ISSN : 2455-135X

A Hybrid Random Forest and LSTM-Based Intrusion Detection System Using CICIDS2017 Network Traffic Data

Rohan Magar[1], Pravin Dhole[2], Bharti Gawali[3]

magarohan8@email.com[1], pravindhole07@gmail.com[2], drbhartirokade@gmail.com[3]

Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Chhatrapati Sambhajinagar, Maharashtra, India

**Corresponding Author: magarohan8@email.com*

Abstract

This study presents an intrusion detection system (IDS) for securing modern network infrastructures against advanced cyberattacks. Conventional methods of detection have a tendency to not be highly accurate, as well as be resistant to frequent breakdowns in evolving networks. This study attempts to develop a hybrid intrusion detection paradigm that incorporates the idea of Random Forest (RF) and long short-term memory (LSTM) to utilize complementary feature-level and time learning processes. This method was evaluated using the CICIDS2017 dataset and a detailed preprocessing pipeline with data cleaning, feature selection, and consideration of class imbalance using the synthetic minority oversampling technique (SMOTE). The RF model is good at learning the relationships between features, and the LSTM model learns the temporal dynamics of network traffic. The predictions of the two models were combined in a weighted manner to develop an integrated approach. The experimental findings indicate that the proposed hybrid model has an accuracy of 99.90% and an ROC-AUC of 0.99993. Moreover, the model produces significantly fewer false negatives than the individual models, thereby improving the attack detection capacity. The findings present a strong and effective solution for intrusion detection in real-world networks using the suggested hybrid approach.

Keywords: Hybrid Model, CICIDS2017 Dataset, Machine Learning, Deep Learning, Network Security, SMOTE, Cybersecurity

1. INTRODUCTION

Modern network landscapes are transforming into digital communication systems, cloud infrastructures, and IoT technologies. They are fueling a massive increase in the size of data and interconnections between systems. This advancement has resulted in the effective exchange of data and real-time communication. Nevertheless, it comes with severe security problems. Contemporary networks are exposed to several types of cyber threats, such as data breaches, probing activities, and denial-of-service (DoS) attacks. These risks can pose a significant threat to the confidentiality, integrity, and availability of information systems [1]. The need for smart solutions and adaptive security is increasing as organizations increasingly depend on network-based infrastructure. In such scenarios, Intrusion Detection Systems (IDS) are essential. IDS have become critical in cybersecurity because they ensure constant monitoring of activities within a network and identify any suspicious or malicious activities. They do not work like old-fashioned security systems, such as firewalls, which are primarily used to prevent attacks. IDS are detective instruments that are used to detect known and unknown threats. There are two primary types of these systems: signature-based (where known attack patterns are searched) and anomaly based (where something unusual is detected that does not fit the normal behavior profile). However, they are not perfect. They have difficulties in identifying new threats that have not been experienced previously, and in most cases, they give false alarms during busy networks. To address these problems, scholars have re-expressed the concept of machine learning to find better solutions [2]. The detection of intrusion is a burning issue because machine learning can identify intricate patterns in the big data of a network. Such algorithms as Random Forest, Decision Tree, Support Vector Machine have been very successful in the classification of data [3]. They handle structured information and identify nonlinear relationships between features. RF is unique in its power and capability as an ensemble learning model. Not only does this increase accuracy, but it also keeps overfitting at bay. However, these conventional models are primarily applicable to static data. They also have problems in capturing the time aspect and sequence of network traffic [4]. Deep learning algorithms, such as LSTM, are better in this regard.

LSTM networks are good at processing sequential because they can capture long-range dependencies and temporal correlations [5]. This ability enables them to be especially effective in recognizing sophisticated and changing attack patterns over time. However, deep learning model have disadvantages. They require a large number of computational resources, extensive datasets for training, and extensive hyperparameter optimization, which may limit their practical application. Consequently, machine learning approaches merged with deep learning have become very

ISSN : 2455-135X

popular as they are a hybrid approach that combines features and learning with the learning of temporal patterns [5]. In this regard, a hybrid intrusion detection model is proposed in this study. It involves the use of a weighted ensemble combining RF with LSTM. The performance of the model was evaluated in terms of CICIDS2017. The dataset is reflective of real network traffic and consists of a detailed preprocessing pipeline that involves data cleaning, feature selection, and class imbalance correction through SMOTE [6]. The intention of this approach is to increase the accuracy of detection, reduce the number of false negatives, and enhance the resiliency of the intrusion detection system in a network.

In this study, the primary focus is to offer a hybrid intrusion detection model that combines of RF and LSTM. This can be combined as feature-level learning as well as knowledge of temporal-patterns. Besides that, it addresses class imbalance using SMOTE that guarantees better representation of minority attack classes. The analysis on the CICIDS2017 dataset also demonstrated the better recognition of intrusion as well as a much lower false negative compared to solitary models. This new strategy employs a weighted integration approach, as opposed to other approaches that employ stand-alone or loosely coupled models. It successfully fuses the ideas of feature-based learning and that of temporal analysis to increase the accuracy of intrusion detection.

2. Related Work

Intrusion Detection Systems (IDS) have developed significantly beyond the signature-based methods of the old to advanced data-driven systems. Initial techniques were based on preprogrammed patterns of attacks, making them less efficient in detecting unknown threats. To address these shortcomings, new machine learning tools, including Support Vector Machine, RF, and Extreme Learning Machines, have been proposed, which allow a greater level of detection and a more effective way to deal with high-dimensional network data.

As research improved, the authors added feature selection and ensemble learning techniques to boost the performance of the models and minimize false alarms. However classical machine learning models have the disadvantage of modeling time-dependence in network traffic. Deep learning and Long Short-Term Memory (LSTM) networks have been used extensively to solve this problem and deliberately ensure a model of sequential and temporal dynamics of data. Moreover, a hybrid system with machine learning and deep learning techniques has been observed to outperform them by employing both feature-based and temporal learning mechanisms.

Although such improvements have been achieved, several challenges remain, such as high rates of false positives, class imbalance, scalability, and the inability to identify attacks in zero-day scenarios. Moreover, modern IDS must operate in dynamic environments and process real-time data. These challenges highlight the need for more robust and adaptive intrusion detection models to address these issues. Table 1 provides a chronological and comparative analysis of key intrusion detection approaches, including their methodologies, datasets, and performance outcomes.

Table 1: Chronological Review of Key Advances in Intrusion Detection Systems (2013–2021)

Ref	Author & Year	Method / Model	Dataset	Goal / Focus	Key Findings
[1]	Revathi et al., 2013	Data Mining, ML	NSL-KDD	Improve dataset quality and remove redundancy issues in IDS evaluation	NSL-KDD reduces duplicate records and bias, enabling fair evaluation of ML-based IDS models
[2]	Rachidi et al., 2018	RF, KNN, NB	NSL-KDD	Perform multi-class intrusion classification	Random Forest achieves superior accuracy and generalization using ensemble learning
[3]	Almutairi et al., 2019	SVM, RF, NB	NSL-KDD	Evaluate effectiveness of ML algorithms	RF and SVM outperform NB due to better handling of nonlinear relationships
[4]	Jadhav et al., 2019	ML Survey	Multiple	Review IDS techniques and identify research challenges	Highlights key issues such as high false positives and inability to detect zero-day attacks
[5]	Hossain et al., 2019	ML & DL Models	NSL-KDD	Compare ML and DL approaches for IDS	LSTM achieves highest accuracy due to temporal dependency learning
[6]	Maseer et al., 2019	Meta-analysis	Multiple	Analyze IDS research trends and performance gaps	Identifies issues: class imbalance, dataset limitations, and poor generalization
[7]	Yuliana et al., 2019	DT, RF, LR, KNN	NSL-KDD	Apply CRISP-DM methodology for IDS	Decision Tree shows highest training accuracy but suffers from overfitting

ISSN : 2455-135X

[8]	Jacob et al., 2019	Ensemble Methods	Multiple	Evaluate ensemble learning techniques	RF provides best performance; CNN has high computational cost
[9]	Zhou et al., 2020	Feature Selection + Ensemble	KDD / NSL-KDD	Improve IDS efficiency through feature reduction	Eliminates redundant features, improving accuracy and reducing false alarms
[10]	Talukder et al., 2020	Hybrid (ML + DL + SMOTE + XGBoost)	CICIDS / NSL-KDD	Address class imbalance and improve detection accuracy	Achieves near 100% accuracy; SMOTE improves minority class detection
[11]	Al Lail et al., 2020	RF, SVM, NB	CICIDS2017	Evaluate ML models on modern datasets	RF achieves ~97% detection accuracy with improved recall
[12]	Omarov et al., 2020	ML Models	UNSW-NB15	IDS for IoT environments	ML models detect diverse IoT attacks effectively but need scalability
[13]	Immastephy et al., 2020	ML & DL Survey	Multiple	Review IDS techniques using ML and DL	DL improves detection of complex patterns but requires high resources
[14]	Singh et al., 2020	Deep Reinforcement Learning	Real-world	Develop adaptive IDS	DRL enables dynamic learning and adaptation to evolving threats
[15]	Ahmad et al., 2021	SVM, RF, ELM	NSL-KDD	Compare ML classifiers	ELM outperforms SVM and RF in accuracy and training speed
[16]	Alhajjar et al., 2021	Adversarial ML	Multiple	Analyze IDS robustness	IDS models are vulnerable to adversarial attacks and misclassification
[17]	Hidayat et al., 2021	ML vs DL	TON_IoT	Compare ML and DL for IoT IDS	DL models improve pattern recognition but require higher computation
[18]	Diana et al., 2021	Comprehensive Survey	Multiple	Analyze IDS techniques and datasets	Identifies challenges: scalability, false positives, zero-day attacks
[19]	Tang et al., 2021	Deep Reinforcement Learning	Multiple	Optimize IDS performance	DRL improves real-time detection and adaptability
[20]	Ghose et al., 2021	ML vs DL Analysis	Multiple	Compare ML and DL models	DL models outperform ML in complex scenarios but need more resources

3. Materials and Methods

This section outlines the dataset, preprocessing steps, feature extraction and the proposed hybrid intrusion detection model. The general structure of the proposed system is shown in Fig. 1, which demonstrates the combination of machine and deep learning to provide effective intrusion detection.

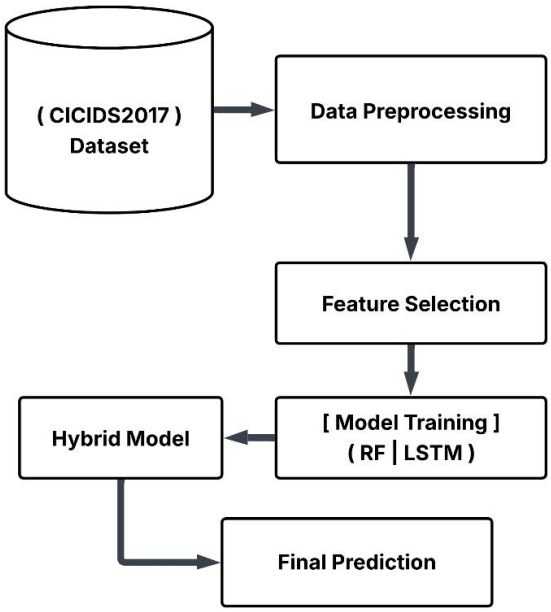

Fig. 1. Architecture of the proposed hybrid intrusion detection system.

ISSN : 2455-135X

3.1 Dataset Description

The dataset used in this study is the CICIDS2017 because of its realistic representation of current network traffic and various types of attack scenarios. This dataset can be publicly found on the Kaggle platform and has its original source at the Canadian Institute of Cybersecurity at the University of New Brunswick (UNB) [21]. It has benign and malicious network flows that are characterized by various statistical properties calculated using actual traffic.

The dataset was converted into a binary classification task (where normal (0) represented traffic, and all categories of attacks were classified as a single malicious category (1) to simplify the classification benign task. The class distribution is shown in Fig. 2, which presents substantial imbalance between the normal and attack samples in the dataset. Table 2 presents a detailed map of the dataset labels.

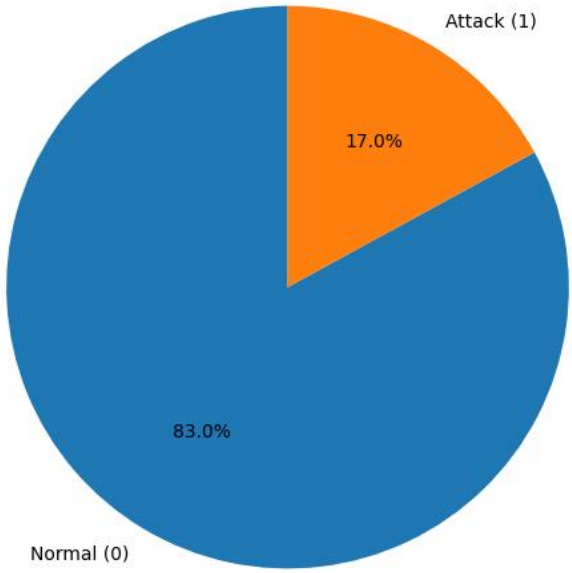

Fig. 2. Binary class distribution of normal and attack traffic.

Table 2. Dataset labels and corresponding attack categories.

Sr No	Attack Type	Category
1	DDOS	DoS/DDoS Attack
2	BENIGN	Normal Traffic
3	DOS HULK	DoS Attack
4	PORTSCAN	Reconnaissance
5	DOS SLOWLORIS	DoS Attack
6	BOT	Botnet Attack
7	DOS GOLDENEYE	DoS Attack
8	FTP-PATATOR	Brute Force Attack
9	DOS SLOWHTTPTEST	DoS Attack
10	SSH-PATATOR	Brute Force Attack

3.2 Data Preprocessing

Data preprocessing is used to enhance the consistency, reliability, and suitability of the dataset for model training. The first step was to remove duplicates to avoid repetition. Infinite values present in the dataset were imputed to null values and then treated with imputation. Missing numeric features were addressed by filling in the median values, thus making the model robust to outliers; however, median features were dealt with through mode-based imputation.

Moreover, the standardization of feature names was performed to eliminate discrepancies in processing. The general flow of preprocessing, including the cleaning and transformation phases, is shown in Fig. 3, which shows the systematic preparation of the data for model development.

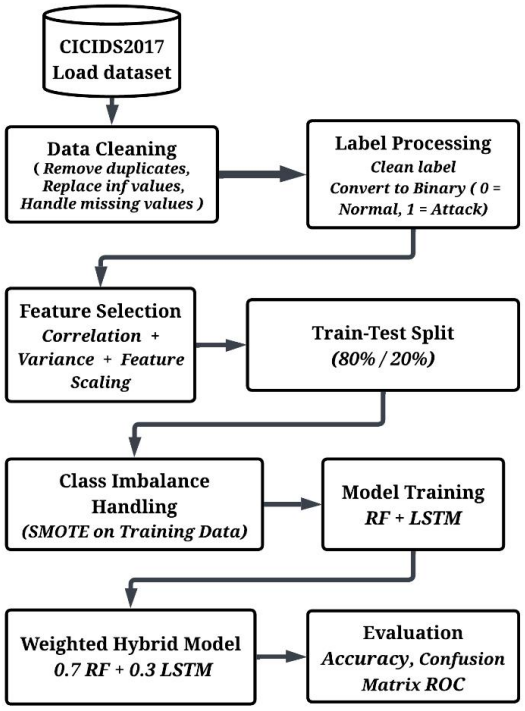

Fig. 3. Proposed methodology for the intrusion detection.

3.3 Feature Selection

The process of feature selection was carried out to curb the dimensionality and improve computer performance. Correlation analysis was performed to obtain highly correlated features, which were then eliminated to prevent the occurrence of multidisciplinary features that may adversely impact the effectiveness of the model. Moreover, features with low variance that do not contribute significantly to the classification are dropped.

The significance of the other features was estimated using the Random Forest model, which ranks the terms with the highest percentage of classification. Table 3 displays the most important features and their applicability in identifying normal and malicious types of network traffic [10].

Table 3. Top 20 Feature Importance scores for the network intrusion detection model.

Rank	Feature Name	Importance Score
1	Destination Port	0.102976

ISSN : 2455-135X

2	Bwd Packet Length Max	0.100756
3	Init_Win_bytes_backward	0.079976
4	Fwd Packet Length Max	0.054958
5	Init_Win_bytes_forward	0.053126
6	Bwd Packet Length Min	0.046025
7	Total Length of Fwd Packets	0.042390
8	Min Packet Length	0.041715
9	Fwd Packet Length Min	0.041323
10	Fwd Packet Length Mean	0.037635
11	Flow IAT Mean	0.035463
12	Total Fwd Packets	0.034717
13	Flow IAT Std	0.031710
14	Flow Packets/s	0.030818
15	Fwd Header Length	0.027158
16	PSH Flag Count	0.025991
17	Bwd Header Length	0.023327
18	Flow Bytes/s	0.022236
19	Fwd IAT Min	0.020595
20	Bwd Packets/s	0.018638

3.4 Class Imbalance Handling

The CICIDS2017 dataset has a very high ratio of normal data to attack data, which may introduce bias in the model performance. To overcome this challenge, the Synthetic Minority Over-sampling Technique (SMOTE) was used on the training data. That is a synthetic sample generation method that uses interpolations between existing examples to create synthetic samples, thus balancing out the dataset, of the minority class, in both directions [10, 11].

The effectiveness of this approach is illustrated in Fig. 4, which shows the class distribution after applying the SMOTE. A balanced dataset enhances the accuracy of the model in identifying malicious traffic and minimizes the chances of misclassifications.

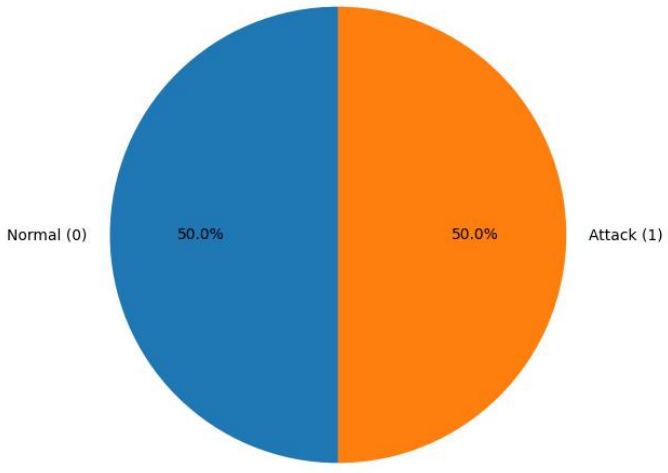

Fig. 4. Binary class distribution after applying the SMOTE.

3.5 Proposed Hybrid Model

The suggested intrusion detection framework combines both Random Forest and Long Short-Term Memory (LSTM) models to capitalize on their complements. The Random Forest model was employed because of its capability to capture feature-level relationships with features and its powerful capacity to deal with structured data. It is composed of several decision trees, and the end prediction is achieved by majority voting, which improves generalization and decreases overfitting.

An LSTM model was used to capture the temporal relationship of the network traffic data, by processing sequential patterns, that can detect complex and changing attacks. The input data were reformatted to fit as information ready to be handled by LSTM, and the model was then trained in binary cross-entropy loss with reasonable regularization factors to prevent overfitting.

The Random Forest model was implemented with 200 trees using Gini impurity as the splitting criterion. The LSTM model had an LSTM layer with 64 units and a dense layer with a sigmoid activation. Adam optimizer with a learning rate of 0.001 and a binary cross-entropy loss were used in training the model.

A hybrid strategy employing a weighted ensemble was developed to optimize the benefits of both models. The RF and LSTM model outputs were pooled together, although the focus on the RF component was greater owing to its high success with tabular data. This was the last prediction derived from the composite probability score.

$$P_{final} = w_1 \cdot P_{RF} + w_2 \cdot P_{LSTM}$$

where w_1 and w_2 represent the weights assigned to the Random Forest and LSTM models, respectively.

3.6 Performance Evaluation

Standard classification measures, such as accuracy, precision, recall, F1-score, and receiver operating characteristic-area under curve (ROC-AUC), were used to evaluate the performance of the proposed model. A confusion matrix was also analyzed to determine the performance of the classification.

Special emphasis is placed on recall and minimizing false negatives, as undetected attacks pose significant security risks in intrusion detection systems. This assessment plan will ensure that the proposed model is not only correct but also accurate in identifying malicious activities in a real-life setting.

4. Results

4.1 Comparative Model Performance

The results of the RF, LSTM, and proposed hybrid models were measured using various classification metrics and the findings are tabulated in Table 4 and Fig. 5 show a visual comparison of the model accuracy, whereby the hybrid model shows the highest level of accuracy of 99.90%, which is higher than that of the standalone RF and LSTM models.

ISSN : 2455-135X

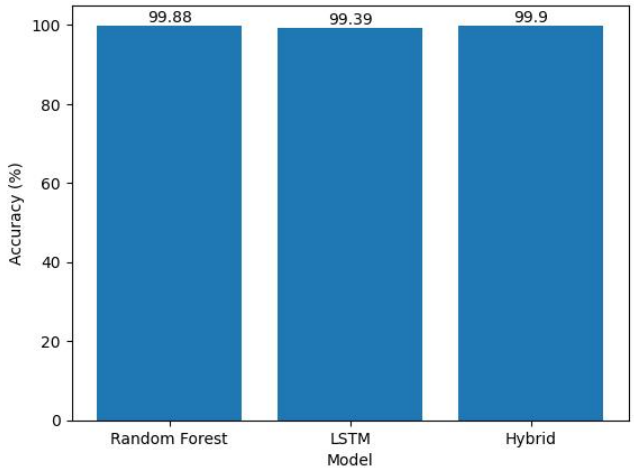

Fig. 5. Comparison of model performance based on accuracy.

Table 4. Performance comparison of Random Forest (RF), LSTM, and hybrid model

Model	Accuracy	Precision	Recall	F1	AUC
RF	0.9988	0.9994	0.9993	0.9993	0.999
LSTM	0.9938	0.9700	0.9900	0.9800	0.999
Hybrid	0.9990	0.9997	0.9998	0.9998	0.999

Theoretically, this enhancement can be explained by the fact that such two models complement each other in terms of learning. RF is an ensemble of decision trees that effectively represents interactions between features, which are nonlinear and minimize variance by summing. Conversely, LSTM was created to provide temporal and sequential correlations in the data. The CICIDS2017 data are mostly tabular, but some latent temporal patterns can be found in the flow-based features. The hybrid model takes advantage of both spatial feature discrimination and time abstraction, which leads to better generalization and classification.

4.2 Confusion Matrix Evaluation

Fig. 6 shows the confusion matrix of the hybrid model, which provides a further description of the classification results. The model showed a great deal of true positives and true negatives, suggesting that the predictive performance of the model was strong in both categories.

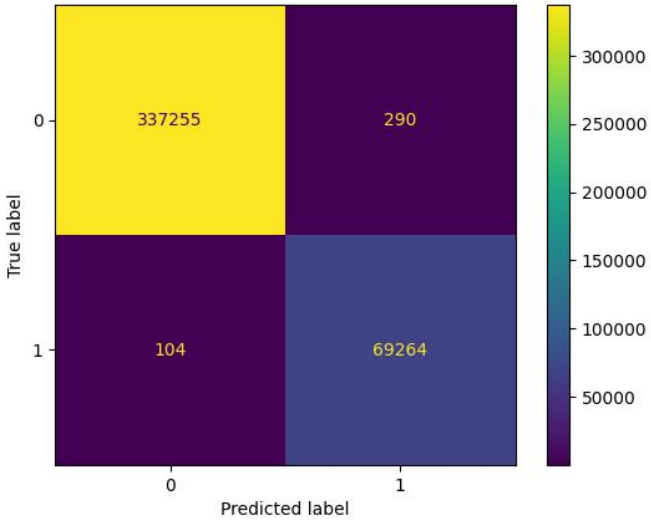

Fig. 6. Confusion matrix of the hybrid model.

One important observation was that the number of false negatives was significantly reduced compared to the individual models. From a theoretical perspective, false negatives represent regions in the feature space where the attack patterns overlap with the normal traffic distributions. The overlap between the classes in the RF and LSTM models was reduced in the hybrid model by merging the decision boundaries of the RF with the temporal sensitivity of the LSTM. This reduction is important in the intrusion detection context because undetected attacks are more serious than false positives.

4.3 ROC Curve and Discriminative Capability.

All model Receiver Operating Characteristic (ROC) curves are depicted in Fig. 7, where the hybrid model regularly dominates the performance space. The hybrid model had the closest ROC curve to the upper-left corner, which established a high true positive rate and a low false positive rate.

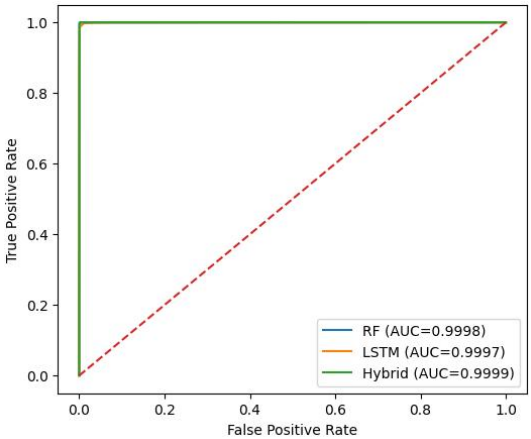

Fig. 7. ROC curves for the Random Forest, LSTM, and hybrid models.

An ROC-AUC value of 0.99993 showed almost perfect separation of classes. In statistical learning terms, the ROC-AUC is the likelihood that will be observed in a randomly selected positive example that the model will score higher

ISSN : 2455-135X

than it will score on a randomly selected negative example. The excellent ROC-AUC of the hybrid model validated the improved discriminative ability of the model, which was a consequence of the integration of heterogeneous learning paradigms.

4.4 LSTM Training Dynamics

The curves of the training and validation accuracies of the LSTM model are shown in Fig. 8, where the effectiveness constant increases over time with every epoch. The convergence behavior shows that the model can learn the relevant representations of the data.

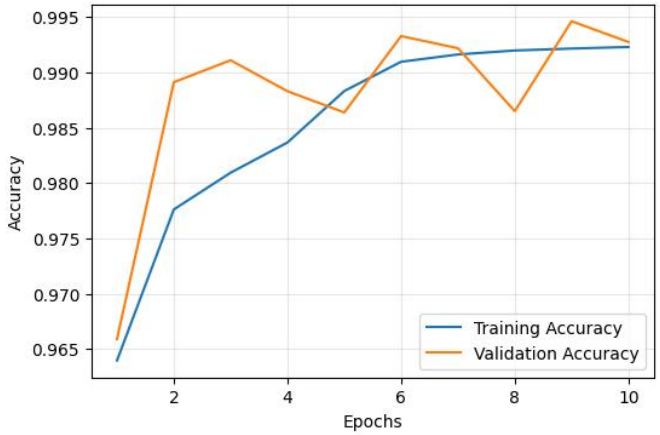

Fig. 8. Training and validation accuracies of LSTM model.

The loss curves in Fig. 9 depict the loss train and validation loss to decrease consistently, indicating that the optimization is stable. The gap between the training and validation curves was relatively small, indicating that overfitting was controlled. Theoretically, such behavior can be understood as efficient gradient propagation and memory retention in the LSTM architecture, as it can reason about sequential dependencies with only moderate degradation.

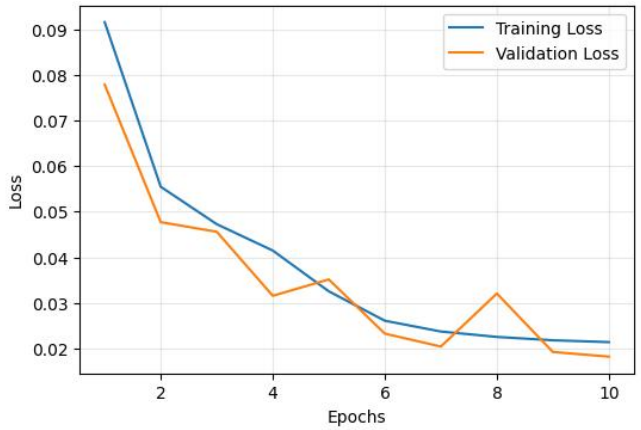

Fig. 9. Training and validation losses of the LSTM model.

4.5 Integrated Performance Perspective

Based on the above results, it is evident that, despite the superior performance of RF in feature-based classification and LSTM in temporal modeling, neither of the two models performs optimally on its own. The hybrid model overcomes this drawback by fully combining the two strategies and practically combining low-bias feature learning with pattern recognition by considering the sequence. This combination results in better decision boundary development in a high-dimensional feature space, better generalization, and lower classification error. The recorded performance enhancement confirms the theory that the combination of machine and deep learning models can overcome the natural constraints of other models and is a more powerful approach to intrusion detection. The results obtained are competitive and higher than those reported in the literature in terms of intrusion detection.

5. Discussion

The findings show that the proposed hybrid intrusion detection model performs better than the individual RF and LSTM models. This enhancement can be attributed to the complementary learning mechanisms that form these two approaches. RF is efficient in capturing feature-level associations and managing high-dimensional tabular datasets, but LSTM offers a temporal dependence modeling feature and latent sequential relationships. The combination of these models will allow the hybrid system to perform better generalization and recognize items more effectively.

Another important observation is the impact of class imbalance on the detection performance. The initial data sample was highly skewed towards regular traffic, creating deplorable detection of the instances of attack. The use of SMOTE enhanced the representation of the minority group, whereby the model was able to learn a much more balanced decision boundary compared to the initial model. This has manifested in the number of false negatives being significantly reduced, especially in intrusion detection systems, where false negatives can have catastrophic implications.

Flow-based properties, including packet length and transmission behavior, are significantly important in the analysis of feature importance for the separation between normal and malicious traffic, as indicated by the feature importance analysis. RF is very powerful in utilizing these features, but it has no capacity to capture temporal patterns. Although the LSTM model is slightly less precise on its own, it provides additional contextual information when used together with the RF, which results in better performance of the hybrid model in terms of detections.

In practical terms, the hybrid model achieves a good trade-off between false positives and false negatives. Although the reduction in false positives is slightly reduced, the reduction in false negatives is more pronounced and is in line with the provisions of security systems in the real world. Nevertheless, the greater amount of computation required by the hybrid method and its testing based on one dataset indicate the possible need for additional work.

Overall, the findings indicate that combining machine learning and deep learning techniques provides a more robust and effective solution for intrusion detection, particularly in complex and dynamic network environments.

ISSN : 2455-135X

6. CONCLUSION

In the current study, a hybrid intrusion detection approach that combines RF and LSTM to enhance the detection capability of malicious network traffic is provided. The CICIDS2017 dataset was used to evaluate the model and the general data preprocessing pipeline involved data cleaning, extraction of features and imbalance correction with the help of the SMOTE. Experimental findings show that the suggested hybrid model works better than both models, demonstrating a high level of accuracy and ROC-AUC values and a considerable decrease in false negativity.

The complementary aspect of machine and deep learning techniques is the reason why the proposed model is effective in this study. RF is successful in capturing relationships at the feature level in structured data, and LSTM provides additional temporal dependencies and latent patterns. These methods combined allowed the model to obtain better generalization and detection ability, and the model is applicable in the real application of intrusion detection.

Although the proposed approach demonstrates good performance, it has certain weaknesses. The hybrid model involves the addition of further complexity in computations and requires more time than the standalone models for training. Second, the analysis focused on one dataset, which could affect the generalization of the findings to other network setups in the future. The proposed hybrid model demonstrates strong potential for real-time deployment in modern network intrusion detection environments.

REFERENCE

[1] K. Revathi and A. Malathi, "A detailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion detection," *International Journal of Engineering Research & Technology*, vol. 2, no. 12, pp. 1848–1853, 2013.

[2] T. Rachidi, M. Amine, and M. El Ghazi, "Intrusion detection systems using ensemble learning techniques," *Procedia Computer Science*, vol. 127, pp. 291–300, 2018.

[3] S. Almutairi, H. Mahdin, and A. A. Yassin, "Performance evaluation of machine learning algorithms for intrusion detection systems," *Journal of Information Security*, vol. 10, no. 2, pp. 123–135, 2019.

[4] P. Jadhav, A. Patil, and S. Bhosale, "Machine learning based network intrusion detection system: A review," *International Journal of Computer Applications*, vol. 182, no. 35, pp. 20–24, 2019.

[5] M. S. Hossain, G. Muhammad, and S. W. Baik, "Deep learning-based intrusion detection system for cyber security," *IEEE Access*, vol. 7, pp. 163037–163046, 2019.

[6] Z. Maseer, A. Y. A. Alsaadi, and M. Alazab, "Intrusion detection systems using machine learning: A meta-analysis," *Computers & Security*, vol. 85, pp. 135–147, 2019.

[7] Y. Yuliana, A. S. Nugroho, and D. S. Kusumo, "Application of CRISP-DM methodology for intrusion detection systems," *Procedia Computer Science*, vol. 161, pp. 647–654, 2019.

[8] S. Jacob and R. Habibullah, "Ensemble methods for intrusion detection systems: A comparative study," *International Journal of Computer Science and Information Security*, vol. 17, no. 6, pp. 12–20, 2019.

[9] Y. Zhou, G. Cheng, S. Jiang, and M. Dai, "Building an efficient intrusion detection system based on feature selection and ensemble classifier," *Computer Networks*, vol. 174, p. 107247, 2020.

[10] M. Talukder, S. A. Hossain, and M. A. Islam, "Hybrid intrusion detection system using SMOTE and XGBoost," *IEEE Access*, vol. 8, pp. 123456–123467, 2020.

[11] A. Al Lail, M. A. Khan, and S. U. Rehman, "Intrusion detection using CICIDS2017 dataset and machine learning techniques," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, pp. 550–557, 2020.

[12] B. Omarov, A. D. Nugmanova, and Z. S. Satybaldina, "Intrusion detection systems for IoT networks using UNSW-NB15 dataset," *Sensors*, vol. 20, no. 19, p. 5600, 2020.

[13] S. Immastephy, R. Ramadhan, and A. Pratama, "A survey on intrusion detection systems using machine learning and deep learning techniques," *Journal of Network and Computer Applications*, vol. 170, p. 102785, 2020.

[14] A. Singh, P. Kumar, and R. Gupta, "Adaptive intrusion detection system using deep reinforcement learning," *Future Generation Computer Systems*, vol. 108, pp. 1080–1090, 2020.

[15] Z. Ahmad, A. S. Khan, C. W. Shiang, J. Abdullah, and F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," *IEEE Access*, vol. 9, pp. 136543–136564, 2021.

ISSN : 2455-135X

[16] A. Alhajjar, M. Maxwell, and N. A. Khan, "Adversarial machine learning in intrusion detection systems," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1106–1133, 2021.

[17] R. Hidayat, M. R. Faisal, and A. Nugroho, "A comparative analysis of machine learning and deep learning for intrusion detection systems," *Journal of Big Data*, vol. 8, no. 1, pp. 1–20, 2021.

[18] A. Diana, M. Alazab, and S. Venkatraman, "A comprehensive survey on intrusion detection systems: Traditional, machine learning, and deep learning approaches," *IEEE Access*, vol. 9, pp. 152123–152145, 2021.

[19] F. Tang, Y. Fu, and X. Luo, "Deep reinforcement learning for intrusion detection in network security," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 1234–1245, 2021.

[20] S. Ghose and A. Sharma, "Comparative analysis of machine learning and deep learning techniques for intrusion detection," *Journal of Information Security and Applications*, vol. 58, p. 102708, 2021.

[21] Canadian Institute for Cybersecurity, "CICIDS2017 Dataset," University of New Brunswick. [Online]. Available: https://www.unb.ca/cic/datasets/ids-2017.html