Baghdad Science Journal

Volume 23 | Issue 2 Article 30

2-24-2026

Enhancing Performance of Intrusion Prevention Systems through Machine Learning: A Comparative Study

Rawaa Hamza Ali
Department of Biology, College of Science, University of Misan, Maysan, Iraq, rawaaha@uomisan.edu.iq

Zainab Saad Karam
Department of Physics, College of Science, University of Misan, Maysan, Iraq, zainab-almosawi@uomisan.edu.iq

How to Cite this Article

Ali, Rawaa Hamza and Karam, Zainab Saad (2026) "Enhancing Performance of Intrusion Prevention Systems through Machine Learning: A Comparative Study," *Baghdad Science Journal*: Vol. 23: Iss. 2, Article 30.
DOI: https://doi.org/10.21123/2411-7986.5222

BAGHDAD SCIENCE JOURNAL 2026;23(2):762–771

RESEARCH ARTICLE

Enhancing Performance of Intrusion Prevention Systems through Machine Learning: A Comparative Study

Rawaa Hamza Ali [1,*], **Zainab Saad Karam** [2]

[1] Department of Biology, College of Science, University of Misan, Maysan, Iraq
[2] Department of Physics, College of Science, University of Misan, Maysan, Iraq

ABSTRACT

This research introduces a Machine Learning-Powered Intrusion Prevention System (ML-IPS) as a robust solution to address the challenges posed by evolving cyber threats. The ML-IPS combines timely threat detection with enhanced accuracy for real-time attack prevention, offering a resilient defense against a broad spectrum of cyber-attack. This study delves into the comprehensive evaluation of a machine learning-powered IPS that ingeniously harnesses the power of advanced algorithms to facilitate real-time threat detection and significantly enhance the overall accuracy of the system. The efficacy of intrusion prevention systems (IPS) that employ machine learning (ML) is greatly dependent on the choice of suitable ML algorithms and the evaluation of their precision and inference duration. This research embarks on an in-depth evaluation of ML models for IPS applications, focusing on a comprehensive comparison of their accuracy and inference time metrics. Additionally, the methodology employs a supervised learning approach using a labeled dataset (CICIDS2017) containing both benign and malicious network traffic, providing a realistic and practical approach. The simulation results demonstrate that Decision tree and random forest algorithms can improve the prevention of attack in real-time by achieving 99.88% accuracy and about 10ms for time of detection. The findings demonstrate that the Intrusion Prevention System (IPS) is adept at promptly identifying and reacting to assaults, thereby furnishing a stronger and more durable safeguard against the ever-changing landscape of cyber hazards.

Keywords: Cybersecurity, Decision tree, KNN, Machine learning, Random forest

Introduction

The contemporary era presents the formidable predicament of ensuring cybersecurity. The world is becoming increasingly reliant on new technologies as a result of the rapid advancement of technology, which has improved living standards and allowed civilization to flourish. Cybersecurity has become increasingly important to protect Internet-connected systems from attack and unauthorized access. [1]

It may be rather difficult to prevent, identify, or mitigate security risks and to manage them swiftly and efficiently. The increasing occurrence of novel threats necessitates a continuous updating of threat detection techniques. In the realm of safeguarding against the ever-increasing and complex network assaults, Intrusion Detection Systems (IDSs) and Intrusion Prevention Systems (IPSs) operate as the principal and indispensable mechanisms. Due to the lack of reliable test and validation datasets, anomaly-based intrusion detection approaches are suffering from consistent and accurate performance evolutions. [2] Signature-based IDS detect unknown attacks accurately, while IDS detect anomalies by comparing previously identified behaviors with current profiles. [3] Machine learning technology has embraced

Received 17 February 2024; revised 10 October 2024; accepted 13 October 2024.
Available online 24 February 2026

* Corresponding author.
E-mail addresses: rawaaha@uomisan.edu.iq (R. H. Ali), zainab-almosawi@uomisan.edu.iq (Z. S. Karam).

https://doi.org/10.21123/2411-7986.5222

many specializations in the field of attack detection, cybersecurity, computer vision, natural language processing, speech recognition and others.[4]

In order to identify the best machine learning strategies for traffic categorization based on execution durations and classification performance measurements—a critical factor for future real-time deployments— A variety of machine learning techniques were considered in this study. Because it includes bidirectional internet flows (derived from traffic captures) with both good traffic and various forms of current violation, the CICIDS2017 dataset was used for this study. A collection of connection-related information that may be utilized to model traffic and differentiate between legitimate flows and assaults describe each traffic flow.[5]

This study addresses an important and relevant problem in cybersecurity, namely the detection and prevention of cyber threats in real-time. Additionally, the CICIDS2017 includes the raw packet-based network traffic captures that were gathered during the dataset construction, allowing the traffic flows to be extracted from them. In order to apply the ML powered IPS system correctly, machine learning algorithms were chosen that firstly detect attacks in real-time so that they can directly prevent these attacks, and secondly, the attack detection process must achieve very high accuracy to be a reliable system. To accomplish these aims, machine learning algorithms were chosen, specifically those that yield the utmost precision and minimize the duration of inference. Therefore, to avoid attacks in real-time, examined several machine learning approaches to identify which ones enable the best traffic classification results based on execution durations and classification performance metrics.

Related studies

In the perpetually evolving domain of cybersecurity, the protection of networks against progressively advanced cyber threats requires the creation of resilient and flexible Intrusion Detection and Prevention Systems. Several related studies have contributed significantly to the development and understanding of Machine Learning-Powered Intrusion Prevention Systems (IPS). Here are some notable examples.

In this study.[6] Detection in contemporary business operations, Intrusion Detection Systems (IDS), which serve the pivotal role of identifying and mitigating security threats affecting confidentiality, integrity, and availability, have assumed paramount significance. While several open-source intrusion detection systems are available, they are often characterized by unique rule and signature syntax, which can hinder their interoperability. This study introduces an innovative intrusion detection approach underpinned by deep learning principles, enabling the classification of diverse attack types without the reliance on human-created signature mappings or rules.

After being exposed to unusual traffic situations via an Internet connection, the ADS was taught to identify anomalies more frequently than the first two separate models, which decreased the overall number of false positives and negatives produced. A high performance intrusion detection approach has been presented by Byoungkoo K, et al.[7] To identify and counteract variant assaults on high-speed lines, the researchers.[8] suggest a sequence detection method based on proposing a hybrid method of supervised machine learning and an unsupervised learning method to build a suitable model. It works by selecting features using the importance decision tree-based method while removing redundant features and detecting outlier/outlier data using the local outlier factor (LOF) method.

According to this study,[9] cybersecurity is becoming a top priority as more and more devices—including smartphones and tablets—connect to 5G networks, and conventional protections are failing miserably. Since these technologies are developing, there has been a lot of interest in combining SDN/NFV-based SFC (Service Functions Chaining) techniques with ML (Machine Learning) to improve security in MEC systems (Multi-access/Mobile Edge Computing). In this study the researchers first develop an elastic architecture to integrate ML with a virtualized SFC in order to provide multiple services at MEC in an intelligent and efficient manner. Afterwards, they proposed an anomaly detection technique based on machine learning (ML) that would serve as a kind of service strategy for SFC classifiers, steering the latter toward quick traffic categorization and consequent redirections of attack flows. Lastly, has been build a corresponding prototype system and conduct extensive tests to evaluate the effectiveness of the proposed approach. Related results demonstrate the feasibility and advantages of the proposed framework and algorithm.

In the present investigation,[10] the advent of a novel technological advancement known as Software-Defined Networks (SDN) possesses the capability to completely transform the manner in which network architecture is generated, upheld, and employed. Old proprietary network architecture is often replaced by open, configurable network architecture. This new, improved technology also adds a new security burden to the network design because of the existing and growing security vulnerabilities. Since the central controller is now the only point of failure, hackers

may more easily attack the network and cause it to become more vulnerable. Integration of the intrusion detection system (IDS) with the SDN architecture is essential to provide a defense against network assaults. A method for intrusion detection based on deep learning has been proposed by Mohammadpour et al.[11] The researchers utilized CNN with the NSL-KDD dataset in the recommended study. Every experiment makes advantage of the proper division of the training and testing datasets. CNN NIDS is utilized for implementation, and two class classifications are employed. Python and the Keras package are used to create the suggested model. The recommended system may identify anomaly-based infiltration more effectively since the authors have included layers like convolutional, fully connected, and pooling in it.

Sharma et al. have presented a methodology that makes use of machine learning (ML) to detect invasive web-based assaults efficiently.[12] The researchers have made an effort to identify and address the primary causes of false positives and false negatives. Every experiment is conducted using the proper training and testing dataset separation. Additionally, CSIC 2010 HTTP was the dataset used in this study. The suggested model is evaluated by using important metrics such as recall, accuracy, and precision. To test the proposed model, ML classifiers such as J48, one rule, and Naïve Bayes are used. After comparing all of the applied strategies, it is determined that the J48 classifier has the best intrusion detection accuracy which is 94.5%. Azam Z et al.[13] studied network intrusion detection mechanisms using machine learning (ML) and deep learning (DL) techniques adopted in IDS systems, and trends and developments in machine learning and machine learning-based network intrusion detection systems (NIDS), through evaluation and selection of the data set. has been using the decision tree as a model to detect anomalies in the results by combining the results of a comparative survey.

Al Lail et al. have devised various security countermeasure strategies referred to as Network Intrusion Detection Systems (NIDS).[14] However, despite these methods, attackers have devised novel tactics to illicitly access resources. This paper suggests the utilization of machine learning (ML) to construct a NIDS system that can identify contemporary attack patterns with a notably high detection accuracy. The study involves the implementation and assessment of multiple ML algorithms, comparing their efficacy on a cutting-edge dataset containing modern attack patterns. The findings reveal that the random forest model surpasses other models, boasting a 97 percent detection rate for modern network attacks. This research demonstrates the feasibility of precise prediction and the attainment of a high detection rate for attacks. These outcomes highlight the potential of ML in developing highly efficient NIDS systems.

These related studies collectively contribute to the understanding and advancement of Machine Learning-Powered IPS, offering insights into various techniques, challenges, and opportunities in the field of intrusion prevention. Scholars have the opportunity to exploit these investigations in order to enrich their own research endeavors and make valuable contributions to the continuous development of cybersecurity tactics.

Materials and methods

The research methodology includes a comprehensive evaluation of machine learning models that can be used with a real-time attack detection system as shown in Fig. 1. A corpus comprising actual network traffic, incorporating both harmless and harmful traffic patterns, was employed to instruct and assess multiple machine learning algorithms as well as neural networks. For real-time Threat Detection. The utilization of a supervised learning approach is employed by the application in order to classify network

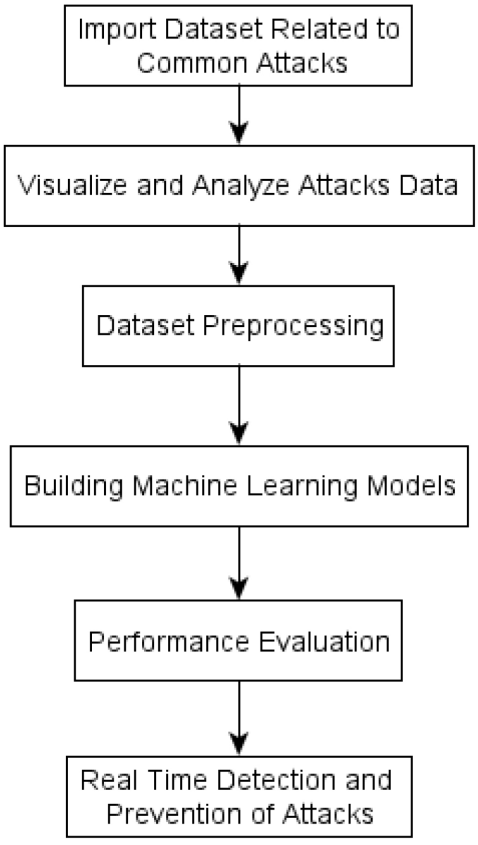

Fig. 1. Proposed methodology.

traffic into two distinct categories, namely normal or malicious. By employing supervised learning algorithms that have been trained using labeled datasets comprising both normal and malicious traffic, the traffic is classified based on predetermined features. These features encompass packet size, destination port, and protocol type. For Enhancing the Accuracy, ML-IPS strives to minimize false positives while maintaining a high detection rate in real-time.

Dataset

The CICIDS2017 dataset is a very convenient dataset for researching Machine Learning-Powered Intrusion Prevention System (IPS). The dataset is vast and varied, comprising of more than 5 million network traffic records categorized into 14 distinct attack types. The organization and cleanliness of the data facilitate the seamless preparation of machine learning models for training purposes. Here are some of the reasons why the CICIDS2017 dataset is a good choice for IPS research:

- Large size: The dataset contains over 5 million network traffic records, which provides a large amount of data for training machine learning models.
- Diversity: The dataset includes a variety of attack types, including both common and rare attacks. The robustness and wide-ranging threat detection capabilities of the models you develop can be guaranteed through this approach.
- The removal of noise and errors through data cleaning and preprocessing facilitates the training and evaluation of machine learning models.
- The balanced representation of the dataset ensures that there is an approximately equal number of records for each attack type, preventing any bias towards a particular attack type in the models you develop.
- In addition, the CICIDS2017 dataset, which is accessible through Kaggle, [15] offers a readily available source for researchers commencing their study on Intrusion Prevention Systems (IPS), thus providing convenience to the research community.
- Attack type: The implemented attacks include Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet and DDoS.

The dataset is divided into two sections: the first section contains 80% of the dataset needed to train the model, and the second section contains 20% of the dataset used to test the model.

Data visualization and preprocessing

Because the simulated data are raw data, preprocessing is necessary to make them suitable for use with machine learning models. Pre-processing involves handling missing values and doing a single hot encoding. The technique employed in addressing missing values is to substitute them with zero, owing to the specificity of the features to the protocol. Prior to constructing analytical models with the dataset, conducting data exploration proves valuable in comprehending and establishing the interrelationship among variables. This examination can be conducted through the amalgamation of multiple attributes or the individual analysis of each attribute. Moreover, techniques such as oversampling, feature scaling, and feature selection are applied to improve model accuracy. Normalization, outlier detection, and encoding categorical variables optimize the dataset for machine learning algorithms.

Building machine learning Ips models

A diverse array of machine learning (ML) models is chosen by drawing from a plethora of influential prior investigations that have furnished valuable perspectives on the advancement of this domain. These studies involve the utilization of Random Forest, Decision Trees, k nearest neighbor (KNN), and Artificial Neural Networks in the shape of Multi-Layer Perceptron (MLP).

Decision tree

A decision tree model [16] is founded on a tree structure that assigns an input feature to each internal node. The root serves as the starting point, where subsets are created from the source set based on specific splitting rules. The internal nodes may have child nodes that are either internal nodes themselves or encompass all potential values of the target class It should be acknowledged that the most basic version of this framework takes the form of a binary classifier. Within the domain of Intrusion Prevention Systems, decision trees are extensively utilized owing to their capacity for interpretation, simplicity of execution, and ability to process data with a high number of dimensions. One of the primary benefits is the model's ease of interpretation; unlike other models, it makes evident why each choice was made. In addition, this particular model exhibits strong performance capabilities when dealing with large datasets and does not necessitate any specific data preprocessing procedures. It is unfortunate, however, that there exist certain limitations.

Specifically, decision trees are not particularly resilient to variations within the training dataset. Moreover, the ideal decision tree derivation procedure is NP-complete. Therefore, rather than looking for the ideal attribute order, a "greedy" method is used to handle this difficulty, where the attribute order for conducting splits at each stage is chosen based on the attribute that produces the best impurity index. Ultimately, overfitting may be readily produced by the answer provided by a decision tree, necessitating additional processes like pruning approaches to combat it. Both "pre-pruning" and "post-pruning" are methods for performing pruning. The main objective is to eliminate nodes that result in a highly accurate division between classes. The task can be accomplished either during the period of instruction or subsequent to the formation of the hierarchy.

Decision trees possess the ability to proficiently acquire knowledge regarding non-linear connections among characteristics in network activity and the designations of attacks. Consequently, they can identify anomalous patterns and categorize network traffic as either harmless or malicious. The interpretability, ease of implementation, and ability to handle high-dimensional data make decision trees an appealing choice for constructing robust and adaptable cybersecurity solutions.

Random forest

A popular example of a model suite is Random Forest,[17] Random forest is an algorithm built according to decision tree technology, which uses several models to improve performance. The features in the original dataset are randomly selected in each sample to build the decision tree model. The most common mode or class predicted by each individual decision tree is the output of random forests, which consist of a number of decision trees.[18] It is based on an ensemble learning approach in which multiple decision tree classifiers $h(x \mid \theta_1), \ldots, h(x \mid \theta_k)$are generated using randomly initialized parameter vectors θ_k. According to recent studies, the Random Forest algorithm integrates bootstrap sampling with random feature selection to construct a diverse set of decision trees. In bootstrap sampling, multiple training subsets are created by sampling the original dataset with replacement, while random feature selection limits the number of features considered at each split, typically to a subset of the total feature space. Each decision tree is trained independently, and the final prediction is obtained through majority voting among all trees. This ensemble strategy improves classification accuracy, enhances generalization capability, and reduces overfitting compared to a single decision tree. However, these improvements are achieved at the expense of higher computational complexity, increased training time, and reduced model interpretability.[19]

K nearest neighbor (KNN)

k-nearest algorithm can be used to classify cases based on the values of variables. The closest data points to the target point k that are looking for provide useful information, which is the basis of the k-NN technique. This non-parametric classification process can be used to estimate the probability density function or direct posterior probability that a given instance belongs to a given class, using the data set generated by the set of classified instances.[20] By using the KNN (K-Nearest Neighbors) algorithm on historical data, it becomes possible to classify actions as customary or deviant. Moreover, this technology can be easily integrated into an intrusion detection system.

Multi-layer perceptron (MLP)

MLP are artificial neural networks and have a supervised learning technique that can be used to learn complex patterns and analyze data interactions is a multi-layer neural network in which information flows unidirectionally from one layer the input to the output layer, passing through the hidden layers. It is can be used to learn complex patterns and analyze data interactions. It consists of several layers and can solve linearly non-separable problems. Three levels are considered within the network: the input layer, which provides the attribute values; A layer of hidden nodes, which is connected to all input nodes; and an output layer, which derives instance classification values based on classes. MLP Classifier can be trained on data to improve attack detection and classify data based on behavior patterns. MLP has a great function in non-linear problems.[21,22]

Performance metrics

The evaluation metrics included detection rate, false positive rate, and scalability. Detection rate measures the proportion of actual attacks correctly identified by the IPS. False positive rate measures the proportion of normal traffic mistakenly flagged as malicious. Scalability assesses ML-IPS's ability to handle increasing network traffic volumes and complexity.

1. True Positive (TP): The quantity of real attacks that a security system accurately identifies as attacks.

2. True Negative (TN): The quantity of typical (non-attack) occurrences that a security system accurately classifies as such.
3. False Positive (FP): The quantity of typical occurrences that a security system mistakenly classifies as assaults; often known as a "false alarm."
4. The number of real assaults that escape detection and are categorized as typical occurrences by a security system is known as False Negative (FN) data.
5. Accuracy: The proportion of successfully recognized occurrences (attacks and normal) to all instances. It evaluates how well the security system performs overall in spotting both kinds of incidents see Eq. (1)

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

6. Precision is the proportion of successfully recognized attacks to all occurrences that a security system has recognized as assaults. It evaluates the system's accuracy in classifying an incident as an attack Eq. (2)

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

7. Recall, also known as Sensitivity or True Positive Rate, is the proportion of successfully detected assaults to all of the real attacks. It assesses the security system's capacity to recognize every assault event see Eq. (3).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

8. F1 Score: Recall and accuracy by harmonic techniques. It works especially well in attack scenarios when it's critical to strike a compromise between precisely identifying assaults (precision) and catching every attack that occurs (recall) see Eq. (4).

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

In the context of cybersecurity, these metrics provide data on the effectiveness of a security system's attack detection and prevention. They support the assessment of the system's accuracy, ability to discriminate between assaults and normal events, and memory and precision balance while reacting to potential threats.

9. The term "training time" refers to the amount of time that a machine learning model uses to learn from training data and adjust its internal parameters. In order to maximize its prediction power, the model looks at the input characteristics and matching target labels at this stage. The steps of feature extraction, model parameter modification, and convergence to an ideal state are all included in the training period.
10. The time that a trained machine learning model spends making predictions or classifications on a separate set of test data is called the testing time. Using the model on unknown data points, this step examines the model's effectiveness and ability to extrapolate. The model makes predictions during testing by drawing on the relationships and patterns it has learnt.
11. The amount of time that a machine learning model takes to forecast or draw conclusions based on fresh, unobserved data sets is known as the "inference time." It is the amount of time that passes between feeding a set of characteristics (data) into a trained model and getting the expected result or forecast.

Results and discussion

This section includes an interpretation of the various experimental outcomes along with a brief discussion of them. The machine learning algorithms RF, DT, MLP and KNN based on the CIC-IDS2017 dataset are evaluated using the parameters shown in the Table 1.

Table 1. Parameters of Ml algorithms.

Algorithm	Parameters
Random Forest	Number of Estimators = 20
KNN	Number of Neighbors = 2
MLP (max_iter = 100)	hidden_layer_sizes = (2, 4)
	max_iter = 20
	activation = relu, solver = adam

The results of evaluation ML-powered IPS that leverages advanced algorithms for real-time threat detection, has been notice that decision tree and random forest have highest detection accuracy at 99.88%, the KNN came in second at 99.45%, MLP classifier due to its incorrect classification has the lowest accuracy at 64.73%. Like that, random forest and decision tree display the highest sensitivity which is 99.92%. the KNN came in second at 99.31%, while MLP has the lowest sensitivity of 94.96%. Table 2 summarizes the facts above.

In terms of the inference time which is important metric for detecting the attack in real-time so should not exceed the value of 100 ms. Decision tree and random forest algorithms have optimum inference time

Table 2. Performance evaluation.

Model	Accuracy	Precision	Recall	F1 Score	Training Time	Testing Time	Inference Time
Random Forest	99.88	99.93	99.92	99.93	380.99	2.56	0.015
Decision Tree	99.88	99.94	99.92	99.93	274.90	1.453	0.01
KNN	99.45	99.85	99.31	99.58	1.97	23311.62	0.824
MLP	64.73	65.31	94.96	77.39	956.59	1.90	0.1041

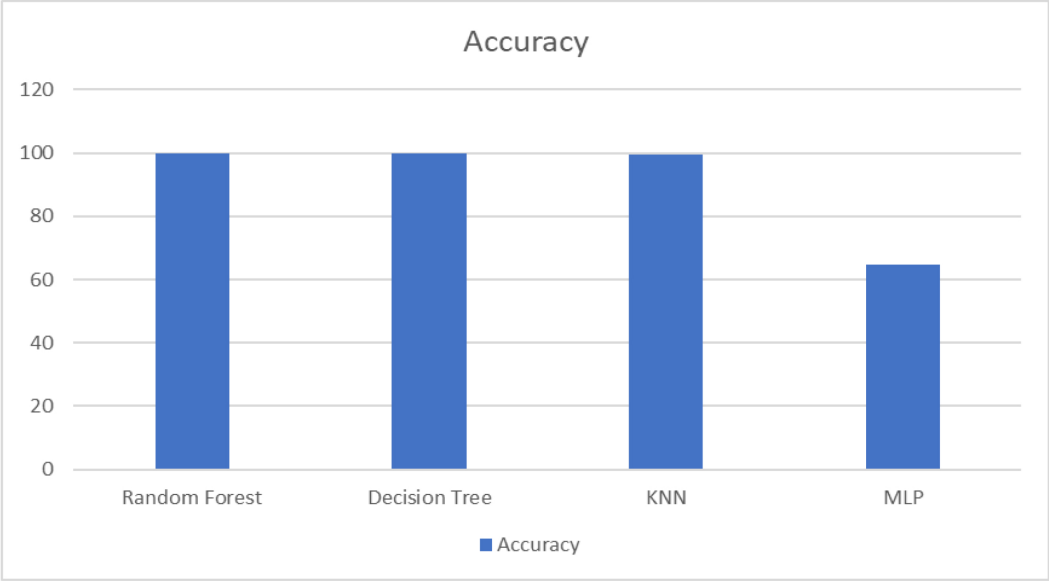

Fig. 2. Accuracy.

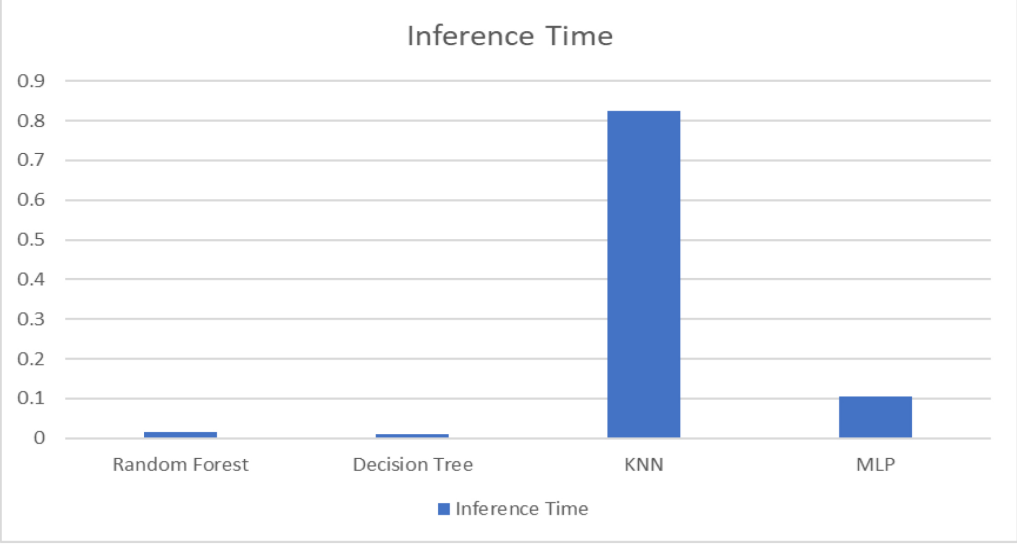

Fig. 3. Inference time.

which is around 10ms for decision tree and 15 ms for random forest this gives the system enough time to prevent these attacks. Conversely, this time for the MLP classifier is 100 ms which is almost accepted but it is about 800 ms for the KNN model which is not acceptable time for detecting the attack in the real-time in order to block it directly by the system.

According to Figs. 2 and 3, the KNN algorithm has a high attack detection accuracy of about 94%, but the attack detection time is large (800 ms) which is not suitable for detecting and preventing attack in real-time, so this algorithm is excluded from the using in this system. Also, the MLP classifier has an acceptable attack detection time, but the detection accuracy

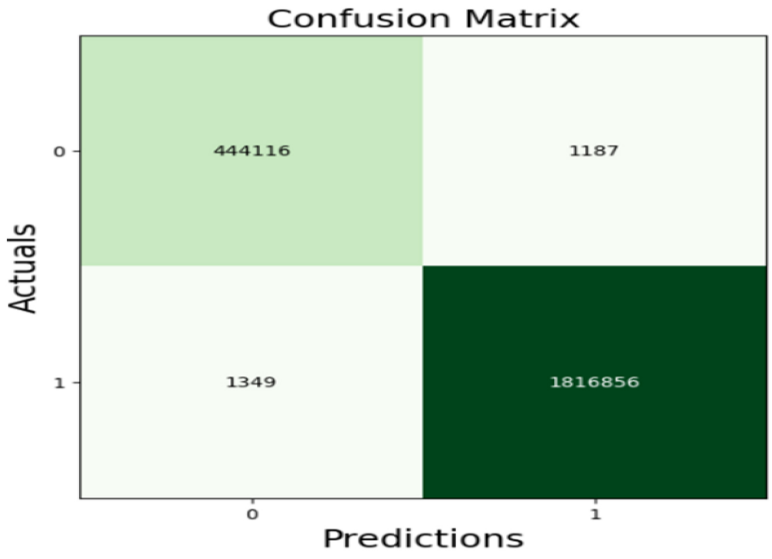

Fig. 4. Confusion matrix for random forest algorithm.

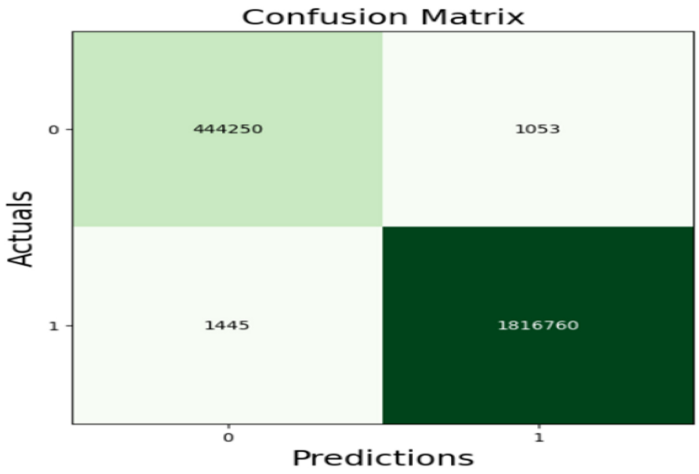

Fig. 5. Confusion matrix for decision tree algorithm.

is reaching 64%, which is not suitable as a value for accuracy in the detection process. Therefore, this classifier is also excluded. Conversely, decision tree and random forest are the best choice for preventing intrusion system in real-time, these algorithms have high accuracy (99.88%) and adequate time (10 ms) for detecting the attack in real-time. In addition, these algorithms can achieve superior detection rates, reduce false positive rates as shown in Figs. 4 and 5.

Conclusion

The proposed ML-powered IPS offers a promising solution to the challenges posed by evolving cyber threats. By combining time of threat detection and enhanced accuracy for real-time preventing of attacks. ML-IPS provides a robust and resilient defense against a wide range of attacks. The evaluation results demonstrate that ML-IPS based on decision tree and random forest hold immense potential in safeguarding networks from threats and ensuring the real-time security and integrity of critical infrastructure. The simulation findings revealed that within this system, the decision tree and random forest algorithms exhibit exemplary performance. This is evident through the high detection accuracy rate of 99.88% and the efficient detection time of approximately 10 milliseconds. These parameters are deemed suitable for the ML-IPS system, thus affirming that the decision tree and random forest algorithms are indeed the optimal selection for a real-time attack detection and prevention system. The future aspects of the study may include studying the impact of different hyperparameter settings or tuning strategies on the performance of the evaluated machine learning algorithms, as hyperparameter tuning is often essential for achieving optimal performance.

Also, exploring the integration of deep learning techniques and dynamic adaptation mechanisms to enhance ML-IPS effectiveness, alongside real-world deployment and evaluation to validate its scalability and performance in diverse network environments.

Authors' declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images that are not ours have been included with the necessary permission for republication, which is attached to the manuscript.
- No animal studies are present in the manuscript.
- No human studies are present in the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee at University of Misan.

Author's contribution statement

R.A. and Z.S. They were involved in conceptualizing the paper, obtaining data, and interpreting the results. They also participated in the drafting of the manuscript and the review process.

References

1. Mijwil M, Aljanabi M, Ali AHJMjoc. Chatgpt: Exploring the role of cybersecurity in the protection of medical information. MJCS. 2023;2023:18–21. https://doi.org/10.58496/MJCS/2023/004.
2. Kumar S, Gupta S, Arora SJIA. Research trends in network-based intrusion detection systems: A review. IEEE Access. 2021;9:157761–157779. https://doi.org/10.1109/ACCESS.2021.3129775.
3. Mahmood RAR, Abdi A, Hussin M. Performance evaluation of intrusion detection system using selected features and machine learning classifiers. Baghdad Sci J. 2021;18(2):884–898. http://dx.doi.org/10.21123/bsj.2021.18.2(Suppl.).0884.
4. Ali AH, Yaseen MG, Aljanabi M, Abed. Transfer learning: A new promising techniques. MJBD. 2023;2023:29–30. https://doi.org/10.58496/MJBD/2023/004.
5. Kurniabudi SD, Damawijoyo IMYB, Bamhdi AM, Budiarto R. CICIDS-2017 dataset feature analysis with information gain for anomaly detection. IEEE Access. 2020;8:132911–132921. https://doi.org/10.1109/ACCESS.2020.3009843.
6. Chockwanich N, Visoottiviseth V. Intrusion detection by deep learning with tensorflow. 2019 21st international conference on advanced communication technology (ICACT). 2019: IEEE. https://doi.org/10.23919/ICACT.2019.8701969.
7. Kim B, Yoon S, Oh J. ATPS–Adaptive threat prevention system for high-performance intrusion detection and response. Managing next generation networks and services: 10th asia-pacific network operations and management symposium, APNOMS 2007, Sapporo, Japan, October 10–12, 2007 Proceedings 10;2007. Springer. https://doi.org/10.1007/978-3-540-75476-3_35.
8. Megantara AA, Ahmad T. A hybrid machine learning method for increasing the performance of network intrusion detection systems. J Big Data. 2021;8(142):1–19. https://doi.org/10.1186/s40537-021-00531-w.
9. Feng B, Zhou H, Li G, Zhang Y, Sood K, Yu S. Enabling machine learning with service function chaining for security enhancement at 5G edges. IEEE Network. 2021;35(5):196–201. https://doi.org/10.1109/MNET.100.2000338.
10. Abubakar A, Pranggono B. Machine learning based intrusion detection system for software defined networks. 7th International Conference on Emerging Security Technologies, (EST 2017), Canterbury, 6–8 September. (EST,2017). IEEE. http://dx.doi.org/10.1109/EST.2017.8090413.
11. Mohammadpour L, Ling TC, Liew CS, Chong CY. A convolutional neural network for network intrusion detection system. APAN. 2018;46(0):50–55.
12. Sharma S, Zavarsky P, Butakov S. Machine learning based intrusion detection system for web-based attacks. IEEE 6th int conference on big data security on cloud (BigDataSecurity), IEEE intl conference on high performance and smart computing, (HPSC) and IEEE intl conference on intelligent data and security (IDS). 2020. IEEE. http://dx.doi.org/10.1109/BigDataSecurity-HPSC-IDS49724.2020.00048.
13. Azam Z, Islam MM, Huda MN. Comparative analysis of intrusion detection systems and machine learning based model analysis through decision tree. IEEE Access. 2023;11:80348–80391. http://dx.doi.org/10.1109/ACCESS.2023.3296444.
14. Al Lail M, Garcia A, Olivo S. Machine learning for network intrusion detection—A comparative study. Future Internet. 2023;15(7):243. https://doi.org/10.3390/fi15070243.
15. Ferraga M, Maglaras L, Moschoyiannis S, Janicke H. Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. J Inf Secur Appl. 2020;50;102419. https://doi.org/10.1016/j.jisa.2019.102419.
16. Rokach L, Maimon O. Data mining and knowledge discovery handbook. 2nd ed. New York: Springer. 2010. https://doi.org/10.1007/978-0-387-09823-4.
17. Mantas CJ, Castellano JG, Moral-García S, Abellán J. A comparison of random forest based algorithms: random credal random forest versus oblique random forest. Soft Comput. 2019;23:10739–10754. https://doi.org/10.1007/s00500-018-3628-5
18. Wotaifi TA, Dhannoon BN. Improving prediction of arabic fake news using fuzzy logic and modified random forest model. Karbala Int J Mod Sci. 2022;8(3):477–485. https://doi.org/10.33640/2405-609X.3241.
19. Salman HA, Kalakech A, Steiti A. Random forest algorithm overview. Babylonian Journal of Machine Learning, BJML. 2024;2024:69–79. https://doi.org/10.58496/BJML/2024/007.
20. Al-Shehari T, Alsowail R. An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques. Entropy. 2021;23(10):1258. https://doi.org/10.3390/e23101258.
21. Nosratabadi S, Ardabili S, Lakner Z, Mako C, Mosavi A. Prediction of food production using machine learning algorithms of multilayer perceptron and ANFIS. Agriculture. 2021;11(5):408. https://doi.org/10.3390/agriculture11050408.
22. Taud H. Mas J. Multilayer perceptron (MLP). Geomatic approaches for modeling land change scenarios. Springer. 2018;27:451–455. https://doi.org/10.1007/978-3-319-60801-3_27.

تعزيز أداء أنظمة منع التسلل من خلال التعلم الآلي: دراسة مقارنة

رواء حمزة علي[1]، زينب سعد كرم[2]

[1] قسم علوم الحياة، كلية العلوم، جامعة ميسان، ميسان، العراق.
[2] قسم الفيزياء، كلية العلوم، جامعة ميسان، ميسان، العراق.

الخلاصة

يقدم هذا البحث نظام منع التطفل المدعوم بالتعلم الآلي (ML-IPS) كحل قوي لمواجهة التحديات التي تفرضها التهديدات السيبرانية المتطورة. يجمع نظام ML-IPS بين الكشف عن التهديدات في الوقت المناسب والدقة المحسنة لمنع الهجمات في الوقت الفعلي، مما يوفر دفاعًا مرنًا ضد مجموعة واسعة من الهجمات السيبرانية. تتعمق هذه الدراسة تحديدًا في التقييم الشامل لنظام IPS الذي يعمل بالتعلم الآلي والذي يستغل ببراعة قوة الخوارزميات المتقدمة لتسهيل اكتشاف التهديدات في الوقت الفعلي وتعزيز الدقة الإجمالية للنظام بشكل كبير. تعتمد فعالية أنظمة منع التطفل (IPS) التي تستخدم التعلم الآلي (ML) بشكل كبير على اختيار خوارزميات تعلم الآلة المناسبة وتقييم دقتها ومدة الاستدلال. يبدأ هذا البحث في إجراء تقييم متعمق لنماذج تعلم الآلة لتطبيقات IPS، مع التركيز على مقارنة شاملة لدقتها ومقاييس وقت الاستدلال. بالإضافة إلى ذلك، تستخدم المنهجية نهج التعلم الخاضع للإشراف باستخدام مجموعة بيانات مصنفة (CICIDS2017) تحتوي على حركة مرور الشبكة الحميدة والخبيثة، مما يوفر نهجًا واقعيًا وعمليًا. توضح نتائج المحاكاة أن شجرة القرار وخوارزميات الغابة العشوائية يمكنها تحسين منع الهجوم في الوقت الفعلي من خلال تحقيق دقة بنسبة 99.88% وحوالي 10 مللي ثانية لوقت الكشف. بالإضافة إلى ذلك، يتم تقييم قدرة IPS على مجموعة متنوعة من الهجمات من خلال استخدام مجموعة بيانات CICIDS2017 التي تتضمن 14 نوعًا من الهجمات. وتوضح النتائج أن نظام منع التطفل (IPS) بارع في تحديد الهجمات والرد عليها بسرعة، وبالتالي توفير نظام أقوى وأكثر متانة. الحماية ضد المشهد المتغير باستمرار للمخاطر السيبرانية.

الكلمات المفتاحية: الأمن السيبراني، التعلم الآلي، شجرة القرار، الغابة العشوائية، KNN.