Developing a Hybrid Machine Learning Model for Enhanced Cyber-Security Threat Identification

Samir Fuad Ibrahim
HiLCoE School of Computer Science and Technology
samir.fuadib@gmail.com

Animaw Kerie
HiLCoE School of Computer Science and Technology
animkmu@gmail.com

Abstract

Malicious use of technology is still a public threat and protecting our digital infrastructure across the globe requires solid intrusion detection solutions. We examine the establishment of a machine learning based model for improved cyber threat analysis utilizing the open-source dataset presented by the CSE-CIC-IDS2018. It uses Feedforward Neural Network (FNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) networks to test these models. Moreover, a CNN-RNN (LSTM) hybrid model is proposed to combine the spatial characteristic extraction feature of CNN and the sequential learning ability of LSTM.

The experimental results show that the LSTM and CNN models performs best with an accuracy index of 97.00% each, and the ROC curve showed very strong classification capability. Up next were the FNN model with an accuracy of 96.98%, and CNN-RNN hybrid model achieving an accuracy of 96.94% and an F1-score of 95.96%. The findings highlight the potential that deep learning techniques hold when it comes to identifying cyber threats, with LSTM and hybrid models outperforming conventional means. This research has implications for the field of AI and cybersecurity by contributing to the development of effective intrusion detection systems and highlighting the importance of optimizing deep learning architectures for threat analysis. In future research, real-time deployment strategies and testing the proposed model on different cybersecurity datasets could be explored for improved applicability in practice.

Keywords: Intrusion Detection System (IDS), Deep Learning, CNN–LSTM Hybrid Model, Cybersecurity Threat Classification, CSE-CIC-IDS2018 Dataset

1. Introduction

Cybersecurity is the process of putting security measures in place to guarantee data availability, confidentiality, and integrity. Developing computer systems that can identify patterns in historical data, learn from it, and draw logical conclusions with little to no human

input is the goal of machine learning, a subfield of artificial intelligence. In response to the increasing risks and vulnerabilities in our digital world, machine learning techniques have evolved into helpful tools for improving cybersecurity measures.[1]

In the current world we live in, the importance of cybersecurity has massively grown, with organizations in various sectors facing highly increased risks from cyber threats. When more businesses and institutions adopt digital platforms, the frequency and complexity of cyberattacks continue to evolve. Although the use of firewalls, Intrusion Detection Systems (IDS), and antivirus programs is widespread, traditional security measures often fall short when it comes to preventing advanced attacks. Modern threats are becoming more and more sophisticated, which is making it easier for attackers to bypass traditional defenses. Much like web and mobile technologies, attack tactics are evolving very quickly to breach systems and evade generic signature-based defenses. Machine Learning's ability to analyse large datasets and identify patterns that may be easily missed by human operators makes Machine Learning a very powerful tool to be used for threat detection. It has emerged as a promising solution to address the limitations of traditional security methods, while also improving the accuracy of threat detection by minimizing false positives and ensuring that security teams can focus on the most significant risks[2-5].

The more cyber threats grow in complexity and frequency, traditional cybersecurity systems such as firewalls and antivirus programs are increasingly becoming ineffective against sophisticated attacks. As organizations and individuals increasingly rely on digital systems, they face a growing risk of cyber-attacks, including but not limited to polymorphic malware, Distributed Denial of Service (DDoS), Bot attacks, SSH-Bruteforce, and FTP-BruteForce. Traditional security measures and prevention techniques are often insufficient to protect against these evolving threats, particularly those originating from compromised internal nodes. Current cybersecurity solutions often generate a high volume of false positives, which can overwhelm security analysts and reduce the efficiency of incident response to genuine threats. Furthermore, the dynamic and evolving nature of cyberattacks, combined with resource constraints, exacerbates the limitations of traditional detection systems. These challenges highlight the growing need for more advanced and intelligent threat detection mechanisms. Machine learning techniques, including models such as Random Forests and neural networks, have demonstrated significant potential in improving detection accuracy by analyzing network traffic data and identifying complex patterns associated with malicious activity in near real time[6, 7].

To address the challenge of developing proactive defense mechanisms against increasingly sophisticated cyberattacks, the following research questions are formulated:

RQ1: Which machine learning models are most suitable for classifying different types of cyber threats?

RQ2: How effective is the proposed model in terms of performance metrics such as accuracy, precision, recall, and F1-score?

RQ3: How does the performance of the proposed model compare with other machine learning models under a consistent experimental framework?

This study makes the following key contributions:

1. Comprehensive Comparative Analysis:

A systematic evaluation of FNN, CNN, RNN, and LSTM models is conducted under a consistent experimental framework using the CSE-CIC-IDS2018 dataset, enabling a fair comparison of their effectiveness in cyber threat classification.

2. Hybrid Model Assessment Beyond Accuracy:

The CNN–LSTM hybrid architecture is implemented and evaluated, demonstrating that while hybridization provides demonstrates competitive performance and the ability to capture both spatial and temporal features, it does not necessarily outperform standalone models in terms of accuracy, highlighting the importance of architectural efficiency.

3. Detailed Class-Level Performance Analysis:

The study provides an in-depth examination of false positive rates (FPR) and AUC scores across multiple attack classes, identifying infiltration attacks as a persistent challenge due to their subtle and stealthy characteristics.

4. Realistic and Reproducible Experimental Framework:

Unlike studies reporting near-perfect accuracy, this work emphasizes transparency by using a publicly available dataset, clearly defined preprocessing steps, and standard evaluation metrics to ensure reproducibility and practical relevance.

5. Insight into Practical Deployment Considerations:

The research highlights the trade-offs between model complexity and performance, offering insights into selecting efficient models suitable for real-world cybersecurity environments with limited computational resources.

2. Literature Review

Machine learning (ML) is a branch of artificial intelligence (AI) that gives systems the ability to learn from data, identify patterns, and make decisions with low human intervention [8]. While the other side of the coin in this paper, namely cyber threats are dynamic by nature which begs the need for organizations to adopt methods that are advanced for the identification of threats and a swift response to said threats [9]. Criminals in the cyber sphere continuously evolve their strategies, employing techniques like artificial intelligence and machine learning to bypass conventional defenses. This escalating complexity necessitates enhanced methods of identification to leverage advanced analytics and automation to detect

threats in real-time. Without making improvements to threat identification capabilities, organizations risk prolonged exposure to vulnerabilities, which might result in significant operational and financial consequences. Due to this Machine learning has become a huge game-changer in cybersecurity, by giving systems the ability to spot, prevent, and respond to threats in ways that adapt over time [10].

Safitra et al [11] offer an in-depth evaluation of the exponential growth of cyber threats as a result of digitalization while noting the need for organizations to grow where digital technologies provide opportunities without compromising security. Their study points out that existing security measures are not enough to tackle the high level of threat that exists today.

Salem et al [12] delve into reviewing the use of artificial intelligence in cyber defense detection and compare the limitations of the older tools with the capabilities that modern tools powered by AI have in responding to threats as they arise. Their study evaluates how three main machine learning techniques are used: using unsupervised methods to identify threats by detecting anomalies and to find threats that are not known, where Decision Trees, Support Vector Machines and Random Forest techniques are used for known threat detection, employing supervised techniques, and lastly, adapting to new threats using reinforcement learning techniques. The authors are also quick to point out the emergence of deep learning as an effective tool in detecting threats, more so with the use of deep neural networks which can learn the necessary features from the input raw data, the Convolutional Neural Networks and Recurrent Neural Networks being especially useful for malware and intrusion detection tasks respectively.

Ozkan-Ozay et al [13] report their observation that there is growing interest in the applications of machine learning in cybersecurity and how its techniques can be used in practical areas like systems for detecting intrusion, classifying malware, assessing vulnerabilities, and the like. The advantages pulled from various machine learning techniques are described. For instance, while supervised learning approaches are typically designed to search for specific threats, such approaches are less useful for real-time threat analysis and response, for which in this case unsupervised learning is more appropriate as it is capable of finding new threats that are not described in its training data, especially in the case of spam filtering and malware. Instead, deep learning is a fast-rising technology mostly applied in intrusion detection and behavioral profiling systems as they excel at processing vast amounts of information. The authors, while analyzing previous literature present that beyond being able to make the most dependable predictions in training supervised systems in the face of known threats; existing models have certain limitations in dealing with threats which are not present in the training data, colloquially referred to as “zero-day” attacks. In contrast to supervised models which are quite rigid in their structures, the unsupervised

however are more adaptable to novel forms of attacks, but at the expense of increased false positives.

Apruzzese et al. [14] Discuss how machine learning can revolutionize the field of cybersecurity, in particular the ways of mitigating the modern cyber threats, especially concerning traditional fixed-rule and manual monitoring approaches. However, there is still a huge disparity between what has been researched and what is deployed in practice. The analysis is built through three main understandings of the application of machine learning to security issues, namely, detection of intrusions on a network, detection of malware and detection of phishing attacks, where the use of machine learning shows a higher detection rate and lower false alarm rate as opposed to the other techniques, with most of its application in clustering and deep learning in the network intrusion detection systems.

The study conducted by Bouchama et al [15] offers an in-depth review of machine learning for cybersecurity domains with an emphasis on supervised, unsupervised and hybrid techniques for the detection of anomalies within network traffic. Their study indicates that, out of all the predictors and prediction paradigms, it is the supervised learning methods that have the competitive edge and in particular Convolutional Neural Networks (CNN) and Deep Neural Networks (DNN) that scored 0.981 and 0.976 accuracy respectively as well as 0.993 and 0.991 detection rates respectively which are far greater than the performance of classical methods such as random forests and Naive Bayes.

Kim et al. [16] applied the Knowledge Discovery and Data Mining (KDD) CUP 1999 and Canadian Institute for Cybersecurity Intrusion Detection System 2018 (CSE-CIC-IDS2018) datasets on different CNN architectures to detect DoS attacks. CNNs exploited spatial feature extraction by representing traffic as RGB and grayscale images. When testing the performance of different CNN architectures, models that were RGB-based achieved accuracy greater than 99% in binary classification and very good performance at multiclass classification, whereas using grayscale and RNN lead to lower performance. It shows CNNs' effectiveness in distinguishing advanced DoS variants while noting RNNs' limitations in multiclass temporal analysis, emphasizing image-based CNNs for improved intrusion detection is highlighted in the study.

Hagar & Gawali [17] contribute to deep learning-based intrusion detection systems by comparing the effectiveness of CNN and LSTM models on the modern CSE-CICIDS2018 dataset. They address key challenges in IDS particularly the class imbalance problem, through the use of upsampling and downsampling techniques. The study provides big results over 14 attack types by achieving results of 98.31% for CNN and 98.15% for LSTM. Interestingly, LSTM achieved lower loss score values at the cost of a significantly longer computation time while CNN showed significantly lower computation time. The study makes significant technical contributions by tuning hyperparameters and reporting

extensive performance metrics (e.g., precision, recall, F1-scores) for every attack class. One limitation to discuss is that the work depends on pre-extracted features instead of raw network traffic data, which the authors admit is a topic they would like to improve on in the future to enable the neural networks to learn their features.

Sarker et al. [18] Address the growing importance of cybersecurity in today's digital landscape by proposing IntruDTree, which is a novel machine-learning-based intrusion detection model. The model aims to enhance prediction accuracy while reducing computational complexity by prioritizing and selecting key security features. By experimenting on cybersecurity datasets, they demonstrate the effectiveness of IntruDTree in detecting cyber intrusions and outperforming traditional machine learning methods.

A survey and analysis of intrusion detection models based on CSE-CIC-IDS2018 [19] collected evaluation on various proposed models as shown in the table 2.1 below:

Table 1. Evaluation of different models on the CSE-CIC-IDS2018 dataset

Authors	Accuracy (%)	Precision (%)	Recall (%)	AUC	Proposed Model(s)
Atefinia and Ahmadi	100.00	100.00	100.00	n/a	Modular Deep Neural Network
Basnet et al.	99.00	n/a	n/a	n/a	MLP
Catillo et al.	99.20	95.00	98.90	n/a	Deep Autoencoder
Chadza et al.	97.00	n/a	n/a	n/a	Baum Welch, Viterbi
Chastikova and Sotnikov	n/a	n/a	n/a	n/a	LSTM
D'hooge et al.	96.00	99.00	79.00	n/a	XGBoost
Ferrag et al.	97.38	n/a	98.18	n/a	RNN, Deep Autoencoder
Filho et al.	n/a	100.00	100.00	n/a	Random Forest
Fitni and Ramli	98.80	98.80	97.10	0.94	Logistic Regression + Decision Tree + Gradient Boosting
Gamage and Samarabandu	98.40	97.79	98.27	n/a	Deep Feed-forward Neural Network
Hua	98.37	98.14	98.37	n/a	LightGBM
Huancayo Ramos et al.	99.99	100.00	99.99	n/a	Random Forest, Decision Tree
Kanimozhi and Jacob	100.00	100.00	100.00	1	MLP
Kanimozhi and Jacob	99.97	99.96	100.00	1	MLP
Karatas et al.	99.69	99.70	99.69	n/a	Adaboost
Kim et al.	99.99	81.75	82.25	n/a	CNN
Li et al.	n/a	n/a	100.00	1	Deep Autoencoder
Lin et al.	96.20	96.00	96.00	n/a	LSTM
Zhao et al.	97.90	98.00	98.00	n/a	Deep Autoencoder

Although the use of Machine Learning (ML) in detecting cyber threats is on the rise, there are considerable gaps in the current research that prevent the practical and the widespread use of these models. One huge issue is the rate of false positives; false positives continue to plague many current systems. Excessive alerts not only lead to alert fatigue among security analysts but also reduce the effectiveness and credibility of the detection system. There is a need for intelligent and context-aware models that can reduce noise and focus on truly

malicious behavior. Another major drawback of the literature is the lack of hybrid approaches. This would be important to demonstrate considering that cyber threats are very dynamic and complex, and no single ML algorithm is likely to model all possible patterns and anomalies effectively every time. An interesting approach could be based on combining different models like CNNs for feature extraction, RNNs or LSTM for sequence learning where they work together to create a more robust and accurate system overall; It is notable that studies that specifically explore the development, assessment and refinement of such hybrid machine learning models, are limited and underdeveloped.

3. Proposed Solution

The research design for this paper mainly constitutes the construction, development, and assessment of a machine learning model with the target of efficiently classifying cybersecurity attacks. By utilizing literature review as a crucial base to the methodology, since real world challenges and practical implementations are required, which are also essential parts for cyber security threat detection. As can be inferred from Figure 1, by following a process of identifying the problem, developing the model, evaluating the model, and validating the model; it facilitates the possibility that the proposed solution is valid and pragmatic in the space of cybersecurity.

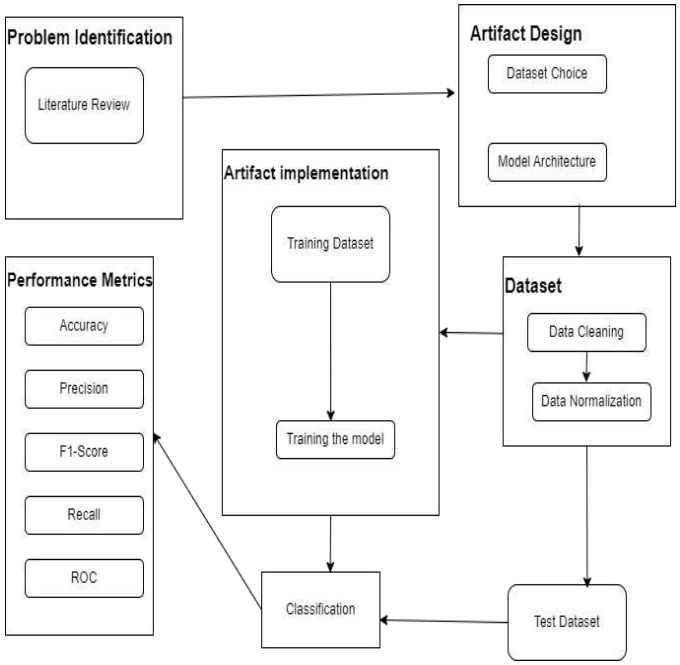


Figure 1. Research Design Process

As you can see from figure 2, the architecture of the proposed model has two key components: a CNN for extracting patterns and an LSTM to capture the temporal dependencies.

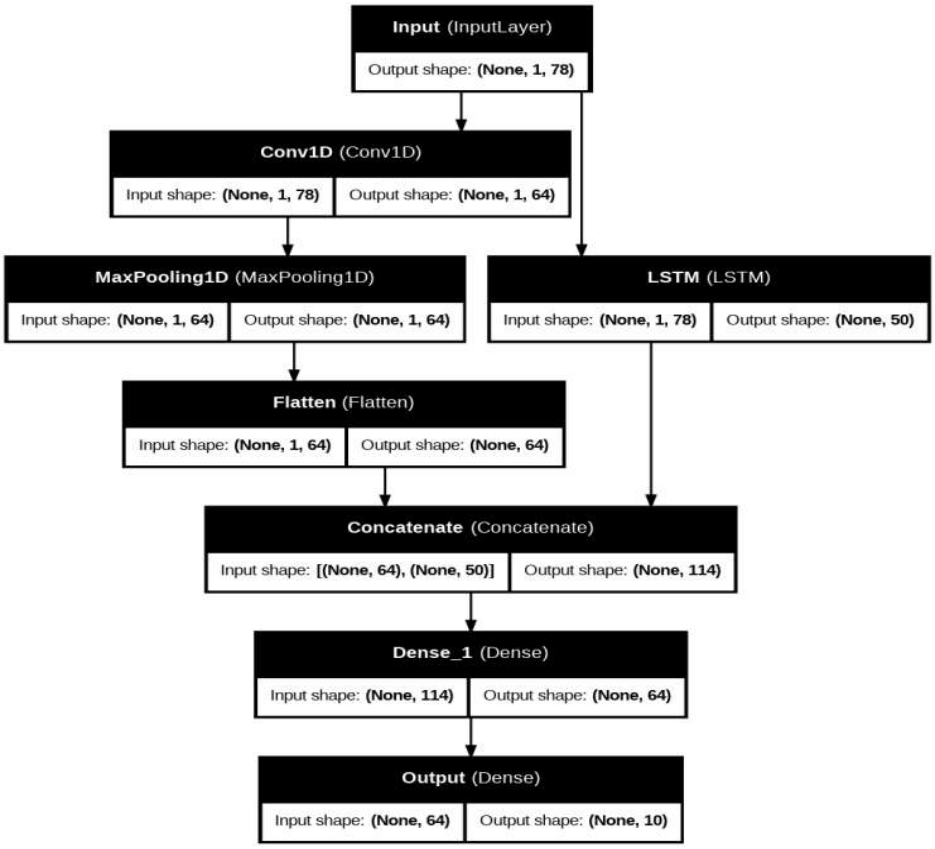


Figure 2. Hybrid Model architecture

A Convolutional Neural Network (CNN) is highly effective at extracting spatial features and local dependencies, particularly from structured tabular data or reshaped time-series input, by learning hierarchical feature representations that capture subtle, localized patterns distinguishing various classes of cyber traffic. Although in cases of recognizing complex, evolving attack behaviors such as infiltration or slow-paced brute force attempts which are an essential capability; CNNs alone are limited in modeling temporal relationships over longer sequences. To address this deficiency, hybrid CNN-LSTM architectures have emerged as a powerful solution. Hence, CNNs are combined with Long Short-Term Memory (LSTM) networks, which excel at capturing temporal dependencies and maintaining contextual information across time steps. By augmenting the model to track the evolution of features over time, LSTM layers provide a deeper understanding of attack sequences. The model will try to leverage the spatial learning capabilities of CNNs alongside the temporal modeling strengths of LSTMs via integrating CNN and LSTM into a hybrid architecture. More accurate and reliable threat classification can come as a result of the model's ability to distinguish between diverse attack types with greater precision while also reducing false positives is enhanced by this fusion.[20-23]

The steps taken to make sure the CSE-CIC-IDS2018 dataset can be well utilized by the model:

- **Data Preprocessing:** preprocessing was performed on the dataset before it was used for model training and evaluation to ensure the integrity and consistency of the dataset. These stages effectively handled any null values while optimizing the dataset based on machine learning needs.
- **Undersampling:** To address the significant class imbalance present in the CSE-CIC-IDS2018 dataset, an undersampling technique was applied to the majority (benign) class. This approach reduces the dominance of benign traffic instances, which would otherwise bias the model toward majority class predictions and hinder its ability to detect minority attack patterns. In the context of this study, the primary objective is to enhance cyber threat identification rather than achieve perfect benign classification. Therefore, reducing the number of benign samples does not necessarily result in a loss of critical information, as benign traffic often contains redundant and repetitive patterns. Instead, undersampling improves the model's focus on learning discriminative features of malicious activities, leading to better detection performance. By prioritizing the representation of attack classes and minimizing majority class bias, this approach aligns with the study's goal of improving the identification of rare but critical cyber threats while maintaining efficient model training.
- **Data Normalization:** Feature values were scaled between 0 and 1 to improve model convergence.
- **Data Reshaping**: The dataset was reshaped to ensure compatibility with CNN and LSTM models.
- **Train-Test Split**: The dataset, obtained from the CSE-CIC-IDS2018 dataset, was preprocessed to address class imbalance, resulting in a balanced label set (y_balanced). Following preprocessing, a stratified 80/20 train–test split was performed to ensure that both the training and testing subsets preserved the class distribution of the processed dataset. This approach minimizes sampling bias and ensures that all classes are proportionally represented in both subsets, enabling consistent learning and reliable evaluation. The 80% training portion provides sufficient data for the CNN–LSTM model to effectively learn complex spatial and temporal patterns in network traffic, while the 20% test set remains large enough to yield statistically meaningful performance metrics.
- **Label Encoding**: The categorical labels were converted into a numerical format for classification.

When it comes to evaluation, the study uses several essential performance metrics in order to evaluate the performance of the proposed machine learning based model. These are accuracy, precision, recall and F1-score along with false positive rate (FPR). In addition to the previously mentioned metrics, the Receiver Operating Characteristic (ROC) curve and the corresponding Area under the Curve (AUC-ROC) are evaluated. As well as the fact that

the proposed model is put head to head with 4 other neural network models to check the viability and standard of the results.

3.1 Model Configuration and Training Setup

To ensure reproducibility and transparency, the models were implemented and evaluated under a consistent experimental setup.

Hybrid CNN–LSTM Model Architecture

The proposed hybrid model integrates convolutional and sequential learning components to capture both spatial and temporal features of network traffic data.

- Input Layer: Accepts preprocessed network traffic features.

CNN Module:

- 1D Convolutional Layer with: 64 filters, Kernel size of 1, ReLU activation
- MaxPooling Layer with pool size of 1
- Flatten Layer to convert feature maps into a one-dimensional vector

LSTM Module:

- A single LSTM layer with 50 units to capture temporal dependencies in sequential data

Feature Fusion:

- Outputs from CNN and LSTM are concatenated into a combined feature vector

Fully Connected Layers:

- Dense layer with 64 neurons and ReLU activation
- Output layer with Softmax activation for multi-class classification

Training Configuration

- The model was trained using the following parameters:
- Loss Function: Categorical Cross-Entropy
- Optimizer: Adam
- Epochs: 10
- Batch Size: 32
- Validation Split: 20%

Data Preparation

- Data normalization using Min-Max scaling (0–1 range)
- Data reshaping for compatibility with CNN and LSTM
- Train-test split: 80% training / 20% testing
- Label encoding applied to categorical outputs
- Undersampling used to balance classes

System Configuration

- Processor: Intel Core i9-12900H
- RAM: 16 GB
- OS: Windows 11 Pro

4. Discussion

In this study, a classification-based experiment is conducted using five machine learning models: Feedforward Neural Network (FNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and a hybrid CNN–RNN (LSTM) architecture, evaluated on the CSE-CIC-IDS2018 dataset.

Table 2. Models Evaluation

Model	Test Accuracy	Test Loss	Precision	Recall	F1 Score
FNN	0.9698	0.0981	0.9622	0.9698	0.9599
CNN	0.9701	0.0937	0.9626	0.9701	0.9624
RNN	0.8021	0.6664	0.7921	0.8021	0.7707
LSTM	0.9700	0.0924	0.9629	0.9700	0.9637
CNN-RNN	0.9694	0.0936	0.9617	0.9694	0.9596

The performance evaluation of five neural network architectures—FNN, CNN, RNN, LSTM, and a CNN–LSTM hybrid—on a multiclass cyber threat classification task provides important insights into their relative effectiveness. As presented in Table 2, noticeable differences in model performance were observed. The FNN, CNN, LSTM, and CNN–LSTM hybrid models all achieved high test accuracy, ranging narrowly between 96.94% and 97.01%, with the LSTM model attaining the lowest test loss (0.0924). These results suggest that models incorporating convolutional or memory-based components are effective in capturing relevant patterns within the data. In particular, CNN and LSTM models demonstrated slightly better performance, indicating the usefulness of spatial feature extraction and temporal dependency modeling in cyber threat classification. The CNN–LSTM hybrid model, despite integrating both approaches, did not outperform the standalone CNN or LSTM models, which may reflect the impact of increased architectural complexity without a corresponding performance gain.

Further analysis of precision, recall, and F1-score highlights the importance of balanced predictive performance in multiclass cyber threat detection. Among the evaluated models, the LSTM architecture achieved the highest F1-score (0.9637), alongside strong precision (0.9629) and recall (0.9700), indicating consistent classification performance across multiple attack categories. The CNN and FNN models followed closely, each demonstrating precision and recall values above 0.96, reflecting their ability to effectively manage both false positives and false negatives. Such balance is particularly important in cybersecurity

contexts, where false negatives may result in undetected threats, while false positives can lead to unnecessary alerts and increased operational overhead.

The CNN–LSTM hybrid model also demonstrated strong performance, achieving an F1-score of 0.9596. However, its marginally lower performance compared to standalone CNN and LSTM models suggests that combining convolutional and recurrent layers does not inherently lead to improved results. This finding underscores the importance of selecting model architectures that align with the characteristics of the task, rather than increasing complexity without clear benefit. In cyber threat classification, where data can be high-dimensional and time-dependent, models that efficiently extract relevant features while maintaining manageable complexity may offer more practical advantages.

Table 3. False Positive Rate of each class across the different models

Class	Label	FNN	CNN	RNN	LSTM	Hybrid
0	Benign	0.0559	0.0522	0.3230	0.0494	0.0563
1	Bot	0.0000	0.0000	0.0178	0.0000	0.0000
2	DDOS attack-HOIC	0.0000	0.0000	0.0014	0.0000	0.0000
3	DDoS attacks-LOIC-HTTP	0.0002	0.0002	0.0019	0.0003	0.0002
4	DoS attacks-Hulk	0.0000	0.0000	0.0016	0.0000	0.0000
5	FTP-BruteForce	0.0000	0.0000	0.0126	0.0000	0.0000
6	Infiltration	0.0021	0.0037	0.0000	0.0051	0.0022
7	SSH-Bruteforce	0.0000	0.0000	0.0033	0.0000	0.0000

A detailed analysis of the false positive rate (FPR), as presented in Table 3, provides important insights into how each model handles the misclassification of benign activity as malicious. In cybersecurity systems, where false alarms can disrupt operations and reduce analyst efficiency, FPR is a critical evaluation metric. Among the evaluated models, the LSTM architecture demonstrated the lowest FPR across most classes. For the benign class, it achieved an FPR of 4.94%, which is lower than that of the FNN (5.59%) and CNN (5.22%) models, and substantially lower than the RNN model (32.30%).

These results are consistent with earlier observations regarding the strong precision and F1-score of the LSTM model, suggesting that it maintains a favorable balance between detecting malicious activity and limiting false alarms. This characteristic is particularly relevant in practical cybersecurity settings, where excessive false positives can lead to alert fatigue and reduce the effectiveness of incident response processes.

Most of the evaluated models maintained a false positive rate (FPR) of zero or near-zero across several attack-related classes, including Bot, DDoS-HOIC, LOIC-HTTP, DoS-Hulk, and SSH-Bruteforce. This indicates a strong capability to avoid misclassifying benign traffic

as these attack types. In contrast, the RNN model exhibited comparatively weaker performance, generating false positives across multiple classes, including 1.78% for Bot, 1.90% for LOIC-HTTP, and 1.26% for FTP-BruteForce. Although these values may appear relatively small, even modest increases in FPR can have significant implications in high-throughput or real-time cybersecurity environments, where false alarms may accumulate and impact the efficiency of incident response processes. These results suggest that the standard RNN architecture is less effective in maintaining low false positive rates compared to the other evaluated models.

In addition to conventional evaluation metrics, the Area Under the Curve (AUC) provides further insight into each model's ability to distinguish between classes across varying decision thresholds. This is particularly relevant in multiclass cyber threat detection, where different attack types may exhibit overlapping characteristics. The CNN, LSTM, FNN, and CNN–LSTM hybrid models achieved high AUC scores ($\geq$ 0.99) for several classes, including Bot, DDoS-HOIC, LOIC-HTTP, DoS-Hulk, FTP-BruteForce, and SSH-Bruteforce. These results indicate strong discriminative capability between attack and benign traffic for these categories. However, as observed in other evaluation metrics, performance varies across classes, highlighting the importance of comprehensive evaluation when assessing model effectiveness in intrusion detection tasks.

A consistent observation across all models is their performance on the Infiltration class, where false positive rates, although relatively low, are higher compared to most other classes. Specifically, the LSTM model recorded an FPR of 0.51%, the CNN 0.37%, while the FNN and CNN–LSTM hybrid models achieved values near 0.2%. This trend suggests that infiltration traffic presents greater classification difficulty relative to other attack types.

Table 4. AUC scores for each class across the different models

Class	FNN	CNN	RNN	LSTM	Hybrid
Benign	0.99	0.99	0.91	0.99	0.99
Bot	1.00	1.00	0.97	1.00	1.00
DDOS attack-HOIC	1.00	1.00	1.00	1.00	1.00
DDos attacks-LOIC-HTTP	1.00	1.00	0.96	1.00	1.00
DoS attacks-Hulk	1.00	1.00	0.98	1.00	1.00
FTP-BruteForce	1.00	1.00	0.99	1.00	1.00
Infiltration	**0.91**	**0.91**	**0.74**	**0.92**	**0.92**
SSH-Bruteforce	**1.00**	**1.00**	**0.98**	**1.00**	**1.00**

As shown in Table 4, all evaluated models exhibit a decline in performance for the infiltration class, which is the only category where a consistent reduction in AUC is observed relative to other classes. The best-performing model, LSTM, achieved an AUC of 0.92, while the CNN–LSTM hybrid model reached the same value, as indicated in Table 5 and Figure 3. The CNN and FNN models followed closely with AUC values of 0.91. This consistent pattern across different architectures indicates that infiltration remains a comparatively challenging class for all models.

This behavior may be attributed to the subtle nature of infiltration attacks, which can exhibit patterns similar to legitimate user activity, making accurate discrimination more challenging. These findings highlight a broader challenge in cyber threat detection; effectively identifying low-intensity or stealthy attack patterns while maintaining low false positive rates. Addressing this issue remains important for improving the reliability of intrusion detection systems. The findings suggest that infiltration detection remains a complex problem, likely due to the limited feature separability associated with this class.

To address this challenge, future research may explore more advanced approaches such as improved feature engineering, temporal anomaly detection techniques, or graph-based methods to better capture relational patterns within network data. Enhancing detection performance for infiltration attacks would contribute to more reliable intrusion detection systems and improve resilience against one of the more difficult-to-detect forms of cyber threats.

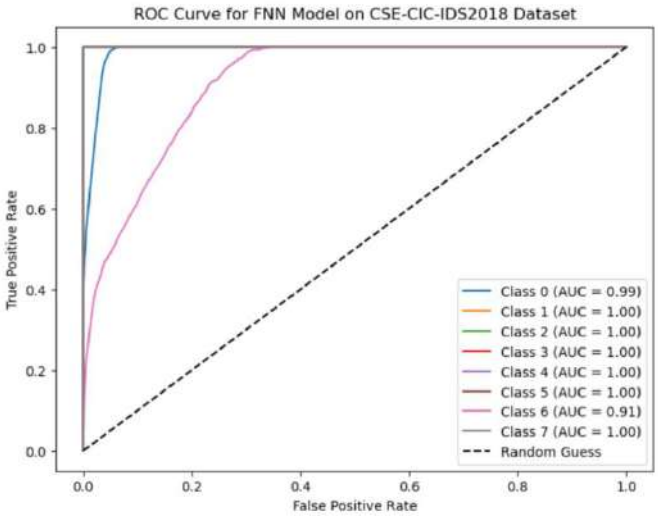


Figure 3. ROC Curve for CNN-LSTM Model



Figure 4. CNN-LSTM Model Performance Confusion Matrix

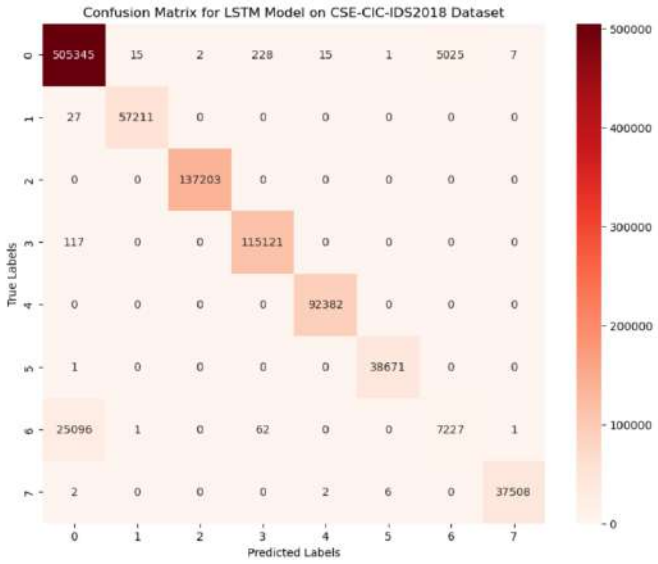


Figure 5. CNN-LSTM Model Performance Confusion Matrix

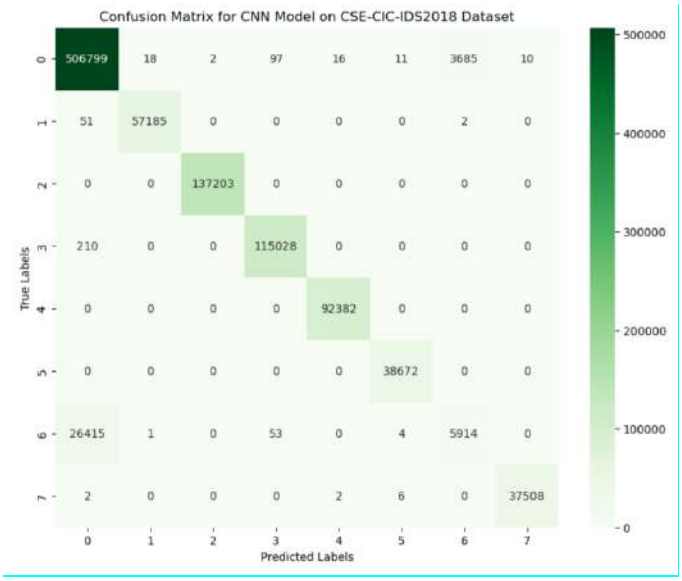


Figure 6. CNN-LSTM Model Performance Confusion Matrix

In the confusion matrices illustrated above in figure 4, figure 5, and figure 6 there is a remarkably high degree of accuracy in identifying dominant classes such as Benign traffic, DDOS attack-HOIC, and DoS attacks-Hulk, evidenced by the densely populated primary diagonals. The models consistently demonstrate an ability to extract spatial and temporal features necessary to distinguish high-volume attack vectors from legitimate network traffic. However, a shared challenge across the architectures is the classification of the Infiltration category (Class 6), which exhibits significant misclassification, often being incorrectly identified as Benign traffic.

From a comparative standpoint, the Hybrid CNN-LSTM model appears to strike a superior balance in feature representation by leveraging the CNN's spatial feature extraction alongside the LSTM's temporal sequence modeling. While the standalone LSTM shows a slightly higher raw count of correctly identified Infiltration instances (7,227) compared to the CNN (5,914) or the Hybrid model (3,962), it also displays higher volatility in misclassifying Benign samples as Infiltration. The Hybrid model optimizes the decision boundaries for the majority of attack classes, resulting in a more robust "clean" diagonal. This analysis suggests that while hybridizing architectures enhances overall classification stability for structured attacks like Brute Force and DDoS, the inherent stealthiness and low-footprint nature of Infiltration attacks remain a significant bottleneck for deep learning-based Intrusion Detection Systems (IDS), necessitating further research into class-balancing techniques or specialized feature engineering.

By employing foundational deep learning architectures, the study achieved high detection accuracy, reaching up to 97.00% using a modest computational setup. The evaluated models

demonstrate the ability to classify multiple types of attacks present in the CSE-CIC-IDS2018 dataset, including DoS, brute-force, and infiltration attacks. The CNN–RNN (LSTM) hybrid model, in particular, combines convolutional and recurrent components to capture both spatial and temporal characteristics of network traffic data, resulting in performance that is comparable to standalone CNN and LSTM models.

Furthermore, the use of standard neural network architectures and widely adopted open-source frameworks supports the reproducibility of the proposed approach. The relatively high performance achieved under these conditions suggests that such models can be effectively implemented without the need for highly specialized computational resources. This highlights the potential applicability of the approach in practical settings, although further evaluation in real-world environments would be necessary to confirm deployment readiness.Despite the wide range of high-performing models reported in the literature with some claiming near-perfect results, our work presents a realistic, transparent, and reproducible approach that achieves competitive performance using foundational deep learning architectures. Some related studies indicated in table 1, such as those by Atefinia and Ahmadi as well as Kanimozhi and Jacob, report 100% accuracy, which, while impressive, raises concerns about potential overfitting, data leakage, or limited generalizability, particularly when model validation details are not clearly disclosed. In contrast, our study employs the publicly available CSE-CIC-IDS2018 dataset, a well-established and diverse benchmark, and follows consistently documented training procedures, including the use of held-out test data and validation splits, to ensure fairness, transparency, and replicability.

5. Conclusion and future work

This study aimed to develop a machine learning–based approach for improving cyber threat identification using the CSE-CIC-IDS2018 dataset. The performance of several deep learning architectures—FNN, CNN, RNN, LSTM, and a hybrid CNN–LSTM model—was evaluated using key metrics, including accuracy, precision, recall, and F1-score. The results indicate that the RNN model performed significantly worse than the other architectures, while the standalone CNN and LSTM models achieved the highest test accuracy (approximately 97.00%). The FNN and CNN–LSTM hybrid models also demonstrated competitive performance, with accuracies of 96.98% and 96.94%, respectively.

The hybrid CNN–LSTM model combines convolutional feature extraction with sequence learning; however, its performance did not exceed that of the standalone CNN and LSTM models. Nevertheless, it achieved strong overall classification results across multiple attack categories, with high true positive rates and high AUC values for most classes. Despite these strong results, certain attack types, particularly infiltration, exhibited comparatively lower performance, indicating that they remain more challenging to detect.

The findings also highlight several practical considerations, including the impact of dataset characteristics, computational requirements, and the need for models to adapt to evolving cyber threats. These factors are important when considering the applicability of such models in real-world cyber security environments.

While this study demonstrates the effectiveness of deep learning approaches for cyber threat classification, further research is needed to improve performance on more challenging attack types and to evaluate models in more realistic settings. Future work may focus on enhanced feature engineering, evaluation on additional datasets, and the integration of approaches that support real-time detection and improved interpretability.

Overall, this research contributes to the development of automated intrusion detection approaches by providing a comparative evaluation of multiple deep learning architectures and highlighting key challenges and considerations in cyber threat classification.

Future Works:

- Extensive hyperparameter tuning and exploration of alternative architectural configurations could further optimize model performance, potentially improving classification accuracy while reducing computational cost.
- Evaluating the proposed model on additional benchmark cybersecurity datasets, such as NSL-KDD and UNSW-NB15, would help assess its generalizability across different network environments.
- Future work may involve deploying the model in real-time network traffic scenarios to evaluate its capability for timely and accurate cyberattack detection.
- Integration of the proposed model into Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS) could be explored to assess its practical applicability in operational cybersecurity environments.

Reference

[1] S. Sur, M. El-Dosuk, S. J. J. o. T. Kamel, and A. I. Technology, "Machine Learning Techniques For Cyber Security," vol. 102, no. 7, 2024.

[2] H. S. Lallie et al., "Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic," vol. 105, p. 102248, 2021.

[3] J. J. E. J. o. T. Bharadiya, "Machine learning in cybersecurity: Techniques and challenges," vol. 7, no. 2, pp. 1-14, 2023.

[4] A. L. Buczak, E. J. I. C. s. Guven, and tutorials, "A survey of data mining and machine learning methods for cyber security intrusion detection," vol. 18, no. 2, pp. 1153-1176, 2015.

[5] I. Ghafir et al., "Detection of advanced persistent threat using machine-learning correlation analysis," vol. 89, pp. 349-359, 2018.

[6] T. Endeshaw, "Jimma Institute of Technology Faculty of Computing and Informatics Intrusion Detection System using Machine Learning for MANET," Jimma University, 2021.

[7] A. Oyetoro, J. Mart, and U. J. S. P. Amah, "Using Machine Learning Techniques Random Forest and Neural Network to Detect Cyber Attacks," 2023.

[8] M. I. Jordan and T. M. J. S. Mitchell, "Machine learning: Trends, perspectives, and prospects," vol. 349, no. 6245, pp. 255-260, 2015.

[9] W. Tounsi, H. J. C. Rais, and security, "A survey on technical threat intelligence in the age of sophisticated cyber attacks," vol. 72, pp. 212-233, 2018.

[10] Y. Xin et al., "Machine learning and deep learning methods for cybersecurity," vol. 6, pp. 35365-35381, 2018.

[11] M. F. Safitra, M. Lubis, and H. J. S. Fakhrurroja, "Counterattacking cyber threats: A framework for the future of cybersecurity," vol. 15, no. 18, p. 13369, 2023.

[12] A. H. Salem, S. M. Azzam, O. Emam, and A. A. J. J. o. B. D. Abohany, "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques," vol. 11, no. 1, p. 105, 2024.

[13] M. Ozkan-Ozay et al., "A Comprehensive Survey: Evaluating the Efficiency of Artificial Intelligence and Machine Learning Techniques on Cyber Security Solutions," 2024.

[14] G. Apruzzese et al., "The role of machine learning in cybersecurity," vol. 4, no. 1, pp. 1-38, 2023.

[15] F. Bouchama, M. J. I. J. o. B. I. Kamal, and B. D. Analytics, "Enhancing cyber threat detection through machine learning-based behavioral modeling of network traffic patterns," vol. 4, no. 9, pp. 1-9, 2021.

[16] J. Kim, J. Kim, H. Kim, M. Shim, and E. J. E. Choi, "CNN-based network intrusion detection against denial-of-service attacks," vol. 9, no. 6, p. 916, 2020.

[17] A. A. Hagar and B. W. J. N. Gawali, "Deep learning for improving attack detection system using CSE-CICIDS2018," vol. 20, no. 6, 2022.

[18] I. H. Sarker, Y. B. Abushark, F. Alsolami, and A. I. J. S. Khan, "Intrudtree: a machine learning based cyber security intrusion detection model," vol. 12, no. 5, p. 754, 2020.

[19] J. L. Leevy and T. M. J. J. o. B. D. Khoshgoftaar, "A survey and analysis of intrusion detection models based on cse-cic-ids2018 big data," vol. 7, pp. 1-19, 2020.

[20] L. L. J. J. o. T. Scientific and A. I. Technology, "HYBRID DEEP LEARNING FRAMEWORK FOR INTRUSION DETECTION: INTEGRATING CNN, LSTM, AND ATTENTION MECHANISMS TO ENHANCE CYBERSECURITY," vol. 103, no. 1, 2025.

[21] Y. Zhang, X. Chen, L. Jin, X. Wang, and D. J. I. A. Guo, "Network intrusion detection: Based on deep hierarchical network and original flow data," vol. 7, pp. 37004-37016, 2019.

[22] J. Kim and H. Kim, "An effective intrusion detection classifier using long short-term memory with gradient descent optimization," in 2017 International Conference on Platform Technology and Service (PlatCon), 2017, pp. 1-6: IEEE.

[23] A. A. Diro and N. J. F. G. C. S. Chilamkurti, "Distributed attack detection scheme using deep learning approach for Internet of Things," vol. 82, pp. 761-768, 2018.