MULTIDISCIPLINARY SCIENCE JOURNAL

RESEARCH ARTICLE
Published Online: September 16, 2024
https://doi.org/10.31893/multiscience.2025129

An improved man-in-the-middle (MITM) attack detections using convolutional neural networks

Mohammed Iddrisu[a] | **Kate Takyi**[a] ✉ | **Rose-Mary Owusuaa Mensah Gyening**[a] | **Kwame Ofosuhene Peasah** | **Linda Amoako Banning**[a] | **Kwabena Owusu Agyemang**[a]

[a]Kwame Nkrumah University of Science & Technology, Department of Computer Science, Ghana.

Abstract The increasing reliance on digital communication networks has made information security a critical concern for individuals, organizations, and governments worldwide. Man-in-the-middle (MITM) attacks are significant, prevalent, and damaging concerning cyber-attacks. Detecting MitM attacks is complex due to their stealthy nature and the sophisticated methods employed by attackers. There is the need for researchers to address this issue using current and novel methods like artificial intelligence. In this paper, an improved MitM attack detection approach using the Convolutional Neural Network (CNN) deep learning algorithm is developed, resulting in an overall detection accuracy of 0.986%. The results confirms that the proposed model is very efficient in comparision to other proposed solutions by other authors.

Keywords: cyber-attacks, deep learning, algorithms, network security

1. Introduction

The increasing reliance on digital communication networks has made information security a critical concern for individuals, organizations, and governments worldwide (Chen et al., 2011). However, this increased connectivity has also led to various cyber threats, with man-in-the-middle (MITM) attacks being damaging forms of cyberattacks (Disha & Waheed, 2022; Zahara et al., 2020). In a MitM attack, an attacker intercepts and alters communications between two parties, often without their knowledge. Detecting MitM attacks is complex due to their stealthy nature and the sophisticated methods employed by attackers. Traditional methods of MitM attack detection usually struggle to accurately identify sophisticated attacks and distinguish them from legitimate network behavior. The existing techniques for detecting MitM attacks rely predominantly on analyzing network traffic patterns and detecting anomalies (Ahmad et al., 2020). However, these methods often face limitations in accurately identifying subtle and complex attack patterns, leading to increased false positives or false negatives.

Additionally, traditional machine learning algorithms employed for MitM attack detection, such as support vector machines (SVMs) and random forests, may struggle to capture the nuanced patterns and dependencies present in network traffic data (Lairab & Azer, 2021). As a result, advanced techniques have been necessary to enhance the accuracy and efficiency of MitM attack detection. In (Hussain et al., 2020), the authors discussed cyber-attacks, focusing on ARP spoofing-based Man-in-the-Middle (MitM) attacks, and proposed the D-ARP. This detection scheme is compatible with the original ARP protocol. D-ARP uses signed ARP packets in parallel to identify and stop ARP spoofing attacks without false positives or negatives. It incorporates features such as DHCP and Nmap to detect attackers' MAC addresses and allows administrators to create a trusted host list. The D-ARP effectively addresses vulnerabilities without altering the original ARP protocol.

Jemili and Bouras (2022) developed a wireless ad hoc network MitM attack detection system and compared two IDS models: one based on principal component analysis (PCA) and the other based on an artificial neural network (ANN). The NS2 tool simulated the network dataset, providing 12 network layer features. We evaluated the IDSs with the SPSS package on 19 samples per group. The PCA-based IDS achieved 76% accuracy and detection rate, while the ANN-based IDS achieved 100%. The study concluded that the ANN algorithm significantly outperforms PCA in detecting MitM attacks. Lan et al. (2020) proposed using artificial neural network (ANN) classification methods for intrusion detection in mobile ad hoc networks (MANETs). This study highlighted the vulnerabilities of MANETs to attacks such as man-in-the-middle (MITM) and black holes due to their wireless and decentralized nature. The authors developed and evaluated the ANN-based intrusion detection system using the NS2 simulation platform, achieving an 88.235% detection rate. The study emphasized time as a crucial factor in detecting and responding to attacks, suggesting this approach as a foundation for future research to enhance MANET security against MITM attacks. In (Majumdar et al., 2021), the authors proposed an energy-efficient framework to prevent MitM attacks in healthcare monitoring systems. The approach uses message authentication codes, more minor signatures, and received signal strength indications for key derivation. The experimental results show a 3% false alarm rate and high detection

 Received: January 29, 2024 | Accepted: August 20, 2024

accuracy, ensuring reliable emergency detection in remote healthcare monitoring. This practical framework enhances data integrity and privacy, improving patient care on the internet of Medical Things (IoMT). In (Ma et al., 2023), a CNN-based anomaly detection model was proposed for IoT security, achieving 98.51% accuracy on the National Infrastructure Database (NID) dataset and 92.85% accuracy on the BoT-IoT dataset. The approach effectively detects intrusions and abnormal traffic patterns, enhancing IoT security and encouraging further research.

The development of an improved MitM attack detection approach using the CNN algorithm is essential for accessing digital communication networks. CNN algorithms can potentially enhance the detection of MitM attacks by effectively analyzing network traffic data and identifying subtle attack patterns that traditional methods may overlook. CNNs are influenced by the architecture and functioning of the visual cortex in the brain, which is known for its ability to detect and identify visual patterns. The CNN architecture has layers of linked neurons that perform operations such as convolution, pooling, and nonlinear activation, leading to the extraction of relevant features from the data fed to it (Sieh & Takada, 2022). Our objectives are threefold:

1. To develop an improved CNN-based model for man-in-the-middle attack detection;
2. To evaluate and compare its performance against traditional MitM attack detection;
3. To analyze the results, draw conclusions, and provide recommendations for further improvements.

These objectives will provide answers to the following research questions:

i. How can the CNN-based model be improved upon and optimized compared to current techniques to increase the precision and effectiveness of man-in-the-middle attack detection?
ii. How does the proposed CNN-based model compare to conventional MitM attack detection techniques in terms of accuracy, false-positive rate, and false-negative rate?
iii. What inferences may be made about the efficacy and applicability of the CNN-based model for man-in-the-middle attack detection based on the evaluation results?

2. Materials and Methods

This study uses CNNs to enhance MitM attack detection by analyzing network traffic patterns. Dataset gathering, preprocessing, CNN architecture creation, and training are performed, followed by validating the approach with real-world data. This method strengthens network security against cyber threats by recognizing complex patterns. Performance evaluation and statistical analysis confirm its success in effective MitM attack detection. Figure 1 illustrates the key ideas and methods we applied in the study.

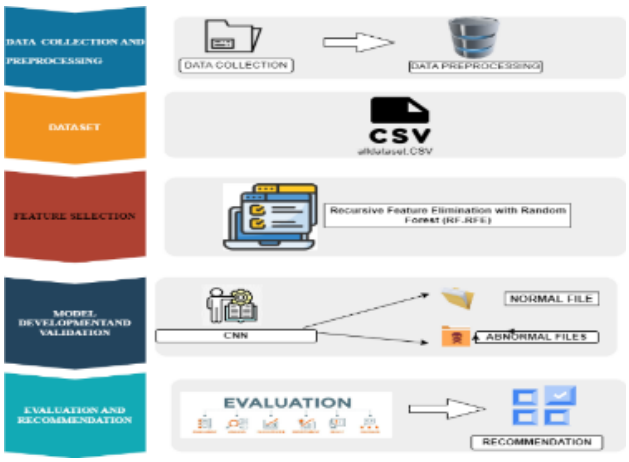

Figure 1 Proposed Framework.

2.1. Data Collection

The dataset used consists of 369,339 samples from nine network attack datasets sourced from an IP-based commercial monitoring system and an IoT device network. The dataset includes both regular activity and Man-in-the-Middle (MitM) attack behavior. Each sample has a ground truth tag indicating the presence of a MitM attack. A total of 353,078 network attack datasets from the monitoring system or IoT network are negative samples, while 16,261 genuine MitM attacks are positive samples. The dataset underwent validation through quality tests such as Kaggle and NDSS to ensure that it accurately represented real-world MitM attacks with an imbalanced distribution of positive and negative samples (Mirsky et al., 2018).

2.2. Dataset Preprocessing

The dataset underwent crucial preprocessing to enhance MitM attack detection with CNNs. Missing data rows were removed, and all columns, except '192.168.20.111', were converted to integers for numerical analysis and CNN compatibility. The dataset is partitioned into 70:10:20 representing training, validation, and testing subsets in percentages, containing approximately 258,537, 36,934, and 73,868 samples, respectively. This preprocessing ensured a clean, well-formatted dataset for successful CNN model research and testing, laying the foundation for improved MitM attack detection using CNNs.

2.3. Feature Selection (RR-RFE)

Feature selection is a critical step when CNNs are used to identify MitM attacks. The dataset is partitioned into features and the target variable. After standardizing and removing missing values from the features, the most significant values are chosen based on their scores. The selected features include source and destination IP addresses, ports, timestamps, packet lengths, and response times. This process reduces dimensionality, improving the CNN model's efficiency in detecting and preventing MitM attacks. We incorporate recursive feature elimination with the random forest (RF-RFE) algorithm to achieve the latter, effectively reducing data redundancy and providing functional feature subsets for evaluation. Figure 2 depicts the RF-RFE procedure in detail.

Input: Training sample: Xg5 = [x3, X53, X53, ---, Xg]

Category: ¥ = [¥1,¥2. Y2 ==» Ym] -

Set of Present Features: s = [1,2, ..., n].

List of sorted features: r = []

Feature Sorting: Train features using the RF classifier a = RF-train (X, y).

Calculate the importance of every feature vector: w.

Repeat until s = []: Find the features with the most negligible importance f = arg min (w).

Update the list of sorted features: r = [s(f),r]-

Eliminate the features with the most negligible importance: s = s(1:—1,f +

1: length(s)). Output: List of sorted features. r

Figure 2 Algorithm for feature selection.

2.4. Deep Learning Model

This research created a CNN-based deep learning model for improved man-in-the-middle attack detection. The model comprises 1D convolution with rectified linear unit (ReLU) activation, max pooling, flattening, softmax output, and a fully linked layer with ReLU activation. We employed the Adam optimizer with categorical cross-entropy loss. Leave-P-out cross-validation (LPOCV) is used to evaluate performance and includes repeatedly partitioning the dataset into training, testing, and validation subsets. This method guarantees consistent and objective validation, improving the CNN model's ability to identify MitM assaults. LPOCV enables a thorough examination of the model's effectiveness and meaningful performance comparisons.

2.5. Performance Metrics

The performance metrics phase uses various measures, such as accuracy, recall, F1 score, precision, ROC curve, and false positive rate, to evaluate the neural network model's performance on test datasets. The model's correctness is determined using the equation TP+TN and assessed using four combinations represented in its outputs by TP, FP, FN, and TN, as in equation 1:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

The calculation rate, or correctness, accurately reflects how many occurrences are detected. The fraction of positive occurrences accurately recognized as positive is represented by the sensitivity, also known as the recall, as shown in Equation 2.

$$\text{Recall} = \frac{TP}{FN+TP} \quad (2)$$

Precision is a statistic that reflects the likelihood of a particular occurrence accurately categorized in Equation 3:

$$\text{Precision} = \frac{TP}{FP+TP} \quad (3)$$

In Equation 4, precision and recall are considered by the F1 score, a statistic that assesses a model's accuracy.

$$\text{F1 score} = \frac{2*Recall*Precision}{Recall*precision} \quad (4)$$

The receiver operating characteristic (ROC) curve shows the algorithm's performance when the true positive rate (TPR) and false positive rate (FPR) are compared at different thresholds of classification, as shown in Figure 3.

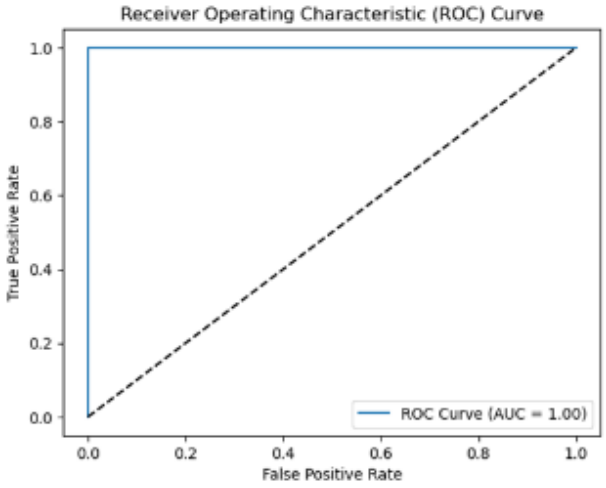

Figure 3 ROC curve showing how feature selection improves MITM attack detection.

3. Results and Discussion

The CNN approach for improving MitM attack detection is evaluated using the following metrics: accuracy, recall, F1-score, and precision. Three hundred sixty-nine thousand three hundred thirty-nine items (369,339) with varied packet-level properties and network patterns make this valuable dataset for in-depth study and the construction of a robust MitM attack detection system. This dataset helps the CNN model recognize unusual patterns and alert possible MitM attacks, improving network security.

3.1. Results of feature selection

For CNN modeling, 33 attributes from 35 columns were selected to identify MITM attacks. The source and destination IP addresses, ports, timestamps, flow characteristics, packet length, and response time data are included. With these attributes, the CNN model detected 100% of MITM attempts. Three hundred sixty-nine thousand three hundred thirty-nine (369,339) records were included. Generalizability requires overfitting and unknown data validation. Feature selection improves MITM attack detection, as shown in Figures 3 and 4 (the confusion matrix and ROC curve).

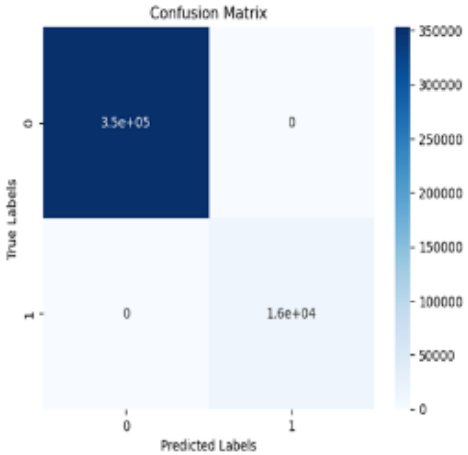

Figure 4 Confusion matrix showing how feature selection improves MITM attack detection.

3.2. Test Results with Variants in Epochs and Batch Sizes

The trained CNN detects MITM assaults with promising results. The model improves across ten epochs, with reduced loss and increasing accuracy. The validation accuracy increased from 98.30% to 98.61%, while the training accuracy increased from 97.87% to 98.71%. Training and validation losses decrease during the epochs, suggesting good learning and convergence.

The overall losses are 0.0438 for training and 0.0479 for validation. While the CNN model has promise, more evaluation and comparison with other methodologies are required for performance validation. Table 1 shows the results for various epochs and batch sizes. Figure 5 shows the training and validation loss of the epoch and batch size. Figure 6 depicts the accuracy of training and validation by epoch and batch size.

Table 1 Training and Validation Accuracies Based on the Number of Epochs and Batch Size.

Number Of Batches	Number Of Epoch	Time Taken (Seconds)	Loss	Accuracy	Vald_Loss	Vald_Accuracy
1	9234	47	0.0736	0.9787	0.0600	0.9830
2	9234	46	0.0552	0.9836	0.0571	0.9829
3	9234	42	0.0514	0.9848	0.0512	0.9859
4	9234	41	0.0490	0.9856	0.0543	0.9850
5	9234	43	0.0482	0.9860	0.0485	0.9862
6	9234	44	0.0463	0.9865	0.0511	0.9853
7	9234	49	0.0456	0.9867	0.0523	0.9859
8	9234	42	0.0466	0.9871	0.0459	0.9858
9	9234	42	0.0445	0.9870	0.0432	0.9873
10	9234	56	0.0438	0.9871	0.0479	0.9861

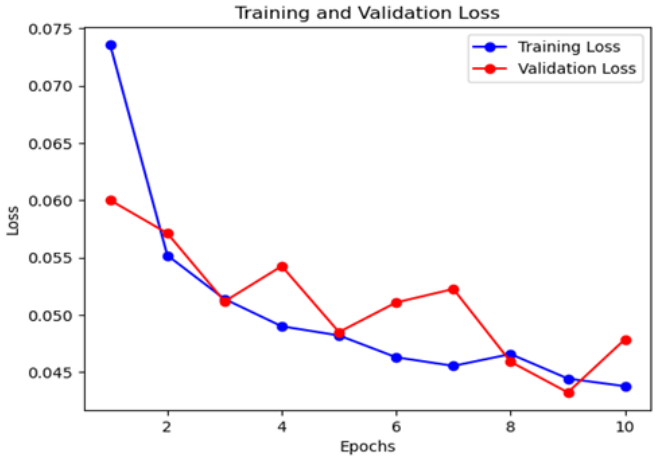

Figure 5 A loss graph of training and validation based on the number of epochs and batch size.

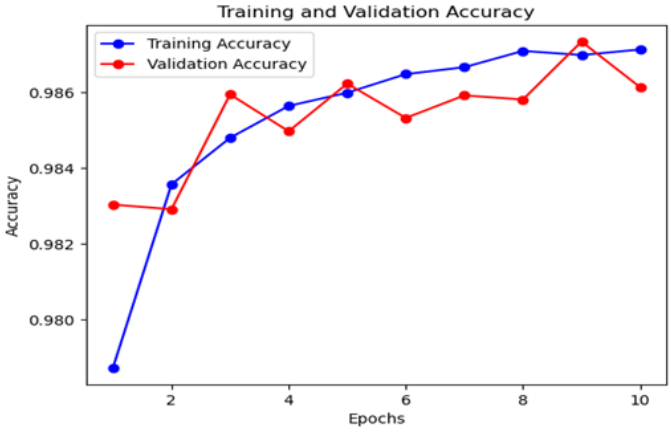

Figure 6 An accuracy graph of training and validation based on the number of epochs and batch size.

Research on improved MITM detection using CNNs has demonstrated outstanding performance and accuracy. The CNN model achieved good accuracy, precision, recall, and F1 scores, showing that it successfully recognized the samples. It is proven that the model can detect genuine Man-in-the-Middle attacks while minimizing false positives. The high recall score demonstrates its capacity to detect a considerable number of actual assaults while reducing false positives. These excellent findings support the effectiveness and promise of the CNN-based approach for detecting and mitigating man-in-the-middle attacks, boosting network security. The findings of the CNN model are shown in Table 2. Figure 7 shows the ROC curve of the CNN model. The confusion matrix of the proposed model is shown in Figure 8. A classification report for the proposed model is shows in Figure 9.

Table 2 Summary of the evaluation metric results from the CNN model.

Model Evaluation	Model Result
Recall Score	0.9861374343423404
Precision Score	
Accuracy Score	0.9861374343423404
F1 Score	
Recall Score	0.9861374343423404

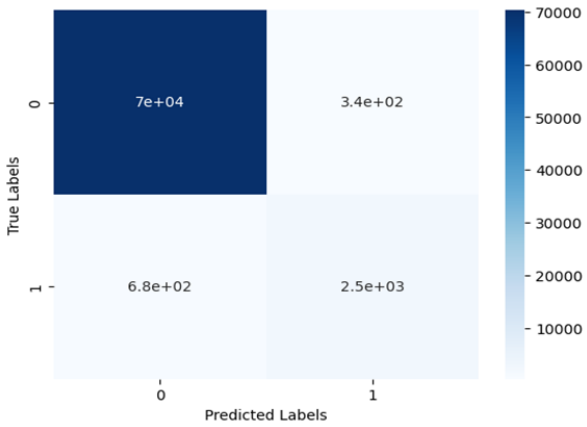

Figure 7 The ROC curve for the proposed model.

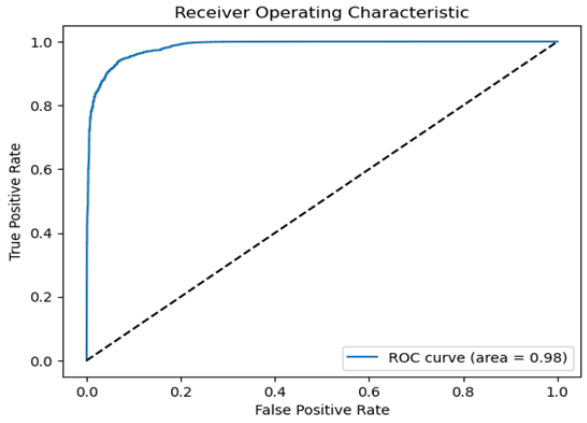

Figure 8 The ROC curve for the proposed model.

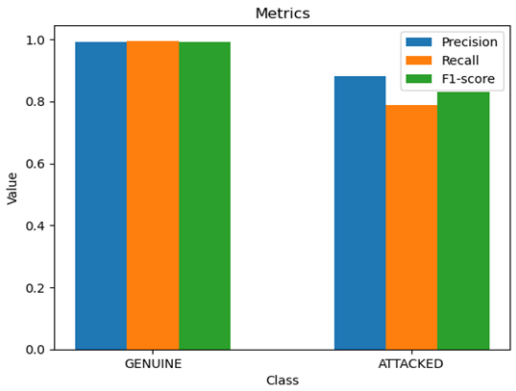

Figure 9 A graph of the classification reports of the proposed model.

3.3. Model Prediction

The model's performance is assessed using 70% training, 10% validation, and 20% testing datasets. The confusion matrix computed results for several labels, with the "GENUINE" class demonstrating good accuracy, recall, and F1 score in CNN testing for enhanced man-in-the-middle attack detection. Despite its relatively strong accuracy (precision), the performance of the "ATTACKED" class needs improvement in terms of recall. The F1 score strikes a compromise between accuracy and recall. The

CNN model reliably recognizes real-world scenarios, although improvements are required to improve the recall score for attacked occurrences. Table 3 gives the categorization results.

Table 3 Model Prediction Results (Generic vs Attacked).

Class	Precision	Recall	FI-Score	Accuracy
Genuine	0.99	0.99	0.99	0.99
Attacked	0.88	0.79	0.83	0.99

3.4. Model Comparison

This study conducted a comparative analysis of various models for detecting man-in-the-middle attacks, employing other methods. The results from the proposed work are compared to those of other works (Saheed & Arowolo, 2021; Saed & Aljuhani, 2022; Sultan et al. 2022) in the literature. These works were selected because they are recent works that align with our work regarding problem statements and metrics for comparison, to the best of our knowledge. In addition, the authors of the work also used traditional machine learning algorithms for MITM detection. The parameters chosen for the comparison were precision, recall, F1 score, and accuracy. Saed & Aljuhani (2022) worked on detecting man-in-the-middle attacks with the random forest algorithm. Sultan et al. (2022) also focused on MITM attacks for IoT devices by exploring various machine learning algorithms and applying the K-nearest neighbor (KNN) algorithm. Saheed & Arowolo (2021) adopted the Particle Swarm Optimization Recurrent Neural Network (PSO-RNN) and other machine learning algorithms to detect cyber attacks, including MITM attacks. In terms of precision, our proposed method and that of Sultan et al. (2022) achieved the best performance, with values of 0.986 and 0.967, respectively. With respect to the recall, accuracy, and F1 score, our proposed model achieved the best results, followed by those of Sultan et al. (2022). The performance of the models can be attributed to the choice of algorithm used by each of the works. KNN has been proven to perform very well when used to construct machine learning models in general (Takyi et al., 2018), while the other methods have some limitations when used to detect MitM attacks (Al-Juboori et al.), justifying why Sultan et al. (2022) performed comparatively well. Table 4 shows a summary of the metric performances of all the models. The proposed model exhibited superior performance compared to other models, achieving precision, recall, F1-scores, and accuracy scores of 0.986. Therefore, we can conclude that the proposed model has the potential to accurately detect MITM attacks in network environments.

Table 4 Model comparison with other works in the literature.

Authors	Methods	Precision	Recall	F1 score	Accuracy
Saheed & Arowolo. (2021)	Particle Swarm Optimization Recurrent Neural Network (PSO-RNN)	0.8563	0.8563	-	0.9508
Saed and Aljuhani (2022)	Random Forest (RF)	0.70	0.866	0.77	0.94
Sultan et al (2022)	using k-Nearest Neighbors (KNN)	0.9674	0.9674	0.9473	0.9586
Proposed Model	CNN	0.986	0.986	0.986	0.986

4. Conclusion

The proposed model for "improved man-in-the-middle attack detection using a CNN-based approach" detected MitM attacks and was evaluated using several metrics. CNN-based MitM attack detection scored 0.9861374343423404, proving its practicality. The technique correctly identified attacked cases (0.88 precision, 0.79 recall, 0.83 F1-score) and detected legitimate communications (0.99). CNN-based network security against MitM attacks shows potential, providing significant information to academics and practitioners. These discoveries may increase the accuracy and resistance of MitM detection.

This research uses a CNN-based technique to improve MitM attack detection. The proposed model outperforms established techniques in terms of F1 score, recall, accuracy, and precision. This demonstrated the possibility of improving overall defenses against MitM assaults. Challenges such as insufficient labeled data, complicated assault patterns, and computing resources, on the other hand, were overcome via dataset curation and CNN architecture design. The CNN-based technique was shown to be reliable in detecting MitM attacks. However, additional improvements may be achieved by expanding the dataset, fine-tuning hyperparameters, and investigating model interpretability methodologies. Despite these limitations, this research advances network security and sets the framework for future MITM attack defense technologies.

4.1. Future work

For further studies and future works, the suggestions for improving the proposed technique include diversifying and extending the dataset for enhanced generalizability, fine-tuning hyperparameters, experimenting with multiple CNN architectures, adopting online training to react to developing attack patterns, and partnering with industry experts among these. These recommendations, when implemented, seek to enhance the model's accuracy and dependability in real-world

circumstances. In addition, we will examine ensemble methods, transfer learning, explainability approaches, real-time deployment, and field assessments to improve the CNN-based MitM attack detection and network security methodology.

Acknowledgment

We express our sincere gratitude to the Department of Computer Science KNUST for the support and resources provided to complete the study.

Ethical considerations

Not applicable.

Conflict of Interest

The authors declare no conflicts of interest.

Funding

This research did not receive any financial support.

References

Ahmad, F., Kurugollu, F., Adnane, A., Hussain, R., & Hussain, F. (2020). MARINE: Man-in-the-Middle Attack Resistant Trust Model in Connected Vehicles. *IEEE Internet of Things Journal, 7*(4)*,* 3310–3322.https://doi.org/10.1109/JIOT.2020.2967568

Al-Juboori, S. A. M., Hazzaa, F., Jabbar, Z. S., Salih, S., & Gheni, H. M. (2023). Man-in-the-middle and denial of service attacks detection using machine learning algorithms*. Bulletin of Electrical Engineering and Informatics*, 12(1), 418-426. DOI: https://doi.org/10.11591/eei.v12i1.4555

Chen, H., Chen, Y., & Summerville, D. H. (2011). A survey on the application of FPGAs for network infrastructure security. In *IEEE Communications Surveys and Tutorials* (p. 541–561). IEEE, 2011. https://doi.org/10.1109/SURV.2011.072210.00075

Disha, R. A., & Waheed, S. (2022). Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. *Cybersecurity, 5*(1),https://doi.org/10.1186/s42400-021-00103-8

Hussain, F., Hussain, R., Hassan, S. A., & Hossain, E. (2020). Machine Learning in IoT Security: Current Solutions and Future Challenges. *IEEE Communications Surveys and Tutorials, 22(3),* 1686–1721. https://doi.org/10.1109/COMST.2020.2986444

Jemili, F., & Bouras, H. (2021). Intrusion detection based on big data fuzzy analytics. In *Open Data*. IntechOpen. DOI: 10.5772/intechopen.99636

Lairab, B. I., & Azer, M. A. (2021). A Distributed Intrusion Detection System for Ad Hoc Networks. In *Proceedings of the 2021 16th International Conference on Computer Engineering and Systems (ICCES)* (p.1-4). IEEE, 2021.https://doi.org/10.1109/ICCES54031.2021.9686166

Lan, H., Zhu, X., Sun, J., & Li, S. (2020). Traffic Data Classification to Detect Man-in-the-Middle Attacks in Industrial Control System. In *Proceedings of 2019 6th International Conference on Dependable Systems and Their Applications (DSA 2019)*, (p. 430–434). https://doi.org/10.1109/DSA.2019.00067

Ma, T., Xu, C., Yang, S., Huang, Y., An, Q., Kuang, X., & Grieco, L. A. (2023). A Mutation-Enabled Proactive Defense Against Service-Oriented Man-in-The-Middle Attack in Kubernetes. *IEEE Transactions on Computers*, 1–14. https://doi.org/10.1109/tc.2023.3238125

Mirsky, Y., Kalbo, N., Elovici, Y., & Shabtai, A. (2018). Vesper: Using echo analysis to detect man-in-the-middle attacks in LANs. *IEEE Transactions on Information Forensics and Security*, *14(6),* 1638-1653.

Saed, M., & Aljuhani, A. (2022, January). Detection of man in the middle attack using machine learning. In *2022 2nd International Conference on Computing and Information Technology (ICCIT)* (p. 388-393). IEEE, 2022.

https://doi.org/10.1109/ICCIT52419.2022.9711555

Saheed, Y. K., & Arowolo, M. O. (2021). Efficient cyber-attack detection on the internet of medical things-smart environment based on deep recurrent neural network and machine learning algorithms. *IEEE Access, 9*, 161546-161554.

https://doi.org/10.1109/ACCESS.2021.3128837

Sieh, W.-B., Leu, J.-S., & Takada, J.-I. (2022). Use chains to block DNS attacks: A trusty blockchain-based domain name system. *Journal of Communications and Networks, 24*(3), 347–356.https://doi.org/10.23919/jcn.2022.000009

Sultan, A. B. M., Mehmood, S., & Zahid, H. (2022, March). Man in the Middle Attack Detection for MQTT based IoT devices using different Machine Learning Algorithms. In *2022 2nd International Conference on Artificial Intelligence (ICAI)* (p. 118-121). IEEE, 2022.https://doi.org/10.1109/ICAI55435.2022.9773590

Takyi, K., Bagga, A., & Goopta, P. (2018, August). Clustering techniques for traffic classification: A comprehensive review. In *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)* (pp. 224-230). IEEE. DOI: 10.1109/ICRITO.2018.8748772

Zahara, L., Musa, P., Wibowo, E. P., Karim, I., & Musa, S. B. (2020). The facial emotion recognition (FER-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (CNN) algorithm based Raspberry Pi. In *2020 Fifth international conference on informatics and computing (ICIC)* (p. 1-9). IEEE. DOI: 10.1109/ICIC50835.2020.9288560