International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(11s) 2026 pp. 449-462
https://doi.org/10.70917/ijcisim-2026-3766
Article

Hybrid Deep Learning Model for Detecting Malicious Activities in Network Traffic

Yousif A. Saadoon[1*], Riam H. Saffar[2]

[1]College of Medicine, University of Misan, Maysan, Amarah, Iraq
[2]Ministry of Education, Maysan Directorate, Maysan, Amarah, Iraq
Corresponding author: Yousif A. Saadoon, Email: yousif@uomisan.edu.iq

Abstract: As cyber threats have evolved, traditional IDS systems struggle to detect malicious network activities. Deep learning has proven to be promising approache for large-scale network traffic analysis, but with redundant features and class imbalance still remaining in the analysis, it creates difficulty in detecting the network trafficMalicious traffic detection using a hybrid deep learning framework combining Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (BiLSTM) and Attention mechanisms is proposed in this research. CNN extracts spatial traffic features, Bi-LSTM models temporal dependency, and the Attention models relevant traffic patterns. Secondly feature selection is utilized so as to diminish dimensionality and enhance computational economy. A Selective SMOTE strategy is used to boost attack minority classes in order to effectively combat class imbalance. The proposed model was evaluated on large scale network traffic dataset and delivered excellent performance with an accuracy around 97.17% and an ROC-AUC around 99.66.

Keywords: Intrusion Detection System (IDS), Deep Learning, Hybrid CNN-BiLSTM, Attention Mechanism, Feature Selection, Class Imbalance, Selective SMOTE, Network Traffic Classification.

1. INTRODUCTION

With the explosion of digital communication, the Internet of Internet Things (IoT), cloud computing, and Internet services, the volume and complexity of traffic have become critical for cyber security [11]. New types of attacks, including Distributed Denial of Service (DDoS), botnets, brute force, and Advanced Persistent Threats (APT) are commonly found against modern networks [12], and can have an impact on confidentiality, integrity and availability of the systems [6].

The classic IDSs are no longer enough, as they require to have a predetermined list of signature and they are not able to detect new or evolving attacks [7-9]. Furthermore, traditional machine learning methods are based on manual feature engineering and lack of the ability to model complex dependencies in network traffic data across both space and time [8-13].

Although these improvements have been made, there are still some significant issues in the malicious traffic detection research that need to be addressed [9-14]. A significant issue is that of having extremely skewed datasets in the cyber security domain where the number of rare classes of attacks is much less than the normal classes of traffic [10]. Such an imbalance can result in models being skewed towards the dominant class, causing the model to detect minority attacks poorly [15]. One of the difficulties is the presence of features that are redundant and irrelevant, leading to higher complexity in the computation and negative effect on the generalization of the model [16].

Considering the aforementioned drawbacks, in this research, a novel hybrid deep learning model named Hybrid Deep Learning Model for Detecting Malicious Activities in Network Traffic is proposed. The proposed framework adopts CNN, BiLSTM, and the Attention mechanisms, working together to efficiently learn both spatial and temporal features of network traffic Besides, feature selection approaches are integrated into the system to diminish dimensionality and increasing learning efficiency. A class imbalance problem is addressed using a Selective SMOTE

(SSMOTE) based approach, which employs an oversampling strategy selectively, on the minority attack categories avoiding excessive data transformation with synthetic over-sampled data and mitigating over fitting risks.

A large scale network traffic dataset with millions of records and several categories of attack is used for evaluation of the proposed model. The results of the experiments show that the proposed approach not only has high classification performance (with high Accuracy, F1 score and ROC-AUC values) but also shows robustness against minority attack classes. Results showed that integration of deep feature extraction, temporal learning, attention mechanisms and selective balancing strategies can significantly enhance the performance of identifying malicious behavior in the current network environment.

In summary, the following can be listed as the main contributions of this study:

- So, the intention was to propose a hybrid CNN-BiLSTM-Attention architecture for detecting malicious traffic.
- Adding feature selection to remove redundant network traffic features and computational efficiency.
- Adding a Selective SMOTE balancing strategy to adjust the data distribution without the consideration of the heavily changed majority class distribution.
- Running many experiments over a large-sized real-world network traffic trace.
- Set a highly competitive detection performance suitable for scientific publication and real world cyber security application.

2. Related Work

In recent years, new technologies such as machine learning and deep learning have been applied to improve the performance of intrusion detection systems. Various studies have proposed hybrid architectures to improve the accuracy of feature learning and classification.

An RNN-GRU based intrusion detection system for IoT environments that can detect attacks in various IoT layers was designed in [1]. It has been tested on the ToN-IoT dataset and made up to 99% accurate in network traffic and 98% in application-layer data. A hybrid CNN-RNN (HDLNIDS) method was suggested in [2], in which CNN is used for capturing spatial feature and recurrent network is used for capturing temporal feature. It was tested on the CICIDS2018 and delivered an accuracy rate of 98.90 percentage and beat some traditional methods and deep learning methods.

An ANN-CNN-LSTM-RNN model for detecting botnets using UNSW-NB15 dataset was presented in [3]. It also showed high accuracy of 96.98% and high ROC-AUC and PR-AUC, which imply its good generalization ability. In [4] they present a hybrid method that used AdaBoost, Random Forest and deep learning based methods which effectively classify the mobile threats for Android platform.

Network for the detection of DoH was proposed in [5], with the use of lightweight classifiers such as Random Forest and AdaBoost Trees, using PCA and random under-sampling, with up to 99 percent of accuracy but less number of features.

Although this is promising, many of the studies that have been done to date still had some constraints on spatial or temporal features, overly feature engineered, or heavily resampled, which may overfit the model. In the light of this, CNN-BiLSTM-Attention ensemble hybrid framework along with feature selection and Selective SMOTE is proposed to effectively preserve both spatial and temporal features of the network traffic while maintaining a proper balance among the class distribution.

Table 1: Comparative Analysis of Previous Studies

Ref.	Methodology	Dataset	Main Techniques	Accuracy	Limitations
[1]	RNN-GRU IDS for IoT	ToN-IoT	RNN + GRU	99%	Focused mainly on IoT environments
[2]	HDLNIDS	CICIDS2018	CNN + RNN	98.90%	Limited handling of

					class imbalance
[3]	ACLR Hybrid Model	UNSW-NB15	ANN+CNN+LSTM+RNN	96.98%	High architectural complexity
[4]	Android Malware Detection	Android Malware Dataset	AdaBoost+RF+DL	High Accuracy	Limited to mobile malware detection
[5]	Malicious DoH Detection	DoH Dataset	RF+AdaBoost+PCA	99.4%–100%	Focused only on DNS traffic
[26]	CNN-BiLSTM-Attention	UNSW	Hybrid + Attention	98%+	Complex
[27]	BiLSTM-Attention	KDDCUP99, NSLKDD, and CICIDS2017	Sequence learning	98%	No CNN
[28]	CNN-BiGRU	IIoT	Temporal + spatial	97–98%	Limited imbalance handling
[29]	CNN-LSTM	CICIDS2017	Hybrid DL	98%	No attention
[30]	GAN-CNN-BiLSTM	Hogzilla	Data augmentation	99%	High cost
[31]	CNN-LSTM	CICIDS2017	hybrid CNN+ LSTM	98%	No feature selection
[32]	Transformer IDS	NSL-KDD, CIC-DDoS 2019 and UNSW-NB15	Long dependency	99%+	Heavy computation
[33]	Federated GNN	Distributed IDS	Privacy-preserving	High	Complexity
Proposed Work	Hybrid IDS Framework	CICIDS-based Dataset	CNN+BiLSTM+Attention+FS+SSMOTE	97%+ AUC≈0.996	Improved minority handling, reduced feature complexity

Table 1 shows that most existing works focus either on spatial or temporal learning individually, while few integrate attention mechanisms and imbalance handling simultaneously, which motivates our proposed model.

3. MATERIALS AND METHODS

The proposed intrusion detection system is the environment in which it is possible to dissuade intrusion by following the multi-stage processing pipeline. Each stage is responsible for a specific task in developing traffic representation for improvement of the quality of traffic and attack detection capability.

The whole flow of the proposed framework can be summarized as follows:

- Real-time Capture of network traffic.
- Data Pre-processing & Data Cleaning.

- Convert strings into numbers: label encoding and normalization.
- Feature selection.
- Dataset balancing using Selective SMOTE.
- Deep learning features extraction by CNN.
- It is a sequential dependency learning via BiLSTM.
- Attention-based feature refinement.
- Multi-class attack classification

The overall architecture of the proposed system is conceptually shown in Figure 1.

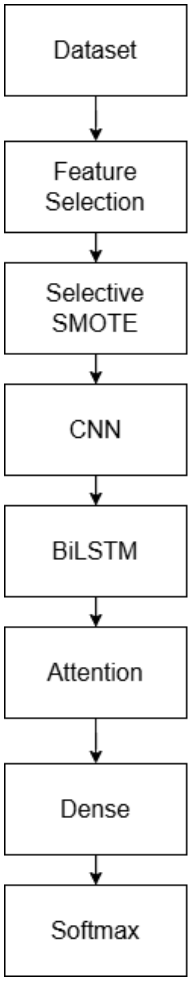

Figure 1: Proposed System

The IDS model proposed in this article aims at being efficient in processing a large amount of traffic on the network, whilst maintaining significant features of an attack.

3.1. Dataset Description

The proposed framework was assessed with CICIDS2017 dataset [17], which is one of the most complete and realistic benchmark datasets in intrusion detection and intrusion prevention security studies and practices.

The dataset originally contained 53 columns, including one column for the target classification. Therefore, 52 columns were used as input features during model training.

The dataset includes: Benign network traffic. There are several contemporary attack types. Several modern attack types. Realistic communication patterns. Statistical flow based traffic features.

Table 2: Dataset Description

Property	Value
Total Records	2,520,751
Total Features	53
Number of Classes	7

There are several malicious activities in the attack category including: Denial of Service (DoS); Brute Force attacks; Web attacks; Distributed attacks; Botnet activities; Infiltration traffic.

The classes are highly imbalanced and makes learning a Difficulty for intrusion detection systems.

Considering that the data is diverse in nature and there are vast number of samples available, only 200,000 samples were considered in experiments with the aim of reducing problem size. To reduce computational complexity while retaining the diversity of the data, a random subset of 200,000 samples was selected for experimental evaluation.

3.1.1 Dataset Sampling Strategy

Given the enormous size of the CICIDS2017 dataset (2,520,751 records), directly training deep learning models on the entire set would have required significant computational resources and training time. Thus, a stratified random sampling technique was used to collect the random sample of 200,000 traffic records ensuring that the samples closely represent the original sample classes. Stratified sampling was done and a fixed random seed (Random State = 42) was applied to obtain the reproducibility. By doing this, the computational burden was greatly decreased and data statistical properties preserved.

Table 3

Stage	Samples
Original Dataset	2,520,751
Sampled Dataset	200,000
Training Set	160,000
Validation Set	40,000

3.2. Data Preprocessing

One of the most important preprocessing stages in intrusion detection systems involves data noise [18], as raw network traffic likely comprises noisy, inconsistent, and incomplete data.

There were numerous operations performed during the pre-processing stage in this study. Missing Value Handling. Infinite Value Removal. Label Encoding.

To get the model convergence in more stable way and to get a better training stability, Normalization using MinMax method was used.

The feature importance scores were estimated using the feature selection Random Forest. Random Forest is a highly useful method for feature ranking because it ranks features based on how well they are able to classify the available observations.

Average impurity reduction over all decision trees was taken to calculate the importance of the features.

The SelectFromModel method, with the median value for feature importance, was used after computing the feature importance values. Any features with importance levels higher than the median value were kept.

The feature selection decreased feature space from: 52 → 26

3.3. Class Imbalance Problem

This imbalance forces deep learning models to favour the majority class over neglecting the minority class attacks.

3.4. Selective SMOTE Balancing

To avoid data leakage, Selective SMOTE was applied only to the training set after the train-validation split. The validation set remained untouched and preserved the original data distribution throughout all experiments.

The balancing strategy consisted of:

Table 3: Selective SMOTE Balancing

Class	**Before SMOTE**	**After SMOTE**
Class 0	130	12,000
Class 1	610	12,000
Class 6	145	12,000

Instead of achieving a perfect balance across all classes to match the size of the largest class, a selective SMOTE strategy was adopted. The sample sizes of the less represented attack classes were increased only to reach an average target size. Such as will not create too many artificial samples, nor engender too large a sample sizing strategy, typically.

Table 4

Class	Before	After
0	130	12000
1	610	12000
6	145	12000

When increasing the oversampling ratio, additional tests further validated that moderate level of selective balancing yielded the optimal balance between minority class representation and generalization of the model.

3.5. Proposed Hybrid Deep Learning Architecture

The proposed architecture of intrusion detection is composed of a set of different deep learning components that aim to take advantage of both space and time traffic features.

The architecture is made up of the following layers: (1) CNN Layer; (2) Batch Normalization; (3) Max Pooling; (4) Dropout; (5) BiLSTM Layer; (6) Attention Layer; (7) Global Average Pooling; (8) Fully Connected Dense Layers; (9) Softmax Classification Layer.

3.6. CNN Layer

A type of net that is extremely capable of learning local spatial patterns in structured data are convolutional neural networks (CNNs [19]. The proposed framework has CNN layer as a major feature extraction element. The CNN layer captures: Local feature correlations; Statistical traffic signatures; Packet-level dependencies; Hidden attack patterns.

The CNN configuration included

Table 4: CNN Configuration

Parameter	**Value**
Filters	64

Kernel Size	3
Activation	ReLU
Padding	Same

The ReLU activation function was chosen because of its computational efficiency [20], and effectiveness in reducing gradient fading problems.

3.7. Batch Normalization and Max Pooling

To help stabilize training and increase convergence speed, Batch normalization was used after convolution.

3.8. BiLSTM Layer

The LSTM network is able to capture the sequential dependencies and temporal patterns [21]. However, most of the classic LSTM methods feed the data in one direction only. Bidirectional LSTM (BiLSTM) has been used to enhance the contextual learning. BiLSTM processes traffic features in both: Forward direction; Backward direction.

In this way, the model can be able to take advantage of richer contextual relationships between the traffic attributes.

The BiLSTM layer contained

Table 5: BiLSTM Parameter

Parameter	Value
Hidden Units	64
Direction	Bidirectional
Return Sequences	True

BiLSTM technology also greatly improves the models' ability to learn advanced time patterns relevant to cyberattacks.

3.9. Attention Mechanism

In recent times, attention mechanisms have been successfully applied to different AI and security applications [22]. BiLSTM layer dynamically calculates the importance score of the learnt traffic representations with the relevance of the attention mechanism [23]. This greatly enhances the ability of the model to be sensitive to complex attack behavior [24].

3.10. Dense Classification Layers

Fully connected with dense layers are fed the feature representations, which are created by sequential learning and feature extraction.

3.11. Training Configuration

The proposed neural network in this research was trained using the improved Adam algorithm [25], due to its ability to adapt to learning while saving on computational tasks. The training parameters are shown in Table 6.

Table 6: Training Parameters

Parameter	Value
Optimizer	Adam
Learning Rate	0.001
Epochs	15
Batch Size	128

Validation Split	20%

Several regularization techniques were applied: Early stopping; Reduce learning rate on plateau; Batch normalization; Dropout regularization.

These techniques improve training stability and reduce overfitting risk.

4. RESULTS

Multiple classification metrics were used to conduct a comprehensive evaluation of the model's performance. The evaluation was based on: Overall classification accuracy; F1-score; Area under the ROC curve; Confusion matrix analysis.

Experimental Environment

All experiments were conducted using Python 3.12 with TensorFlow/Keras for deep learning implementation and Scikit-Learn for preprocessing and evaluation. Training was performed on Google Colab utilizing an NVIDIA T4 GPU, 12 GB GPU memory, 32 GB RAM, and Intel Xeon CPUs. The average training time per experiment ranged between 20–40 minutes depending on the oversampling configuration.

4.1. Experimental Results

The proposed hybrid model was able to obtain the intrusion detection performance, which is quite competitive, using the CICIDS2017 dataset. The results for the last evaluation are summarized in Figure 2:

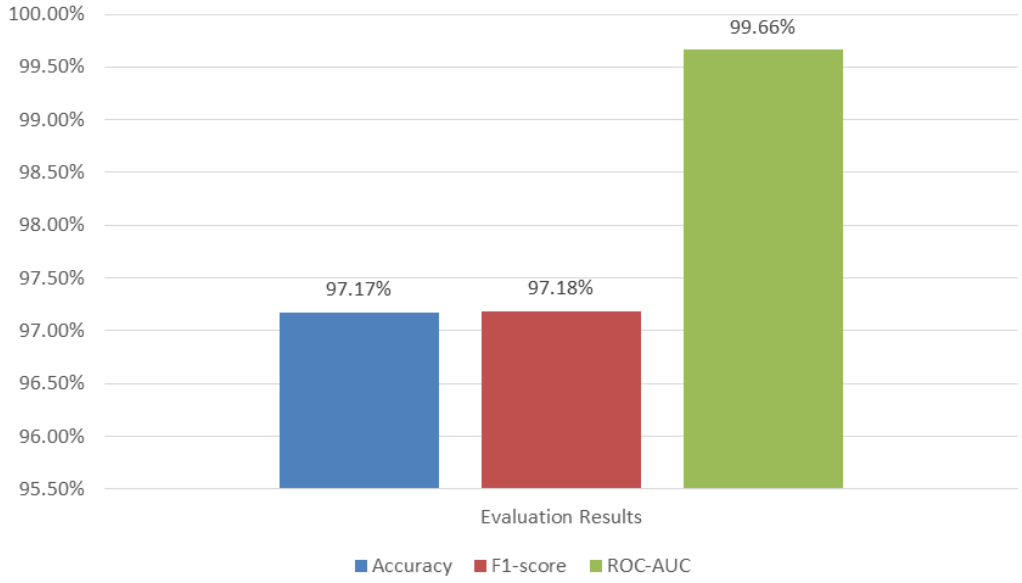

Figure 2

The outcomes of this study show the efficiency of combing the feature selection, selective SMOTE, CNN, BiLSTM, and Attention mechanism together to develop a single intrusion detection system.

The proposed CNN–BiLSTM–Attention model was evaluated on a large-scale network penetration dataset comprising 2,520,751 records with 53 native features the computational efficiency was enhanced by feature selection after pre-processing in which 26 features were determined to be most important and included for discriminatory information.

To overcome the huge class imbalance problem, a selective SMOTE technique was employed only to the minority classes, so as to preserve the class distribution structure of the mainstream classes and get a more close-to-balanced training distribution for the minority classes.

Final data set was split to training and testing set in a ratio of 80:20. Then the model was trained for 15 epochs with the Adam optimizer with an adaptive learning rate schedule.

The proposed model achieved the following performance:

- **Accuracy:** 0.9716
- **Precision (Weighted):** 0.9743
- **Recall (Weighted):** 0.9716
- **F1-score (Weighted):** 0.9718
- **ROC-AUC:** 0.9966

Macro-averaged metrics further indicate the challenge of class imbalance:

- **Macro Precision:** 0.6572
- **Macro Recall:** 0.6527
- **Macro F1-score:** 0.6492

The difference between weighted and macro-averaged results highlights the impact of class imbalance, where the model performs strongly on majority classes while maintaining moderate performance on minority attack categories.

4.2. Classification Performance Analysis

The majority of attack types fared well on the classification report.

The model performed very well in terms of detection in the case of: Normal traffic; Common attack classes; Medium-frequency attack categories.

Nearly perfect classification performance was obtained for several classes for example.

Table 7: Representative per-class F1-scores

Class	F1-score
Class 2	0.99
Class 3	0.98
Class 4	0.99
Class 5	0.98

The results show that the proposed framework is able to detect complex traffic behaviors and is able to distinguish between benign and harmful traffic.

4.3. Impact of CNN Feature Extraction

The CNN layer was a key element of the learning of the local traffic structures and statistical packet patterns. The convolutional filters allowed the model to learn discriminative traffic representations, without the need for manual feature engineering, by itself.

CNN-based feature extraction improved: Spatial pattern learning; Noise reduction; Feature abstraction; Classification robustness.

Adding CNN layers was found to increase the model's capacity to learn the defense attack signatures it was able to uncover in network flow statistics.

4.4. Impact of BiLSTM Sequential Learning

The BiLSTM layer enhanced the model's ability to capture temporal dependencies between traffic attributes. BiLSTM has the ability to maintain sequential contextual information, unlike a traditional feedforward neural network.

This significantly improved: Traffic behavior understanding; Sequential pattern modeling; Long-range dependency learning; Detection of sophisticated attacks.

The inference of attacks whose sequence of traffic was complex was helpful with BiLSTM.

4.5. Impact of Attention Mechanism

The key feature in facilitating better feature weighting and context understanding was the attention mechanism. Attention gave the model the special opportunity to concentrate on the very informative traffic patterns, minimizing influence from unimportant information.

The integration of attention mechanisms improved: Detection sensitivity; Contextual learning; Feature prioritization; Attack discrimination capability.

The attention mechanism also improved model interpretability by emphasizing critical traffic representations.

4.6. Impact of Feature Selection

The feature selection substantially reduced computation complexity and enhanced generalisation of the presented model.

There were several advantages to doing this, namely reducing the number of features from 52 to 26: Faster training; reduced memory consumption; Lower overfitting risk; improved convergence; Better feature interpretability.

The obtained experimental results confirm that it is possible to enhance IDS learning behavior by eliminating redundant traffic attributes.

4.7. Impact of Selective SMOTE

The SMOTE algorithm which was executed by selecting with that method improves the ratio of minority attacks without too much distortion of the distribution. Compared to aggressive balancing, the proposed Selective SMOTE could maintain the realistic traffic characteristics while improving the minority learning.

The strategy to balance improved: Minority attack recall; Model robustness; Learning stability; Attack diversity representation.

But the extreme minority was still a big challenge because of the extreme imbalance and lack of original traffic from these minority groups.

4.8. Comparative Discussion

The proposed framework has several important advantages over conventional machine learning based IDSs.

Table 8: Comparative Discussion

Traditional IDS	**Proposed Hybrid IDS**
Manual feature engineering	Automatic deep feature extraction
Weak temporal learning	BiLSTM sequence modeling
Limited feature focus	Attention-based learning
Sensitive to imbalance	Selective SMOTE balancing
High overfitting risk	Improved generalization
Limited scalability	High robustness

The hybrid integration of CNN, BiLSTM, and Attention mechanisms enables the proposed model to achieve superior intrusion detection performance.

This excellent performance of the proposed framework is largely believed to be due to the synergic face-to-face combination of the components' strengths. The CNN layers are effective at encoding local spatial patterns of data traffic, and the BiLSTM layers are responsible for encoding temporal order in the network streams, in both directions.

In addition, the attention mechanism highlights dynamically the most relevant data traffic features, enabling the model to include only the patterns associated to attacks while blocking irrelevant data traffic information. Feature selection reduces both the noise and computational complexity; the selective SMOTE technique represents the less prevalent attacks, without significantly changing the original structure of the data distribution.

The results show that the proposed hybrid architecture is able to effectively capture both spatial and temporal dependencies in network traffic data. The convolutional layers of the CN do local feature pattern extraction, and the BiLSTM part captures dependencies in the data streams. The attention mechanism is also used to pay attention to the most relevant features, which improves performance further.

Although it shows a fairly high precision, weak performance on highly underrepresented classes because of class imbalance is observable, indicating the class 0 and 6 as weak classes with performance levels of 0.86 and 0.89, respectively, compared to the other classes.Overall high precision, the lower efficiency on highly under-represented classes due to class imbalances is interesting, indicating the class 0 and 6 as weak classes with precisions of 0.86 and 0.89, respectively, compared to the other classes. This restriction is seen in the overall mean metrics, which will be much lower as compared to the weighted mean metrics.

Conversely, the use of selective SMOTE does not overly skew the original distribution of data but adequately de-represents underrepresented classes, so that the learning behaviour is more stable than for simple oversampling.

Statistical Considerations and Limitations

Given the huge amount of data and the difficulty of computing, the experiments were repeated on a fixed seed of random numbers and with only 1 train iteration. Although multi-run statistical validation was recommended for more generalisability claims, this model showed consistent convergence behaviour across epochs, which points to a consistent learning dynamics.

In future work, the use of k-fold cross validation and multiple runs of random seeds will be added to yield statistical confidence intervals and robustness analysis.

Limitations

While the findings are encouraging, there are a number of constraints. First, experiments were carried out on a single reference set, potentially not sufficiently representative of all the network environments in practice. Second, the representation for less common attack categories was enhanced by using the selective SMOTE technique, however, these attack categories still had poor detection accuracy. Last, the model has not yet been tested in a real production setting or explored in an “off-line” test.

Practical Deployment Considerations

The proposed CNN-BiLSTM-Attention framework exhibits promising advantages, including high detection accuracy and the capability to detect multiple categories of attacks, and it holds great potential to be applied to the real-world Cybersecurity domain. The feature selection makes the numbers of input features from 52 to 26 that decreases the computational complexity while shortens the inference time. This reduction will allow it to be deployed in environments such as Security Operations Centers (SOCs), enterprise networks, cloud infrastructures, and Internet of Things (IoT) environments where huge amounts of network traffic need to be continually analyzed.

Moreover, the hybrid design integrates the spatial and temporal analysis of traffic, which makes possible the detection of known and unknown attack behaviours. Using the Selective SMOTE technology reduces the number of under-represented attack classes during training but still preserves the distribution of the dataset. The proposed framework can be implemented together with the existing Intrusion Detection System (IDSs) to make intelligent traffic monitoring and early detection of attacks in practical application scenarios. There is still a space for further optimizations to achieve faster inference response time and cater to high-throughput network settings in the context of immediate deployment, though.

In practice, the proposed approach can assist security analysts distinguish suspicious data flows and reduce the work done to detect malicious actions in the data flow. The high ROC-AUC score of the framework shows a good level of differentiating between normal and malicious data traffic, which is essential in environments where false alarms incur high operational costs. There is, however, a need for more validation on another dataset and real-time network traffic to see to the widespread use up to the point of industry-level implementation.

4.9. Research Contribution

The main findings of this research can be summarized as follows: A hybrid intrusion detection system architecture for deep learning, which integrates CNN, BiLSTM, and Attention, has been proposed. Dimensionality Reduction using Random Forest Based Feature Selection. Using Selective SMOTE for solving severe class imbalance. The intrusion detection accuracy of the data set is high for CICIDS2017. Maximise the strength and the generalization power of the model. Development of an IDS system that can scale to the modern cyber environment.

5. CONCLUSIONS

This paper introduces a novel hybrid deep learning detection system of malicious activity in a network traffic by combining convolutional neural network (CNN), Bidirectional Long Short-Term Memory (BiLSTM) and attentional mechanisms. The proposed model was specially designed to solve the various issues regarding the existing IDS models, including high dimensional feature space, intricate traffic behavior, and class imbalances in cybersecurity datasets.

With this structure one implementing layer is added to the framework, which is feature selection, to reduce the amount of calculation and to guarantee the learning effectiveness and maximize network traffic properties information. Furthermore, a Selective SMOTE technique to control oversampling of minority attack classes while avoiding oversampling of major classes was used to prevent over-fitting without impacting their original distribution.

This hybrid architecture was proven effective in both experimental evaluations to captures both spatial and temporal dependencies, whilst the network traffic flows. CNN layers are successfully applied to derive local traffic patterns and a BiLSTM layer is applied to learn sequential relationship among the traffic features. Furthermore, the attention mechanism improved the model's attention to the most relevant traffic behaviors correlated with malicious activities.

The outcomes show that the proposed framework for multi-class malicious traffic classification evaluated with high performance metrics of Accuracy, Precision, Recall, F1 Score and ROC-AUC, which shows its capabilities to be efficient in terms of multiclass malicious traffic classification. The study also identified the advantages of the strategy optimization and feature selection combination with regard to enhancing the detection ability, especially for minority attack classes.

The proposed system in this paper is versatile and scalable and has good potential for implementation in today's cyber security environments, with intelligent intrusion detection.

Future Work. The proposed framework had some success, but there are several research avenues that can lead to an even more successful and useful system. When paired with future developments in architecture, such as Transformer-based systems and self-attention networks, it can be possible to improve the learning of long-range relationships. Future work could explore how to integrate Transformer-based architectures and self-attention networks with the present models to improve the understanding of the behavior of network traffic and the ability to learn long-range dependencies.

Currently, transformers have raised a great deal of effective models that have been proven to be useful in sequential data analysis, and could be further applied for ID performance enhancement.

Second, further experiments can be carried out on various benchmarks like CICIDS2018, UNSW-NB15, and BoT-IoT, to test the proposed framework's ability to generalize across other network settings and attack conditions.

Third, future work could investigate more sophisticated methods for handling foreboding, such as adaptive oversampling methods, methods using generated traffic, and cost-sensitive learning methods, for better detection of very rare attacks and zero-day threats.

The other line of inquiry is to build an intrusion detection system that can be deployed with real streaming network traffic and via online learners. This would enable the proposed framework to work well within the dynamic and high-speed network environment, where the traffic is updated periodically.

In addition, an explainable artificial intelligence (XAI) suite might also be incorporated to aid in explaining decisions for detection. It would be a positive step to create human understandable explanations for the malicious traffic classification, enhancing the validity of traffic classification and facilitating implementation in critical Cyber security infrastructures.

Finally, the computational effectiveness of the model could be optimized; moreover, light versions of the model can be deployed on a device, like at the edge of networks, IoT, or low-resource/infrastructure Cybersecurity systems.

To conclude, the suggested blend of deep learning and classic approaches provides an excellent basis for future intelligent IDS studies and presents various potential paths for further network security and cyber threat detection development.

References

1. Khan, N.W.; et al. A hybrid deep learning-based intrusion detection system for IoT networks. Math. Biosci. Eng. 2023, 20, 13491–13520.
2. Qazi, E.U.H.; Faheem, M.H.; Zia, T. HDLNIDS: Hybrid Deep-Learning-Based Network Intrusion Detection System. Appl. Sci. 2023, 13, 4921. https://doi.org/10.3390/app13084921.
3. Ali, M.; Shahroz, M.; Mushtaq, M.F.; Alfarhood, S.; Safran, M.; Ashraf, I. Hybrid Machine Learning Model for Efficient Botnet Attack Detection in IoT Environment. IEEE Access 2024, 12, 40682–40699. https://doi.org/10.1109/ACCESS.2024.3376400.
4. Pardhi, P.R.; Rout, J.K.; Ray, N.K.; Sahu, S.K. Classification of malware from the network traffic using hybrid and deep learning-based approach. SN Comput. Sci. 2024, 5, 162. https://doi.org/10.1007/s42979-023-02516-3.
5. Abu Al-Haija, Q.; Alohaly, M.; Odeh, A.J. A lightweight double-stage scheme to identify malicious DNS over HTTPS traffic using a hybrid learning approach. Sensors 2023, 23, 3489. https://doi.org/10.3390/s23073489.
6. Abdulganiyu, O.H.; Tchakoucht, T.A.; Saheed, Y.K. A systematic literature review for network intrusion detection system (IDS). Int. J. Inf. Secur. 2023, 22, 1125–1162.
7. Sowmya, T.; Mary Anita, E.A. A comprehensive review of AI based intrusion detection system. Measurement: Sensors 2023, 28, 100827.
8. Heidari, A.; Jamali, M.A.J. Internet of Things intrusion detection systems: a comprehensive review and future directions. Cluster Comput. 2023, 26, 3753–3780.
9. Zipperle, M.; Gottwalt, F.; Chang, E.; Dillon, T. Provenance-based intrusion detection systems: A survey. ACM Comput. Surv. 2022, 55, 1–36.
10. Diana, L.; Dini, P.; Paolini, D. Overview on intrusion detection systems for computers networking security. Computers 2025, 14, 87.
11. Lone, Aejaz Nazir, Suhel Mustajab, and Mahfooz Alam. "A comprehensive study on cybersecurity challenges and opportunities in the IoT world." Security and Privacy 6.6 (2023): e318.
12. Amoo, Olukunle Oladipupo, et al. "Cybersecurity threats in the age of IoT: A review of protective measures." International Journal of Science and Research Archive 11.1 (2024): 1304-1310.
13. Kumar, Avinash, and Jairo A. Gutierrez. "Impact of machine learning on intrusion detection systems for the protection of critical infrastructure." Information 16.7 (2025): 515.
14. Shanmugam, Vaishnavi, Roozbeh Razavi-Far, and Ehsan Hallaji. "Addressing class imbalance in intrusion detection: A comprehensive evaluation of machine learning approaches." Electronics 14.1 (2024): 69.
15. Abdelkhalek, Ahmed, and Maggie Mashaly. "Addressing the class imbalance problem in network intrusion detection systems using data resampling and deep learning: A. Abdelkhalek, M. Mashaly." The journal of Supercomputing 79.10 (2023): 10611-10644.
16. Al-Qarni, Elham Abdullah, and Ghadah Ahmad Al-Asmari. "Addressing imbalanced data in network intrusion detection: A review and survey." International Journal of Advanced Computer Science & Applications 15.2 (2024).
17. https://www.unb.ca/cic/datasets/ids-2017.html
18. Ayantayo, Abiodun, et al. "Network intrusion detection using feature fusion with deep learning." Journal of Big Data 10.1 (2023): 167.
19. Fan, Ching-Lung. "Multiscale feature extraction by using convolutional neural network: Extraction of objects from multiresolution images of urban areas." ISPRS International Journal of Geo-Information 13.1 (2023): 5.
20. Bai, Yuhan. "RELU-function and derived function review." SHS web of conferences. Vol. 144. EDP Sciences, 2022.
21. Al-Selwi, Safwan Mahmood, et al. "LSTM inefficiency in long-term dependencies regression problems." Journal of Advanced Research in Applied Sciences and Engineering Technology 30.3 (2023): 16-31.
22. Ravi, Vinayakumar, et al. "A Multi-View attention-based deep learning framework for malware detection in smart healthcare systems." Computer Communications 195 (2022): 73-81.
23. Sivamohan, Sivanandam, and S. S. Sridhar. "An optimized model for network intrusion detection systems in industry 4.0 using XAI based Bi-LSTM framework." Neural Computing and Applications 35.15 (2023): 11459-11475.
24. Kalita, Emi, et al. "Predicting student academic performance using Bi-LSTM: a deep learning framework with SHAP-based interpretability and statistical validation." Frontiers in Education. Vol. 10. Frontiers, 2025.
25. Reyad, Mohamed, Amany M. Sarhan, and Mohammad Arafa. "A modified Adam algorithm for deep neural network optimization." Neural Computing and Applications 35.23 (2023): 17095-17112.
26. D. Shou, C. Li, Z. Wang, and Y. Wang,“An Intrusion Detection Method Based on Attention Mechanism to Improve CNN-BiLSTM Model,” The Computer Journal, vol. 67, no. 5, pp. 1851–1865, 2024.
27. J. Zhang et al.,“A Network Intrusion Detection Model Based on BiLSTM with Multi-Head Attention Mechanism,” Electronics, vol. 12, no. 19, 2023.
28. K. Yang, J. Wang, and M. Li,“An improved intrusion detection method for IIoT using attention mechanisms, BiGRU, and Inception-CNN,” Scientific Reports, vol. 14, 2024.

29. Gueriani, H. Kheder, and A. C. Mazari,“Enhancing IoT Security with CNN and LSTM-Based Intrusion Detection Systems,” arXiv, 2024.
30. Benchama and K. Zebbara, “Novel Approach to Intrusion Detection: GAN-MSCNN-BiLSTM with LIME Predictions,” arXiv, 2024.
31. T. N. Ghorsad and A. V. Zade,“Hybrid CNN+LSTM Deep Learning Model for Intrusions Detection Over IoT Environment,” IJRITCC, 2023.
32. C. Xi et al.,“A Novel Multi-Scale Network Intrusion Detection Model with Transformer,” Scientific Reports, 2024.
33. J. Wang et al.,“Federated Learning for Network Attack Detection Using Attention-Based Graph Neural Networks,” Scientific Reports, 2024.
34. Z. He, X. Wang, and C. Li,“A Time Series Intrusion Detection Method Based on SSAE, TCN and BiLSTM,” Computers, Materials & Continua, 2024.
35. F. Al Tfaily et al.,“Graph-Based Federated Learning Approach for Intrusion Detection in IoT Networks,” Scientific Reports, 2025.