Forest (RF). It is highly paramount that relevant specific parameters are used, and some of the activities carried out during the model configuration as shown in Table 4. In this research, a hybrid approach combining RNN and RF for Network Intrusion Detection Systems (NIDS) was implemented. This hybrid approach leverages the temporal modelling strength of RNNs and the classification strength of Random Forests. The reason for combining the RNN and RF is that RNN has the ability to learn temporal patterns (e.g., how a session evolves over time), and RF excels at classifying based on structured features (e.g., flow statistics). Their combination yields a stronger and more accurate detection model, especially for advanced and stealthy attacks.

3.3.1 Description of Models for Hybridisation

The following models were selected to be hybridised together for the development of the hybridised model for the detection of network intrusions, as seen from their impact in other literature, are Recurrent Neural Networks (RNN) and Random Forest (RF).

Table 4. Configurational Parameters

S/N	Parameters	Value
1	Batch Size	128
2	Number of Features	7
3	Metrics Used	Accuracy, Precision, True Positive Rate or Recall, False Positive Rate, False Negative Rate
4	Epochs	20
5	Activation Function	ReLU
6	Loss Function	Categorical Cross-Entropy
7	Optimizer	'adam'
8	Number of units in the RNN layer	50
9	Number of units in the dense layer	7
10	Number of Estimators for Random Forest	100

3.4 Recurrent Neural Networks (RNNs)

A neural network in which the input for the current step is provided by the output of the preceding phase. In machine learning, inputs and outputs have historically been independent of one another, making it challenging to predict a future state from a prior one. However, RNN was developed with a "hidden state" feature that aids in the memory of sequence information.

In contemporary machine learning, diverse techniques are employed to manage various data types. One particularly challenging data type to handle and predict is sequential data. Unlike typical datasets where features are assumed to be order-independent, sequential data has inherent order dependencies that must be preserved and understood. Recurrent Neural Networks (RNNs) were developed to manage such data effectively.

An RNN comprises units with fixed activation functions, one for each time step in the sequence. Each unit maintains an internal state, known as the hidden state, which embodies the information about the past sequence the network has processed up to that point. This hidden state is continually updated at each time step, reflecting the evolving knowledge of the network regarding the sequence. This mechanism enables RNNs to leverage historical information to predict future data points in the sequence.

Furthermore, RNNs employ a training method called Back-Propagation Through Time (BPTT), an extension of the standard back-propagation algorithm. BPTT adjusts the network's weights based on the errors propagated backwards through time, allowing the network to learn from past data points effectively. This training approach is crucial for the network to accurately capture and utilise the temporal dependencies inherent in sequential data. Figure 3 shows the Recurrent Neuron and Unfolding.

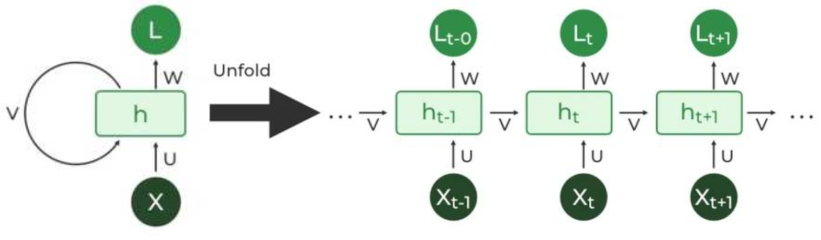


Figure 3. Recurrent Neuron and Unfolding

3.4.1 Random Forest (RF)

A Random Forest is an ML algorithm that leverages the collective power of an ensemble of decision trees. During training, it constructs a multitude of decision trees, each built on a random subset of data points and using a random selection of features for splitting decisions. Each tree uses a selection of data points generated at random and features to make predictions. This randomness helps prevent the trees from becoming too focused on specific details in the training data, improving their ability to handle new, unseen data.

This element of randomness injects diversity into the forest, impeding any single feature from dominating the learning process and potentially introducing bias. Furthermore, to enhance this diversification, a second layer of randomness is introduced at each node within the trees. Here, a subset of features is randomly selected as candidates for splitting the data. This dual approach using random data subsets and random feature subsets fosters a collection of trees with unique decision-making capabilities, ultimately strengthening the overall model's robustness. Figure 4 shows the Random Forest Model Working algorithm.

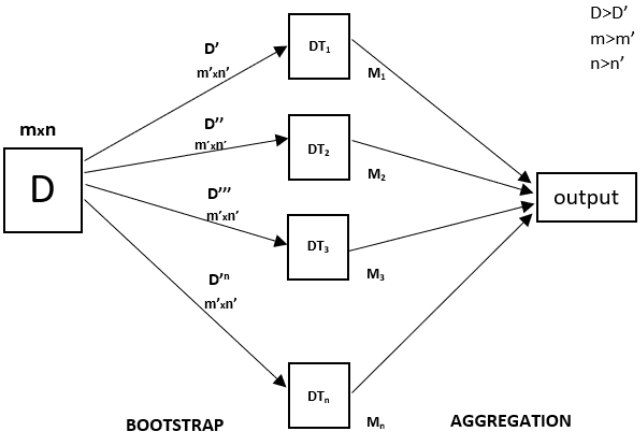


Figure 4. Random Forest Model Working

3.5 Performance Evaluation Metrics

Key performance indicators such as Accuracy, Precision, True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), False Negative Rate (FNR), and F1-Score are used to assess the model's performance and offer more context for its operation after it has been trained and validated as presented in Equations 1-6, [12-15].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$F1\,Score = \frac{2\times Precision \times Recall}{Precision+Recall} \tag{3}$$

$$TPR = \frac{TP}{TP+FN} \tag{5}$$

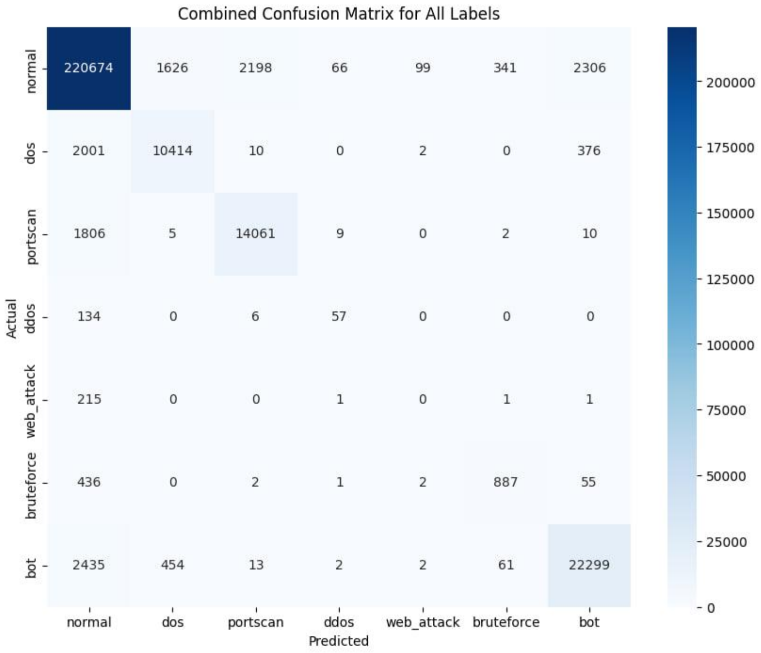
$$TNR = \frac{TN}{FP+TN} \tag{6}$$

TP represents True Positive; TN represents True Negative; FP represents False Positive; FN represents False Negative.

4. RESULTS AND DISCUSSION

Table 5 shows the model’s training performance result and Table 6 shows the interpretation of the result for the above experimental setup. The result of the hybridised intrusion detection system (IDS) experiment demonstrated impressive performance across the training, validation, and testing phases. The system achieved a remarkable training accuracy of 99.76%, indicating its capability to effectively learn and model the patterns within the training data. In the validation phase, the IDS maintained a high accuracy of 96.14%, showcasing the potential of the model to adapt to novel threats. A tabular representation of the performance of a machine-learning model, specifically on a set of test data, is called a confusion matrix. In essence, it divides the examples into these four categories so that accurate and inaccurate predictions may be distinguished easily.

This tool is particularly valuable for evaluating the effectiveness of classification models, which are designed to predict categorical labels for each input instance. By offering a comprehensive view of prediction results, the confusion matrix aids in identifying specific areas where the model excels or needs improvement, thus serving as a crucial metric for model assessment and refinement. Figure 5 shows the multiclassification for the hybridised model's confusion matrix for all labels



Actual \ Predicted	normal	dos	portscan	ddos	web_attack	bruteforce	bot
normal	220674	1626	2198	66	99	341	2306
dos	2001	10414	10	0	2	0	376
portscan	1806	5	14061	9	0	2	10
ddos	134	0	6	57	0	0	0
web_attack	215	0	0	1	0	1	1
bruteforce	436	0	2	1	2	887	55
bot	2435	454	13	2	2	61	22299

Figure 5. Multiclassification for Hybridised Model's Confusion Matrix for all Labels

Table 5. Report on Classification from Hybridised Model Testing

Hybridised RNN and Random Forest						
	Precision	Recall	F1-Score	True Negative Rate (TNR)	False Positive Rate (FPR)	False Negative Rate (FNR)
Normal	0.98	0.98	0.98	91.63%	8.366%	2.30%
DoS	0.89	0.89	0.89	99.48%	0.51%	11.02%
PortScan	0.91	0.90	0.91	99.46%	0.53%	9.6%
Web Attack	0.42	0.30	0.35	99.97%	0.02%	69.5%
Botnet	0.03	0.01	0.02	99.95%	0.04%	98.6%
DDoS	0.71	0.67	0.69	99.86%	0.13%	33.1%
BruteForce	0.90	0.92	0.91	99.01%	0.98%	8.2%
Macro Average	0.69	0.67	0.68			
Weighted Average	0.96	0.96	0.96			

Similarly, the testing phase yielded an accuracy of 96.08%, in contrast to other works as shown in Table 7, further affirming the system's robustness and reliability in detecting and classifying various types of network intrusions. These results highlight the efficacy of combining Recurrent Neural Networks (RNN) for feature extraction with Random Forest classifiers for intrusion detection

Table 6. Interpretations of classification from Hybridised Model Testing

Class	Total Samples	Correctly Predicted	Main Misclassifications	Issues
Normal	Very High	220,674	Portscan, Bot	Minor false positives
DoS	~12,793	10,414	Normal, Bot	~19% false negatives
Portscan	~15,893	14,061	Normal	~11% false negatives
DDoS	~197	57	Normal	High false negative
Web Attack	~218	1	Normal	Critical detection failure
Bruteforce	~1,384	887	Normal, Bot	Moderate errors
Bot	~25,266	22,299	Normal, DoS	Good, but some overlap

Table 7. Comparison with other works

S/N	Authors	Machine Learning Technique	Number of Datasets Used	Results
1	Ours	Hybridised Model of RNN and RF	1	Precision: **96.0%** Accuracy: **96.08%** F1-Score: **96.0%** TPR: **96.0%** TNR: **97.8%** FPR: **1.4%** FNR: **3.29%**
2	[7]	BiLSTM	1	Precision: 86.38% Accuracy: 90.73% F1-Score: 89.65% TPR: 93.17%
3	[9]	DNN and LSTM-RNN	3	Accuracy: 98.68 F1 Score: 98.83 FPR: 2.47
4	[16]	OC-SVM and Supervised Random Forest	1	Accuracy: 95.95% Recall: 99.56% FPR: 0.44% FNR: 7.8%

5. CONCLUSION

The achievements of this research are significant. The hybrid model demonstrated an impressive accuracy of 96.08% during testing, showcasing its ability to classify most network traffic instances correctly. With precision and recall at 96.0%, the model effectively minimised false positives and false negatives, achieving a balance crucial for reducing unnecessary alerts and ensuring that actual threats are not overlooked. The F1 score also agrees with this opinion at 96.0%, indicating that the model effectively handled the tradeoff between precision and recall. In addition, the model had maintained low FPR and FNR—1.4% and 3.29%, respectively, meaning the system was reliable and trustworthy. The model also achieved high sensitivity and specificity with a TPR of 96.0% and a TNR of 97.8%.

Implementing the hybrid model combined RNNs and Random Forests, leveraging the sequential data processing strengths of RNNs and the robust classification capabilities of Random Forests. The Random Forest model successfully classified the features that the RNN model had extracted from the data. The system is flexible and all-encompassing because it was taught to identify several network assaults, such as DoS, DDoS, BruteForce, PortScan, Bot, and Web attacks. There are also some limitations to this model's performance, including the fact that the model is highly dependent on the quality and quantity of the training data, and limited datasets can restrict the model's generalizability. While effective, the hybrid model can be computationally intensive and require significant training and real-time deployment resources.

Although minimised, false positives and negatives indicate room for improvement in the model's detection capabilities. Future implementations should aim to collect more diverse and high-quality datasets to improve the model's generalizability. Integrating real-time data processing capabilities will enable the system to respond immediately to ongoing threats. Ensuring that the infrastructure supporting the NIDS is scalable to handle high volumes of network traffic without compromising performance is also crucial, along with seamless integration with existing security infrastructure to provide a cohesive defence mechanism.

ACKNOWLEDGEMENTS

The authors acknowledge the Covenant University Centre for Research, Innovation, and Discovery (CUCRID), Ota, Ogun State, Nigeria.

REFERENCES

[1] L. Boukela, G. Zhang, M. Yacoub, and S. Bouzefrane, “A near-autonomous and incremental intrusion detection system through active learning of known and unknown attacks,” Oct. 2023, doi: 10.1109/SPAC53836.2021.9539947.

[2] N. Omer, A. H. Samak, A. I. Taloba, and R. M. Abd El-Aziz, “A novel optimized probabilistic neural network approach for intrusion detection and categorization,” *Alexandria Engineering Journal*, vol. 72, pp. 351–361, Jun. 2023, doi: 10.1016/j.aej.2023.03.093.

[3] W. F. Kamil and I. J. Mohammed, “Deep learning model for intrusion detection system utilizing convolution neural network,” Open Engineering, vol. 13, no. 1, Jan. 2023, doi: 10.1515/eng-2022-0403.

[4] S. Ouiazzane, M. Addou, and F. Barramou, “Cyberthreat real-time Detection based on an Intelligent Hybrid Network Intrusion Detection System.”

[5] O. F. Isife, K. Okokpujie, I. P. Okokpujie, R. E. Subair, A. A. Vincent, M. E. Awomoyi, “Development of a malicious network traffic intrusion detection system using deep learning.” International Journal of Safety and Security Engineering, Vol. 13, No. 4, 2023, pp. 587-595. https://doi.org/10.18280/ijsse.130401

[6] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, and F. Ahmad, “Network intrusion detection system: A systematic study of machine learning and deep learning approaches,” Transactions on Emerging Telecommunications Technologies, vol. 32, no. 1, Jan. 2021, doi: 10.1002/ett.4150.

[7] Y. Fu, Y. Du, Z. Cao, Q. Li, and W. Xiang, “A Deep Learning Model for Network Intrusion Detection with Imbalanced Data,” Electronics (Switzerland), vol. 11, no. 6, Mar. 2022, doi: 10.3390/electronics11060898.

[8] A. Hussain, F. Aguiló-Gost, E. Simó-Mezquita, E. Marín-Tordera and X. Massip, "An NIDS for Known and Zero-Day Anomalies," 2023 19th International Conference on the Design of Reliable Communication Networks (DRCN), Vilanova i la Geltru, Spain, 2023, pp. 1-7, doi: 10.1109/DRCN57075.2023.10108319.

[9] A. K. Silivery, R. M. Rao Kovvur, R. Solleti, L. S. Kumar, and B. Madhu, “A model for multi-attack classification to improve intrusion detection performance using deep learning approaches,” Measurement: Sensors, vol. 30, Dec. 2023, doi: 10.1016/j.measen.2023.100924.

[10] V. Hnamte, H. Nhung-Nguyen, J. Hussain, and Y. Hwa-Kim, “A Novel Two-Stage Deep Learning Model for Network Intrusion Detection: LSTM-AE,” IEEE Access, vol. 11, pp. 37131–37148, 2023, doi: 10.1109/ACCESS.2023.3266979.

[11] S. D. Pande, G. R. Lanke, M. Soni, M. A. Kulkarni, R. R. Maaliw, and P. P. Singh, “Deep Learning-Based Intrusion Detection Model for Network Security,” in Lecture Notes in Networks and Systems, Springer Science and Business Media Deutschland GmbH, 2023, pp. 377–386. doi: 10.1007/978-981-99-3177-4_27.

[12] K. Okokpujie, I. P. Okokpujie, O. I. Ayomikun, O.I., A.M. Orimogunje, A.T. Ogundipe “Development of a web and mobile applications-based cassava disease classification interface using Convolutional Neural Network”. Mathematical Modelling of Engineering Problems, 10(1): 119-128, 2023. https://doi.org/10.18280/mmep.100113.

[13] K. Okokpujie, I. P. Okokpujie, A. T. Ogundipe, C. D. Anike, O. B. Asaboro, A. A. Vincent, A.A., “Development of a sustainable Internet of Things-based system for monitoring cattle health and location with web and mobile application feedback”. Mathematical Modelling of Engineering Problems, 10(3):740-748, 2023. https://doi.org/10.18280/mmep.1003023.

[14] J. O. Olayiwola, J. A. Badejo, K. Okokpujie, and M. E. Awomoyi, “Lung-Related Diseases Classification Using Deep Convolutional Neural Network”. Mathematical Modelling of Engineering Problems, 10(4): 1097, 2023. https://doi.org/10.18280/mmep.100401

[15] K. Okokpujie, S. Apeh, “Predictive modeling of trait-aging invariant face recognition system using machine learning”. In: Kim, K., Kim, HY. (eds) Information Science and Applications. Lecture Notes in Electrical Engineering, vol 621. Springer, Singapore, 2020. https://doi.org/10.1007/978-981-15-1465-4_43

[16] S. Kumar and K. J. Singh, “Intrusion Detection System Using Supervised Machine Learning,” in Lecture Notes in Electrical Engineering, Springer Science and Business Media Deutschland GmbH, 2024, pp. 399–409. doi: 10.1007/978-981-99-4713-3_38.

[17] Canadian Institute for Cybersecurity, University of New Brunswick, "Canadian Institute for Cybersecurity," [Online]. Available: https://www.unb.ca/cic/. [Accessed: Nov. 1, 2024].

[18] Canadian Institute for Cybersecurity, "CICIDS 2017 Dataset," University of New Brunswick, 2017. [Online]. Available: https://www.unb.ca/cic/datasets/ids-2017.html. [Accessed: Nov. 1, 2024].

[19] X. J. Li, M. Ma, and Y. Sun, “An Adaptive Deep Learning Neural Network Model to Enhance Machine-Learning-Based Classifiers for Intrusion Detection in Smart Grids,” *Algorithms*, vol. 16, no. 6, Jun. 2023, doi: 10.3390/a16060288.

BIOGRAPHIES OF AUTHORS

Mary Oluwafeyisayo Ogundele holds a BSc in Electronics and Computer Engineering, a master’s degree in information technology and a Masters of Business Administration (MBA) amongst other professional certifications. She currently leads the Network and User Administration Team at Unified Payments Systems, a prominent fintech organization in Nigeria. She served at Nigeria's Apex Bank, where she was responsible for administering and managing critical network and security solutions. She is a member of the Nigeria Society of Engineers (NSE), Council for the Regulation of Engineering in Nigeria (COREN) and a fellow of The National Institute of Professional Engineers and Scientists (NIPES).

Dr. Kennedy Okokpujie holds a Bachelor of Engineering (B.Eng.) in Electrical and Electronics Engineering, Master of Science (M.Sc.) in Electrical and Electronics Engineering, Master of Engineering (M.Eng.) in Electronics and Telecommunication Engineering and Master of Business Administration (MBA), Ph.D in Information and Communication Engineering, besides several professional certificates and skills. He is an associate professor in the Department of Electrical and Information Engineering at Covenant University, Ota, Ogun State, Nigeria. He is a member of the Nigerian Society of Engineers and the Institute of Electrical and Electronics Engineers (IEEE). His research areas of interest include Biometrics, Artificial Intelligence, and electronics and communication engineering. He can be contacted at email: kennedy.okokpujie@covenantuniversity.edu.ng.

Ijeh Princess Adaora is a postgraduate student currently pursuing a Master's degree in Computer Science. She holds a Bachelor's degree in Computer Science from the Federal University of Petroleum Resources, Effurun, Delta State, Nigeria. For her undergraduate research, she developed a web-based electronic management application to improve administrative efficiency. Her current research interests include Cybersecurity and Artificial Intelligence, focusing on developing smart, data-driven systems. Adaora is dedicated to advancing intelligent technologies through secure and innovative software solutions. She aspires to contribute to cutting-edge research in intelligent systems and secure software design, with the long-term goal of enhancing digital infrastructure in developing and developed regions. adaora.ijehpgs@stu.cu.edu.ng

Engr. Dr. Imhade Princess Okokpujie, a Senior Lecturer/Researcher at Afe Babalola University, holds a PhD in Mechanical Engineering from Covenant University, focusing on nano-lubricant in advanced manufacturing. She has authored over 186 scholarly papers and a book on modern optimisation techniques. Dr. Okokpujie has secured three research grants and received multiple awards, including Covenant University's Chancellor's Exceptional Researcher of the Year (2018, 2019) and Afe Babalola University's outstanding research award. She specialises in advanced manufacturing, nano-lubricants, and energy systems, and has been awarded the Global Excellence Stature Fellowship Grant at the University of Johannesburg. A registered COREN engineer and NSE member, she has held leadership roles in her Institution, APWEN and NSE, mentoring numerous students and young researchers. Dr. Okokpujie is passionate about girl-child education and women's development.