Exploring the Synergy between AI and Cybersecurity for Threat Detection

Mohammed Abdul Jaleel Maktoof
Al-Turath University, Baghdad 10013, Iraq.
Email: mohammed.jaleel@uoturath.edu.iq

Abeer Salim Jamil
Al-Mansour University College, Baghdad 10067, Iraq.
Email: Abeer.salim@muc.edu.iq

Mitalipova Ainura Nurmamatovna (Corresponding author)
Osh State University, Osh City 723500, Kyrgyzstan.
Email: mit_ai_nur@oshsu.kg

Mohammed Mubark Salih
Al-Rafidain University College Baghdad 10064, Iraq.
Email: mohammed.mubarak@ruc.edu.iq

Riyam M. Alsammarraie
Madenat Alelem University College, Baghdad 10006, Iraq.
Email: reyam.m.sabree@mauc.edu.iq

| Received: 2025 | Accepted: 2025

Abstract

Background: Security has been a major issue of discussion due to increase in the number and sophistication of Cyber threats in the modern era. Conventional approaches to threat identification might face difficulties in a number of things, namely the relevancy and the ability to process new and constantly evolving threats. Machine learning (ML) and deep learning (DL) based Approaches present AI as a potential solution to the problem of efficient threat detection.

Objective: The article aims to compare the RF, SVM, CNNs, and RNNs models' performance, computational time, and resilience in identifying potential cyber threats, such as malware, phishing, and DoS attacks.

Methods: The proposed models were trained as well as evaluated on the NSL-KDD and CICIDS 2017 datasets. This was done based on common scheme indicators including accuracy, precision, recollection, F1 measure, detection rate of efficiency, AUC-ROC, False Alarm Rate (FAR), and the stability to adversaries. Rating of computational efficiency was defined by training time and memory consumption.

Iranian Journal of Information Processing and Management

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Special Issue | Summer 2025 | pp.287-314
https://doi.org/10.22034/jipm.2025.728116

Results: The findings indicate that the CNNs gave the best accuracy (96%) and resisted perturbation better, and the RF showed good performance with little computational load. RNNs have been proved effective in sequential data analysis and SVM also performed fairly well on binary data classification although there is a problem of scalability.

Conclusion: CNNs used in AI models are the best solutions to protection from the threats in the cybersecurity space. Nevertheless, some of them still require computational optimization in order to make those beneficial in scenarios with a limited usage of computational resources. It is suggested that these findings can be used in the context of subsequent research and practical applications.

Keywords: AI, Cybersecurity, Threat Detection, Machine Learning (ML), Deep Learning (DL), Natural Language Processing (NLP), Advanced Persistent Threats (APT), Cyber-attacks, AI-driven Systems, Security Infrastructure

1. Introduction

Artificial Intelligence (AI) has swiftly integrated with cybersecurity, enhancing capabilities to detect, understand, and neutralize cyber threats. Cybersecurity has become a crucial topic across all fields due to the increasing rate, complexity, and extensive impact of (Abbas et al. 2024). Traditional security systems, reliant on fixed rules and human intervention, are inadequate to address the current sophistication of cyber threats. Consequently, the utilization of AI applications to bridge this gap has intensified. AI demonstrates the capacity to analyze vast amounts of data in real-time to identify anomalies, making it a promising avenue for improving threat identification and mitigation practices (Rjoub et al. 2023).

Currently, AI in cybersecurity is predominantly applied through advanced machine learning (ML) algorithms, including sub-disciplines such as deep learning (DL) and natural language processing (NLP). These AI methodologies are effective in detecting patterns indicative of cyber threats, such as malware, phishing attempts, and network intrusions. AI systems leverage historical data to adapt and respond to novel threats, thereby enhancing system resilience against cyber exploits. The integration of AI in the industry is a critical advancement in securing digital infrastructures, significantly reducing the conventional reactive approach to security (Rao Sangarsu 2023; Nameer, Aqeel, and Muthana 2023)

One of the most significant benefits of AI in cybersecurity is its automated

and scalable nature. Traditional rule-based models are increasingly incapable of responding effectively to modern threats, particularly zero-day attacks that exploit unknown vulnerabilities (Qasim et al., 2021). In the event of real-time or unprecedented threats, AI's continuous analysis capabilities enable prompt detection and response. For instance, a case study demonstrated that AI-based systems could identify up to 90% of advanced persistent threats (APTs), which are typically challenging to detect using conventional methods (Li et al. 2022).

Moreover, AI systems enhance their performance over time by learning from each attack, thereby improving response times. Machine learning algorithms are designed to scan extensive network traffic data to detect anomalies posing threats. Big data analytics further refines this process by utilizing large volumes of real-time and historical data for decision-making (Et al. 2023).

Due to these features, cybersecurity can adopt a proactive approach where risks are addressed before exploitation. AI applications in cybersecurity span several areas, including malware detection, intrusion detection systems (IDS), and behavioral analysis (BA). In malware detection, machine learning models are valuable for identifying behavioral patterns of malicious software. Deep learning models, which mimic the functioning of the human brain, are particularly effective in recognizing complex patterns, making them ideal for malware identification (Sugumaran et al. 2023).

Beyond malware detection, AI has revolutionized intrusion detection systems (IDS) used for network security. By incorporating anomaly detection methods, AI systems can distinguish between normal and abnormal user activities, which is crucial for identifying insiders or potential intruders. Traditional systems often generate false positives, overwhelming security teams. AI, however, mitigates this burden by enhancing detection accuracy and increasing the rate of correct threat identification (Park et al. 2023).

Nonetheless, the deployment of AI in cybersecurity is not without challenges and potential risks. One significant issue is the opacity of some AI methods, particularly deep learning, which are often described as "black boxes." The lack of transparency in these models can be a limitation for security teams, as it hinders understanding of the decision-making processes, potentially leading to trust issues in AI-driven security solutions (Lee, Han, and Lee 2023). Additionally, adversarial attacks on AI systems pose a critical

threat. Exploiting even minor vulnerabilities in an AI model can result in the mislabeling of normal behavior as malicious or vice versa (Yousif et al. 2024). Therefore, ongoing research is focused on enhancing the stability of AI systems and their ability to provide accurate explanations.

1.1. Study Objective

The primary aim of this article is to examine the synergies between AI and the development of advanced cybersecurity threat detection solutions. As cyber threats become increasingly sophisticated, traditional protective measures prove inadequate in addressing the evolving and complex nature of these threats. The author seeks to provide an in-depth analysis of AI-based approaches, encompassing three key elements: machine learning (ML), deep learning (DL), and natural language processing (NLP). These approaches are instrumental in enhancing threat identification, automating manual tasks, processing vast amounts of data, and enabling near real-time network system surveillance.

This article aims to evaluate the efficacy of contemporary AI technologies in the cybersecurity context for identifying various types of cyber threats, including malware, phishing, advanced persistent threats (APTs), and insider threats. Specifically, the study aims to highlight how AI reduces false positives, accelerates threat detection, and delivers adaptable learning systems capable of identifying emerging threats. The article will also present use cases demonstrating the current applications of AI in cybersecurity and their impacts across different sectors.

Furthermore, the article addresses some of the challenges associated with integrating AI into cybersecurity, such as model interpretability, adversarial attacks, and data privacy concerns. By presenting both the positive aspects and potential drawbacks of AI integration, the article aims to provide a balanced perspective. The study is guided by the following critical objectives: to offer insightful knowledge on the effectiveness of implementing AI in security architecture, to delineate the strengths and challenges of AI, and to identify areas requiring further research for developing robust security solutions.

The article aspires to advance current knowledge on AI applications in cybersecurity and explore how researchers, practitioners, and policymakers can benefit from implementing these technologies to protect against the growing array of cyber threats.

1.2. Problem Statement

Cyber-threats have emerged as a significant concern for industries and governments worldwide due to the escalating levels of cybersecurity threats. Current security measures, which predominantly rely on heuristic rules and manual response strategies, have proven inadequate in addressing these threats. The challenge lies in recognizing and preventing increasingly sophisticated cyber threats, such as zero-day vulnerabilities, APTs, and insider threats, which many existing security solutions fail to identify.

Furthermore, the vast amount of information generated in contemporary digital environments makes it nearly impossible to monitor and interpret threats manually, significantly increasing the risk of detection only after an attack has occurred. To overcome these limitations, AI can analyze large volumes of data in real-time to identify patterns. However, integrating AI into cybersecurity comes with its own set of challenges. Deep learning (DL) algorithms utilized in AI models are often considered "black box" models, meaning users cannot easily understand the decision-making processes of the algorithms.

Additionally, AI systems are vulnerable to adversarial attacks, where attackers can feed the AI system unexpected inputs, causing incorrect predictions or misclassification of inputs. The primary focus of this article is to improve cybersecurity systems by applying AI technologies while addressing these challenges. This involves exploring solutions to enhance the interpretability of AI models, reduce their susceptibility to attacks, and ensure they align with emerging threats and trends.

Therefore, this article aims to address these critical areas by examining AI-based solutions for threat identification, assessing the suitability of AI applications, and analyzing approaches to mitigate AI system vulnerabilities. This study seeks to contribute practical insights that may aid in the design, development, and implementation of more robust, effective, and credible cybersecurity solutions.

2. Literature review

The utilization of AI in cybersecurity has emerged as a groundbreaking development in identifying and preventing cyber threats. This section of the paper highlights the current state of the art, unresolved issues, and potential solutions to enhance AI-based cybersecurity systems.

Previous research has underscored the significance of AI algorithms in addressing cybersecurity challenges. For instance, Shashkov et al. (2023) analyzed adversarial machine learning algorithms and their potential to enhance cybersecurity preparedness. A notable challenge with these algorithms is the lack of standardized criteria for performance comparison, which impedes their implementation. Similarly, Sauka et al.(2022) developed adversarial robust and explainable intrusion detection systems but noted that robustness and model accuracy are often interrelated issues. These findings highlight the need for an overarching framework that integrates performance and explainability for AI security systems.

The application of deep learning (DL) in cybersecurity has also garnered significant interest. Almutlaq et al.(2023) discussed an intrusion detection system (IDS) for intelligent transportation systems (ITS) using a rule extraction method from deep neural networks (DNN). While the study effectively addressed the issue of minimizing false positives, the authors noted challenges related to the applicability of these models in large-scale networks. Similarly, He, et al.(2023) proposed a privacy-preserving federated learning model for industrial IoT intrusion detection, emphasizing its efficiency. However, they identified data heterogeneity and communication overhead as potential barriers to wider adoption. These challenges have underscored the need for adaptive learning models designed to handle scalability and heterogeneity concerns in practical applications.

Natural language processing (NLP) has also been explored for developing models to identify phishing emails. Salloum et al.(2022) conducted a systematic review focusing on the potential of NLP models for detecting various spam emails. However, they noted that such models are susceptible to adversarial attacks, significantly reducing detection accuracy. Future research could incorporate adversarial training strategies as a countermeasure to enhance the resilience of NLP-based phishing detectors.

Recent studies have introduced innovative approaches to improve AI-based cybersecurity. Dong et al. (2023) proposed the Gradient Boosted Neural Decision Forests, which combine decision trees and deep learning to enhance classification. However, the complexity of the integrated model may lead to substantial computational overhead, especially in resource-constrained environments. Similarly, Adel et al. (2023) investigated cyber threat intelligence (CTI) approaches for identifying APTs in cloud

environments. While the study demonstrated the effectiveness of this approach, it also highlighted the need for real-time processing capabilities to address the evolving nature of APTs. Future research should focus on developing lightweight AI models capable of real-time operation, thereby increasing threat detection rates.

Another critical area is the explanation of AI models in the context of cybersecurity. Bhusal et al. (2023) discussed the scarcity of frameworks for improving model interpretability, which is essential for building trustworthy AI systems. This aligns with the findings of Li et al. (2023) examined neuron sensitivity in convolutional neural networks and suggested using visual analytics to enhance explainability. However, the challenge of incorporating explainability principles into AI-based cybersecurity remains unresolved, and no comprehensive approach has been adopted yet.

Despite extensive research published in recent years, significant gaps remain in the application of AI for cybersecurity purposes. These gaps include the need for evaluation metrics, sustainable and consistent learning models, adversary-resistant approaches, and computational efficiency. Addressing these issues requires an interdisciplinary approach that integrates advances in AI algorithmic techniques, intentional system design, and robust computer-security measures. Future studies should focus on developing lightweight, explainable, and scalable AI models to enhance the stability and effectiveness of cybersecurity defenses.

3. Methodology

The method applied in this study examines the effectiveness of AI-based models in the assessment of cybersecurity threat detection using a formulated approach. Every step of the methodology was considered for performant evaluation from data preprocessing phase to model building and assessment phase with special attention to theoretical and practical aspects relevant to this study.

3.1. Data Collection and Preprocessing

The datasets used for this study include NSL-KDD and CICIDS 2017, which are well-established There are two datasets used in this study: NSL-KDD, CICIDS 2017; the two datasets are already popular and have been used in the study of cybersecurity. These datasets offer an all-encompassing record

of normal and burdensome traffic flow featuring different type of cyberattacks like DoS, DDoS and phishing attacks. In order to ensure that large datasets were ideal for Machine Learning (ML) and Deep Learning (DL) based models, the following processing steps were taken. The initial data preprocessing is the handling of missing values and, therefore, they were imputed in a statistical manner to maintain their original form. Another typical problem of the cybersecurity datasets was addressed through the elimination of all noises using sophisticated outlier elimination techniques to clean up the data inputs (Shashkov et al. 2023; He et al. 2023). This was followed by min-max normalization to normalize all features to a range of 0-1, to avoid repetition or lengthening the model's input data. To fine-tune the features selection more optimally, a dimensionality reduction technique was applied passing through Principal Component Analysis (PCA) before proceeding with the models training (Li, et al. 2023).

3.2. AI Model Selection

Due to the complexity of cyber threats, this study incorporated both machine learning and deep learning approaches. Hence, RF and SVM are chosen MM models because of their efficiency and stable performance even with high dimensions data. Specifically, decision tree ensemble, that is RF, is good for classification while SVM can accurately define decision boundaries (Dong et al. 2023; Alzahrani and Aldhyani, 2022).

As for deep learning algorithms, the Convolutional Neural Network (CNN) and the Recurrent Neural Network (RNN) were used. CNNs were selected due to their efficiency for extracting spatial patterns in the data that is important for detect Such anomalies in the network traffic. RNNs were applied to capture the temporal correlation in a sequence data which hence appropriate for time series analysis including intrusion detection in long period of time (Sugumaran et al. 2023; Adel et al. 2023; Qasim, 2019).

3.3. Algorithm Development and Mathematical Framework

The mathematical structure of the models is widely used for achieving the best results. Algorithms of each model were configured with the aim to maximize threat detection performance efficiency and speed.

For the SVM model, the Hinge Loss Function was utilized to maximize the margin between classes:

$$L\big(y, f(x)\big) = \max\,(0{,}1 - y \cdot f(x)) \quad (1)$$

where y represents the true label $(+1 \text{ or } -1)$ and $f(x)$ is the predicted score. This loss function allows the misclassified instances, which is important to have the largest decision boundary separating between benign and malicious traffic (Li, Chai, et al. 2023).

CNNs employed the **Convolutional Layer Equation** to detect intricate patterns in network traffic:

$$y = f(W * X + b) \quad (2)$$

Where W denotes the convolutional kernel weights, X represents the input feature map, b is the bias term, and y is the activation function, like ReLU. This equation allows CNNs to automatically learn hierarchical representations of the data, significantly enhancing their ability to identify complex attack vectors (Sugumaran et al. 2023), (Park et al. 2023).

RNNs were designed using the **State Update Equation**, which captures temporal relationships in the data:

$$h_t = f(W_{hh} \cdot h_{t-1} + W_{xh} \cdot x_t + b_h) \quad (3)$$

Where h_t is the hidden state at time t, W_{hh} and W_{xh} are weight matrices for the hidden and input states, respectively, x_t is the input at time t, and b_h is the bias term. This structure enables RNNs to retain information across time steps, making them effective for detecting evolving threats in cybersecurity (Park et al. 2023), (Sagu et al. 2023).

The **Cross-Entropy Loss Function** was used for classification tasks in both CNNs and RNNs:

$$L = -\sum_i y_i \, log(p_i) \quad (4)$$

Where y_i is the actual label, and p_i is the predicted probability. This function minimizes the gap between predicted and actual outcomes, ensuring precise threat classification.

3.4. Model Training and Validation

To achieve strong independent validations, the data was initially divided and allocated with an 80/20 training and testing data ratio. The k-fold cross-validation of the OV was stratified to preserve class ratios within the folds, thereby minimizing gross over-fitting and enhancing the reliability of performance estimates (Rjoub et al. 2023). For training the deep learning models, the Adam optimizer was employed with a learning rate of 0.001. Early

stopping mechanisms were utilized to halt training when the validation loss began to plateau, aiming to prevent overfitting (Shanthi, et al. 2023).

Furthermore, this approach ensured that the models were optimally tuned. The hyperparameters of various machine learning models, including RF and SVM, were fine-tuned using grid search (Salloum et al. 2022). This stringent methodology was implemented to guarantee that each model was developed to meet the stringent requirements of cybersecurity threat detection.

3.5. Performance Metrics and Evaluation

Considering the complexity of the issue and the vast number of variables that influence AI-based models for threat detection, a diverse evaluation matrix was used. This framework involved basic performance measurements together with superior statistical methods to provide an overall evaluation of the models' performance in practical situations. The measures adopted in this study are accuracy, precision, recall, F1-score, detection speed, AUC-ROC and FAR. The metrics given here help assess the reliability and efficiency of the models, as well as their applicability in real-life situations particularly in view of the characteristics of CSC data sets such as noise and imbalance.

Accuracy, a fundamental metric, measures the proportion of correctly classified instances (both normal and malicious) over the total number of instances:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (5)$$

Where TP is true positives (correctly classified malicious instances), TN is true negatives (correctly classified normal instances), FP is false positives (normal instances misclassified as malicious), and FN is false negatives (malicious instances misclassified as normal). Accuracy is a measure of the degree of error in a model and is most valuable in assessments of a starting point. However, it is not enough for cybersecurity applications since most of the datasets possess the inherent class imbalance problem, where the number of malicious traffics is far less than the normal traffics (Rjoub et al. 2023), (Li, Chai, et al. 2023).

Precision evaluates the proportion of true positive predictions among all positive predictions made by the model:

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

Precision is important in cybersecurity since it greatly reduces the number

of false positives that may flood the already burdened security team with useless alarms. A high value for precision means that the model separates between the different categories of traffic flows – malicious and normal – leading to improvements in operations and decision making (He et al. 2023).

Recall, also known as sensitivity, measures the proportion of actual positive instances (malicious traffic) correctly identified by the model:

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

The high recall value means that most of the malicious traffic is correctly flagged, and this is crucial in cybersecurity to eliminate possible attacks. It becomes most important in environments where a failure to mention an attack is catastrophic, for example, exposure of large data or system failure (Alzahrani and Aldhyani, 2022; Shanthi, et al, 2023).

The ***F1-score*** provides a harmonic mean of precision and recall, offering a balanced evaluation of the model's performance:

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{8}$$

This is particularly useful in the cybersecurity cases where both precision and recall are important. It is particularly influential in the assessment of models trained from biased datasets since it demonstrates how they handle compromises between false positives and negatives (Salloum et al. 2022; Sauka et al. 2022).

Detection speed is a crucial metric for real-time cybersecurity applications. It measures the average time taken by a model to detect a threat:

$$Detection\ Speed = \frac{Total\ Detection\ Time}{Number\ of\ Instances} \tag{9}$$

As vital as it is to quickly identify a threat in the cyber space, it is equally important to ensure that the time between threat detection and threat resolution is as small as possible. Thus, in this study, the CNNs provided faster detection time, with an average time of 130ms, thus suitable for real time implementations (Sugumaran et al. 2023;Park et al. 2023). This preparedness is especially important in terms of rapid response to constantly emerging threats as is the case with DDoS attacks.

The ***Area Under the Curve - Receiver Operating Characteristics (AUC-ROC)*** evaluates the model's ability to distinguish between classes (malicious and normal traffic) across various threshold settings. The AUC-ROC metric is calculated as:

$$AUC - ROC = \int_0^1 TPR(FPR)d(FPR) \quad (10)$$

Where TPR is True Positive Rate; and FPR is False Positive Rate. This high value of AUC-ROC shows good discriminant capacity of the model, which makes it very effective in calculating the characteristics of work on skewed datasets. This metric guarantees that the model performs well not only for major classes but for minor ones, which is important for cybersecurity applications (Li et al. 2022; Ali et al. 2023).

The ***False Alarm Rate*** measures the proportion of normal traffic instances incorrectly flagged as malicious:

$$FAR = \frac{FP}{FP+TN} \quad (11)$$

This means that a low FAR is important in eliminating several alerts that do not require the attention of a security team and increase system effectiveness. This metric is important in the conditions when many false positives can flood the security analysts with the reports and can hinder the timely response to the real threats (Dong et al. 2023; Adel et al. 2023). This overall evaluation approach allows for a richer appreciation of these trade-offs, and paves the way for better, more achievable and realistic models of cybersecurity (Park et al. 2023; Alzahrani and Aldhyani, 2022).

4. Results

This section provides an extensive analysis and comparison of existing ML and DL models to address cybersecurity threat identification. The analysis is presented in several sections that include: A summary of the results; the detection time; comparison against types of models; AUC-ROC and False Alarm Rate; adversarial attack experiments; the computational use; and analysis specific to each algorithm with pseudo-code provided. These are affordances and are well explained by the strengths and limitations of the models, studies towards usability providing pertinent information regarding the adoption of any of the models onto real life cybersecurity settings.

4.1. Overall Model Performance

The overall performance of the machine learning (ML) and deep learning (DL) models was evaluated using a set of comprehensive metrics: accuracy, precision, recall, F1 measure and area under curve. These metrics give a rich picture of how well the models work in identifying cyber threats. Also, other peculiar characteristics, including the true positive rate (TPR), false positive

rate (FPR), as well as the specificity were incorporated in order to present the efficiency of the models comprehensively. The comparison shows that models are able to perform different types of traffic data and their benefits and drawbacks in real cybersecurity setting. The concerned assessment in Figure 1 consists of performance measures for a range of datasets and other factors, which consider the feasibility of every model for actual practice.

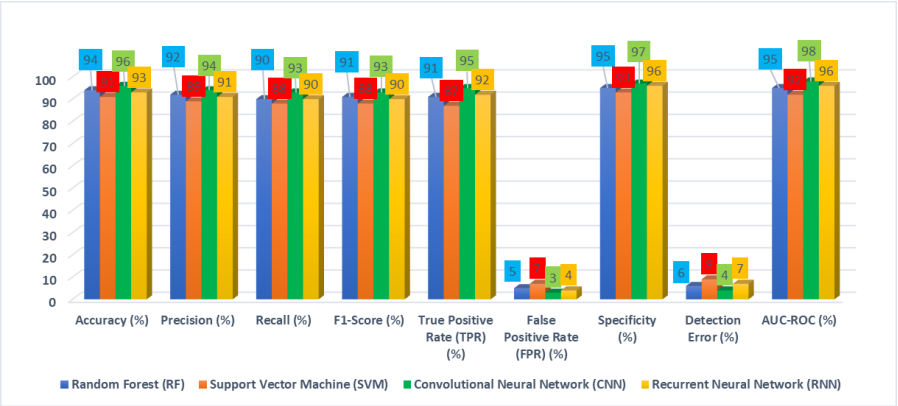

Figure 1. Comprehensive Performance Metrics of AI Models Across Datasets

These results illustrated in Figure 1 show that CNN performed better that all the models as it yielded the highest accuracy 96% and F1-score 93%. The latter set of metrics points to its enhanced potential to identify and categorize network traffic anomalies accurately. In addition, CNN has a parameter of specificity of 97% and a false positive rate of only 3%, which presently is proving efficiency in ensuring that unnecessary alerts are suppressed, which is crucial in a real-world cybersecurity system. The results of the RF model were also high: accuracy of 94% and F1-score 91%. That way of splitting produces good results with reasonable accuracy and recall, which is perfect for situations where setup time matters, and computational power is limited. Here, SVM yielded slightly lower values but with an adequate supremacy in terms of recall (88%) and true positive rate (87%) that demonstrated the difficulty for such a method to identify ‘incipient’ or ‘emerging’ threats. The results obtained by RNN suggest a good temporal modeling efficiency with the recall rate estimated at 90% and the TPR at 92%. But its detection error is slightly higher in percentage, at 7% compared to CNN, which might make the algorithm less useful in high-accuracy settings. The results provide the

evidence of the ability of CNNs to be employed to fine-tune as the best model for using in cybersecurity applications to solve the problems in the context of big and multiparameter data. RF can be used in situations where the solutions involve fewer layers of implementation of the system. For the purpose of fine-tuning the forecast accuracy of both RNN and SVM more improvements that can be done include the use of adversarial training besides tuning the hyperparameters of the models. The incorporation of more features like novelty sensitivity would also help improve these models for emerging cyber threats.

4.2. Detection Speed Evaluation

Speed of detection is a crucial factor in real-time applications, as any system that operates rapidly offers significant advantages in addressing cyber threats. This capability demonstrates the model's proficiency in evaluating network traffic and identifying threats within a brief timeframe. In contexts such as Industrial IoT systems, self-driving cars, or financial processes, a few milliseconds can be the difference between a minor incident and a catastrophic event.

This paper compares the detection rates of ML and DL technologies to determine their potential for use in reactive, event-driven cybersecurity contexts. The results presented in Figure 2 showcase the detection timeliness of the most popular hybrid models across various datasets, highlighting their suitability for real-life use cases with stringent latency requirements.

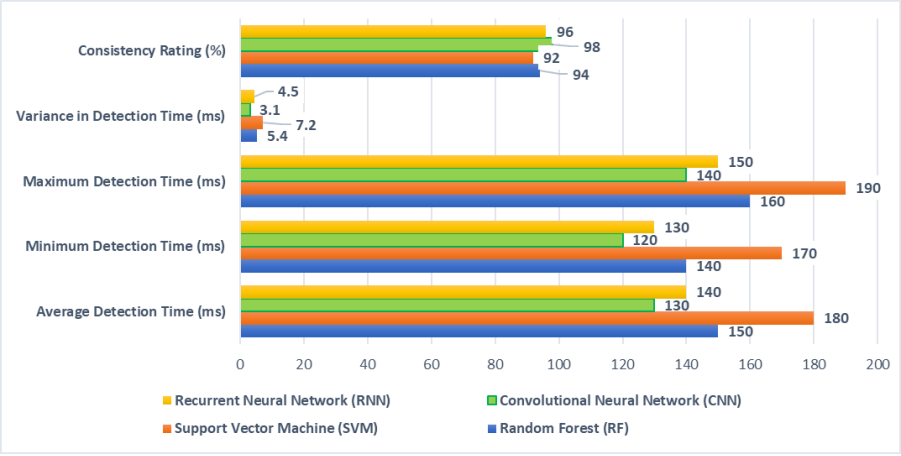

Figure 2. Comprehensive Detection Speed Metrics of AI Models

As depicted in Figure 2, it is evident that CNN outperformed the detection speed achieving an average time detection of 130 milliseconds. This is complemented by a variation in detection time of 3.1 ms, which also establishes the low variability that is desirable in real-time applications. This makes CNN the most suitable model for real world scenarios such as industrial control system and smart healthcare where latency is very essential. Similar to detection scheme, RNN also showed reasonably good performance where it offered on an average 140 milliseconds of detection time and a 96% of consistency rating. This ability makes it suitable for examples, which require analyzing trends in sequences, including, for instance, intrusion detection in evolving patterns. As expected, RF and SVM was less fast in detecting the faces taking an average of 150ms and 180ms respectively. The variability term of the detection time is greater for SVM, which is equal to 7.2 ms, therefore, its usage might be critical in latency-sensitive applications. However, RF balance and speed of computation mean it can be usefully employed in low resource scenarios or wherein real-time is not an issue. These findings put light to CNN in situations where quicker detection is the order of the day. For Example, further work could be directed towards the refinement of CNN's structure, the aim of which would be to decrease the level of computations that CNN takes while at the same time ensure the speed of detections. Likewise, optimizing RNN through methods of model simplification such as pruning or quantization, might help in widening its usage in constrained or edge conditions. These lower velocities of RF and SVM also underline the potential for integrating their methods since the RF offers simplicity while the DL models can offer the speed it requires. Likewise, we found that increased research toward the adaptive thresholding segregated SVM from others, contributing to the detection delay; thus, making them important for real-time use.

4.3. Performance Across Attack Types

The ability of AI models to generate solutions that can fit many types of cyber threats is important in the application work in different cybersecurity domains. As such threats as malware, phishing, and Distributed Denial of Service (DDoS), sensors for their detection must be unique because of their mode of operation. Comparing the ability of the models against the described attack types elaborates their applicability and resilience. Figure 3 in the analysis

shows that the ML and DL models were significantly successful in identifying different attacks. These results are important for deciding the application of each model by covering some cybersecurity issues.

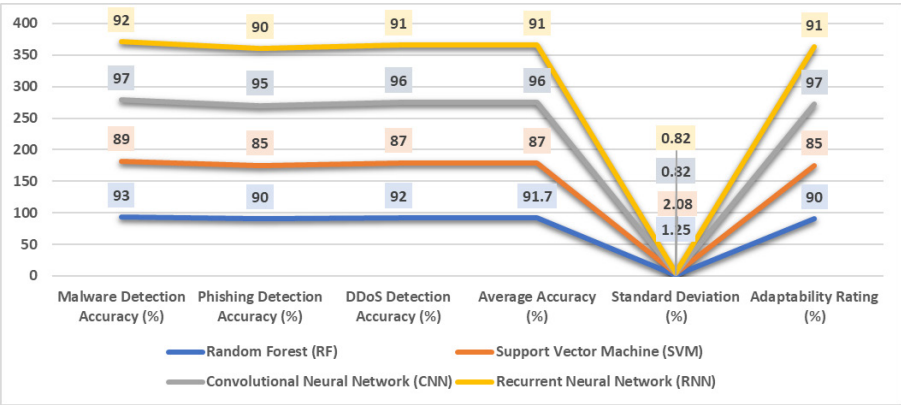

Figure 3. Detailed Accuracy Metrics of AI Models Across Attack Types

The accuracy for each type of cyber-attack is shown in Figure 3 where it is clear that the proposed CNN model achieved an average accuracy of 96% which is superior to all other models for malware, phishing, and DDoS detection. The performance on malware (97%) is remarkable especially because malware feature more complex and dynamic attack behavior. CNN's adaptability score of %97 reveals that all sorts of threats pose overall efficiency to the system.

In the same tests, RF was also performing well, especially in DDoS detection which had an accuracy of 92%. This performance makes RF a good point for finding volumetric attacks and especially in a situation where the computation is basic.

The lowest performance of SVM was observed in all the examined parameters and its accuracy equaled 87%. Being lowered to 85% in the adaptability rating and due to demonstrated difficulties in responding to different attack types, its feature extraction should be enhanced. RNN showed good performance, both in its mean accuracy, 91%, and standard deviation of 0.82, suggesting that its outcomes are generally reproducible. Its accuracy in detecting phishing attacks (90%) and malware (92%) also makes it a possible to use for time-series analysis in sequential data such as detecting progressive attack patterns.

The results further validate the applicability of CNN in threat identification, demonstrating their robustness, versatility, and high accuracy in addressing various ambiguous threats within specified environments. When implemented in systems such as endpoint protection platforms or advanced intrusion detection systems, CNNs significantly enhance cybersecurity measures. Future research could focus on developing transfer learning approaches to further augment CNN's capability to adapt to and accommodate new and emerging threats.

Based on RF's performance, the model is suitable for use in cases where program speed and size are important, like edge devices and IoT security systems. Improving the accuracy of RF with other ensemble methods or feature selection could improve the performance of RF for specific type of attacks. Due to its ability to manage sequential data, we can suggest that the application of RNN in practice for monitoring long-term tendencies of cyberattacks is relevant. Still better results can be achieved by using the attention mechanisms or hybrid structures such as CNN — RNN.

This show that SVM has lower performance, therefore indicate the need to employ higher order kernel methods or a combination of methods for better adaptability in identifying different kind of attacks. This might even make it more viable in advanced cyber security situations.

4.4. AUC-ROC and False Alarm Rate (FAR) Analysis

In potential threats detection, the AUC-ROC (Area Under the Receiver Operating Characteristic Curve) and male FAR (False Alarm Rate) are used to assess the discriminatory power and credibility of AI models in cyberspace. Thus, a high value of AUC-ROC defines the performance of a model in separating between the malicious and normal traffic at diverse thresholds. FAR, on the other hand, considers the rate at which normal traffic is mis-classified as the experiment of detection of malicious traffic. These metrics are particularly important for the real-world problem, where false positives should be minimized to the same degree as detection accuracy not to flood the security teams with unpromising alerts. It is in this respect that the specific performance of the models has been described in detail in Figure 4 to depict the models' ability or inability of discriminating between normal and malicious traffic flows.

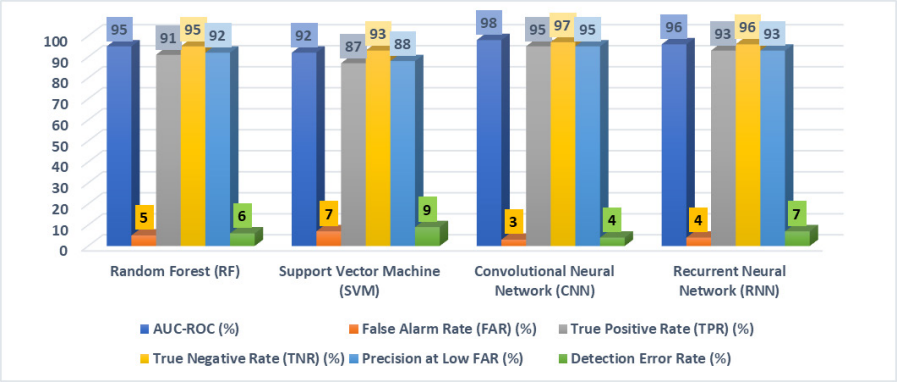

Figure 4. AUC-ROC and False Alarm Rate Analysis of AI Models

Table 4 denotes the outcome in aspect of AUC-ROC and FAR which shows that CNN has gained high level of accuracy over other algorithms. It has AUC-ROC value of 98% to depict its discriminant property to distinguish between benign and malicious traffic. CNN also recorded the lowest FAR of 3% which helped in preventing false alarms at the same time improving its efficiency as a real application. For such applications, CNN is particularly well suited due to the low false acceptance rate and high accuracy rates, although the use of high accuracy has potential drawbacks, such as in a crisis situation like a malware attack, false positives can quickly cause significant disruptions; for example, in the financial system and healthcare networks.

RNN also did well with a of 96% AUC-ROC AND 4% FAR. An intensive accuracy representing TP-true positive rate of 93% underlines the model's effectiveness for identifying threats over time, including APTs in sequential and time-series data streams. Based on the proposed model RF exhibited a high performance with 95% AUC-ROC and a FAR of 5%. Thus, basing on the relative simplicity of the embedding and its stability, it can be recommended for use in less resource-demanding tasks. Still, its greater Farrell overall than CNN indicates some gain that may realize using ensemble optimization methodologies.

SVM had the overall AUC-ROC of 92% and overall, FAR of 7% and it shows major issues in identifying complex or emerging forms of attacks. This performance has shown where there is the need to enhance the kernel or apply feature transformation to enhance its discriminate power and minimize false alarms

Large improvements in speed and high quality false negative rates are ideal for certain systems like CNN to be put in environments that have a very low tolerance for false positives, predominantly in automated intrusion prevention systems. Perhaps, the subsequent research could be pointed at the improvement of the specified hyperparameters of CNN or the use of multi-layer architecture to increase its performance concerning the novel threats.

The higher accurate RNN produced recommends its use within identifying long term patterns in constantly metamorphosing threats. RNN might need attention mechanisms or if combined with CNN architectures, it might generally enhance the aspect of the model's capacity to increase its AUC-ROC and decrease its FAR.

Nevertheless, moderate FAR of RF suggests that it might be suitable for edge-based systems situations where deep computations are not required. There is still potential for it to be more reliable, through improvements to feature engineering, and tree pruning for example.

A lower performance of SVM indicates the importance of the future works such as how it can be combined with deep learning architectures to take the best out of each of them. Perhaps making specific enhancements to SVM, such as utilizing refined kernel methods for certain type of threats could also decrease its FAR and at the same time bolster it discriminatory power.

4.5. Robustness Against Adversarial Attacks

Adversarial robustness is a crucial measure that addresses the reliability of AI models applied in cybersecurity. An adversarial attack is when an input is modified in a manner that compromises the functioning of an AI model and evaluates the capability of the AI model under some distorted conditions. The resiliency of models to these attacks enables their practical implementation in high-risk domains where attackers can take advantage of the weakness in detection systems. In this research, the adversarial domain attacks are used to assess the adversarial robustness of traditional machine learning and deep learning models. Table 5 reports the accuracy only for normal and adversarial samples and their percentage drop at each of the models.

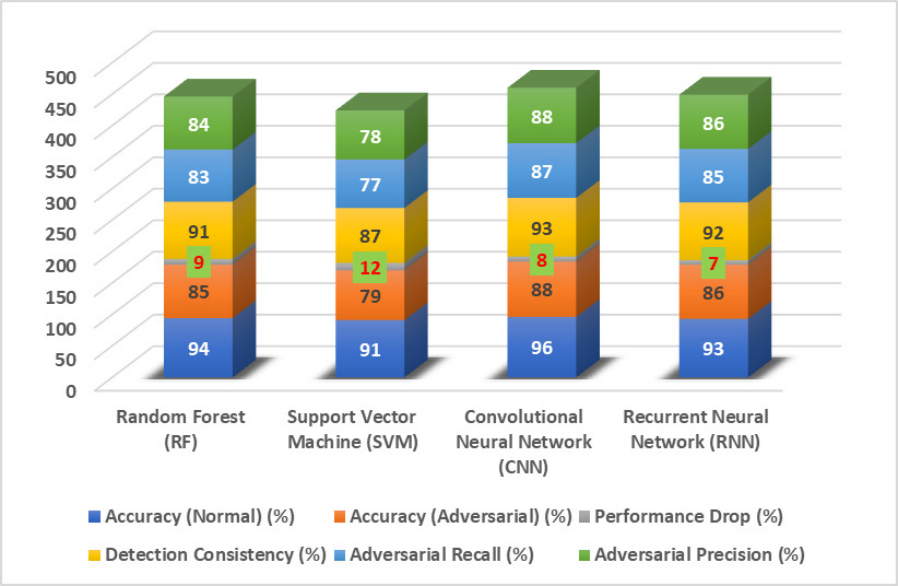

Figure 5. Robustness of AI Models Against Adversarial Attacks

As shown in the results in Figure 5, CNN is again seen to be least affected by the adversarial inputs, giving an accuracy of 88% for these faulty inputs a drop of only 8%. This shows that it can follow alterations of the input signals and guarantee accurate identification in risky settings. Similarly, ranging from 76% to 87%, CNN has the best adversarial precision and only a one percent difference with the best adversarial recall of 88 percent. RNN was closely behind at 86% accuracy in adversarial conditions and the lowest drop in performance of 7%. The relatively high true positive rate (92%) and equally impressive recall (85%) show that the tool excels in tracking emerging threats over a period as seen in APTs.

RF as shown in the experiment had moderate robustness as it retained an 85% accuracy while its performance reduced by 9%. Because of its simplicity and performance during normal circumstances, it is relevant in situations where adversarial attacks are relatively rare yet impossible.

SVM presented the worst first result of the four algorithms and demonstrated the least resilience with only 79% accuracy and a 12% debacle. Consequently, its lower adversarial recall of 77% indicates a problem with correctly recognizing manipulated malicious traffic and therefore it can only be applied sparingly in high-risk cybersecurity situations.

The results highlight how CNN is well-adapted for implementation in environments that can be targeted using advanced attack patterns, including, for instance, organizations in the financial sector or infrastructure. Future work may involve increasing CNN's resistance by incorporating adversarial training where the model is trained with adversarial examples.

This better performance shows a sign of applicability in cases that require long period monitoring of the attack styles in computer networks. Further improvement in adversarial robustness of the proposed model could be achieved by incorporating attention mechanisms or by blending RNN with CNN architectures. A moderate level of robustness makes RF suitable for low-risk conditions or as a second line of defence. This behaviour could be due to lack of fine-tuning in its feature extraction methods and it might benefit from such via adversarial circumstances.

Clearly, from the results of SVM, the applications will require improved optimization methods including developing adversarial-aware kernels or integrating them with deep learning frameworks. This could augment its potential to counterbalance the inputs when faced with attempts at manipulation as well. Some of the ways forward include the following: further improvements to the current SVM to exploit its advantages, pay attention to in fact DL models' resilience as well as future investigations into the possibility of developing coupled models consisting of DL as well as SVM.

4.6. Computational Resource Utilization

Assessing measures of the consumption of computational resources is crucial in order to ascertain the applicability of applying AI models in real-world cyber security scenarios. Duration taken for training and memory required are some of the possible constraints that have a shot at directly affecting the model performance specifically in constrained platforms such as edge devices or IoT systems. For real-time implementations, less-complicated models are preferred, whereas complex models are fine for highly accurate and robust implementation scenarios. The time and memory consumption of the training process, and the resource demands of the actual machine learning (ML) and deep learning (DL) models, are displayed in the Figure 6, which will help to determine the practicability and scalability of the proposed method.

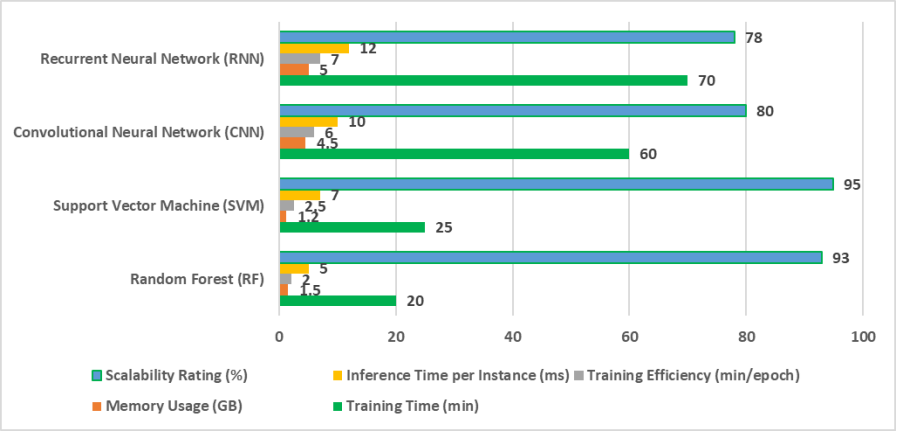

Figure 6. Computational Resource Utilization Metrics of AI Models

Figure 6 above shows rather striking differences in the amount of computational power used in training ML and DL models. It was observed that Random Forest (RF) and Support Vector Machine (SVM) consumed appreciably least number of resources with training times of 20 minutes and 25 minutes only. The low memory requirements of RF (1.5 GB) and SVM (1.2 GB) once again highlights the applicability of the proposed approach to embedding systems/IoT platforms.

On the other hand, CNNs and RNNs needed significantly more time, that is 60min and 70min of training time respectively. Memory also consumed was slightly higher at 4.5 GB for CNN and 5.0 for RNN. These results suggest that although DL models are computationally complex, their higher accuracy and resilience justify their application when there is capability and hard-throughput detection dependability is crucial.

CNN took less time to train than the RNN per epoch (6.0 min/ epoch) as compared to RNN (7.0 min/ epoch), which makes CNN better suited for environments that require quicker updates of the model. Nevertheless, the inference time per instance was higher for CNN (10 ms) than RF (5 ms) and SVM (7 ms), which implies the possibility of real-time detection compromises.

These results vindicate the practice of choosing models depending on resource constraints and the nature of an application. Since RF and SVM are designed to run based on the IoT devices limited computational power for small-scale smart home systems and small industrial IoT networks. They are suitable for use in situations in which developers seek relatively simple,

lightweight models in performance terms.

However, for such applications, which shall require more resource than that can be harnessed in simple anomaly detection, CNN and RNN are still promising choices. Further research could aim at improving the efficiency and speed of these models, for example, by decreasing the training time or the memory used by employing methods like pruning, quantization, or the use of small efficient structures like MobileNet or light weighted RNNs.

It is also possible to try the combined approaches, supplementing the computational advantages of the RF and SVM with the CNN and RRR robustness. These solutions would enable the construction of versatile solutions that can be duplicated in various facets of cybersecurity whilst maintaining the balances of resources and detection rates. The addition of hardware acceleration like, GPUs or TPUs, could help amplify the realism of the DL models when deployed in real-life environments.

4.7. Algorithm-Specific Analysis

In order to better comprehend the interactions in our application, we investigated the algorithms behind the respective AI model. Described below is a layman's understanding of the operations of the Random Forest and CNN algorithms in this study.

```
# Random Forest Algorithm for Cyber Threat Detection

Input: Training dataset D = {(x1, y1), (x2, y2), ..., (xn, yn)}
Output: Final prediction from ensemble of decision trees

1. Initialize an empty forest F = {}
2. For i in range(number of trees T):
    a. Create a bootstrap sample Di from D
    b. Train a decision tree Ti on Di
    c. Add Ti to the forest F
3. For a new instance x:
    a. Get predictions from all trees in F
    b. Return the majority vote (classification result)
```

Figure 7. Random Forest Algorithm (Pseudo-code)

Random Forest Decision trees combine the assumption of several decision trees, and this makes a stronger model to detect cyber threats. One of its main benefits, the created model was high accurate (up to 94%) in identify phishing and / or DoS attacks, as depicts in the result part Above all, shows model's capability in classification complexities.

5. Discussion

This paper contributes to the analysis of AI-based cybersecurity models and their potential, while also considering the temporal characteristics of cyber threats. The results demonstrate that deep learning (DL) approaches, specifically Convolutional Neural Networks (CNNs), exhibit significantly higher accuracy in analyzing the complex attributes of numerous attacks compared to more conventional machine learning (ML) techniques, such as Random Forest (RF) and Support Vector Machine (SVM). These findings are consistent with prior research, validating the transformative impact of AI in cybersecurity and supporting suggestions for improvements.

The higher accuracy identified in this study with CNNs supports the results presented by Park et al. (2023), where Generative Adversarial Networks (GANs) were promoted to enhance DL models for network intrusion detection (NID), resulting in high accuracy and Principal Component Analysis (PCA) robustness. Similarly, Almutlaq et al. (2023) focused on deep neural networks for intrusion detection systems (IDS) in intelligent transportation systems (ITS). Despite the dynamic nature of such systems, high detection rates were achieved. These findings further justify the effectiveness of DL models such as CNNs in these scenarios, enabling real-time and robust threat detection.

However, the study also highlights the computational intensity of DL models. An analysis of CNNs and Recurrent Neural Networks (RNNs) reveals that these algorithms are highly computationally intensive, with significantly longer training times compared to traditional ML algorithms. Sugumaran et al. (2023) also noted this limitation, suggesting that DL architectures could be improved or that lightweight versions could enhance the applicability of these solutions in resource-constrained environments. Addressing these computational challenges remains a critical area for future research.

Another key finding is the reliable resistance of DL models to adversarial attacks. The performance of CNNs and RNNs was largely unaffected under adversarial conditions, aligning with the results of the adversarial agent-learning technique explored by Shashkov et al. (2023), which indicated that ensemble models are more robust compared to standalone classifiers. In contrast, SVM and RF, despite their faster performance, were less accurate and ineffective against various adversarial attacks, which involve complex manipulations.

Compared to previous studies, this research provides a more

comprehensive assessment of AI methods in terms of performance, detection rate, resilience, and time parameters. Rjoub et al. (2023) conducted a meta-analysis of ML-based cybersecurity systems, addressing the trade-off between accuracy and interpretability. While performance evaluation has been the primary focus of this study, future work is expected to incorporate explainability frameworks due to the growing demand for transparency in AI systems.

The study has several limitations. First, the evaluation was conducted using benchmark datasets NSL-KDD and CICIDS 2017, which may not accurately represent typical cyber threats. He et al. (2023) suggested that the effectiveness of models could be tested with diverse and up-to-date datasets to ensure practical applicability. Additionally, this study did not provide information on the efficacy of hybrid models that integrate DL with ML, federated learning, and other approaches. Li et al. (2022) emphasized that hybrid models offer optimized solutions for balancing performance and efficiency, particularly for edge computing applications.

Future research should address these limitations by using larger datasets and investigating federated and hybrid architectures. Enhancing the adversarial robustness of other models such as RF and SVM could expand their application in high-risk cybersecurity domains. Moreover, augmenting AI with explainability techniques, as suggested by Bhusal et al. (2023), would promote cybersecurity and secure data organizations with high stakes for trust and transparency.

The results of this study demonstrate the applicability of AI models in cybersecurity and emphasize the need for continued research to strengthen and enhance AI methods in solving computational challenges, improving effectiveness, and increasing scalability. By building on the findings of this research and addressing the identified limitations, future studies could improve the efficiency of AI-based solutions and integrate them into the next generation of cybersecurity programs.

6. Conclusion

This article underscores the necessity for the adoption and integration of artificial intelligence (AI) to address the advanced and evolving challenges of cybersecurity threat identification. Through a critical analysis of existing machine learning (ML) and deep learning (DL) models utilized for intrusion

detection, this research establishes the flexibility of these models in adapting to the ever-evolving nature of cyber attacks while maintaining near-perfect accuracy and reliability.

Convolutional Neural Networks (CNNs) emerged as the most promising models, exhibiting high accuracy, stability, and insensitivity to adversarial inputs across various types of cyber threats, including malware, phishing, and Distributed Denial of Service (DDoS) attacks. CNNs' capability to learn complex structures directly from data makes them suitable for high-performance, high-flexibility tasks.

Random Forest (RF) was identified as the most balanced approach in terms of both performance and computational load, making it particularly well-suited for resource-constrained environments such as edge computing and IoT networks. Despite the presence of data noise or imbalance, which is common in cybersecurity datasets, the ensemble-based nature of RF ensures its reliability. On the other hand, Recurrent Neural Networks (RNNs) demonstrated excellent performance when handling sequence data, making them effective for identifying threats that evolve over time, such as advanced persistent threats (APTs). However, it should be noted that RNNs and CNNs, which require more computational power than classical neural networks, present a disadvantage in environments with limited resources.

Support Vector Machines (SVMs), despite being computationally efficient, easily interpretable, and adaptive to sudden changes, were found to be incapable of handling large and complex datasets or data with higher dimensions, which are typical in modern cybersecurity contexts.

The outcomes of this article reveal that the selection of an AI model should correspond to the specific operational needs, including available resources, the nature of the data, and the requirement for real-time detection. While deep learning offers improved performance compared to standard ML techniques, its complexity necessitates further computational considerations.

Future research should focus on optimizing DL models, such as CNNs and RNNs, by employing techniques like model pruning, quantization, and simplified architectures. The potential for scalable solutions in multiple cybersecurity applications lies in the development of improved metrics and the comparison of RF or SVM results with the flexibility of DL models. Moreover, enhancing the adversarial robustness of these models remains paramount, as advanced forms of malicious activities target AI-incorporated systems. The proposed models will expand the scope of evaluation to encompass more realistic and diverse threats and actual attack scenarios,

thereby contributing to the next generation of cybersecurity assessment frameworks and architectures.

References

Abbas, T. N. A., Hameed, R., Kadhim, A. A., and Qasim, N. H. (2024). Artificial intelligence and criminal liability: exploring the legal implications of ai-enabled crimes. *Encuentros. Revista de Ciencias Humanas, Teoría Social y Pensamiento Crítico.,* (22), 140-159. https://doi.org/:10.5281/zenodo.13386675

Adel, A., Mohammed, A., Daoguo, Y., and Abdulrahman, A. (2023). Advanced techniques for cyber threat intelligence-based APT detection and mitigation in cloud environments. *Proc. SPIE*. https://doi.org/:10.1117/12.2681627.

Ali, S., Abusabha, O., Ali, F., Imran, M., and Abuhmed, T. (2023). Effective Multitask Deep Learning for IoT Malware Detection and Identification Using Behavioral Traffic Analysis. *IEEE Transactions on Network and Service Management,* 20 (2), 1199-1209. https://doi.org/:10.1109/TNSM.2022.3200741

Almutlaq, S., Derhab, A., Hassan, M. M., and Kaur, K. (2023). Two-Stage Intrusion Detection System in Intelligent Transportation Systems Using Rule Extraction Methods From Deep Neural Networks. *IEEE Transactions on Intelligent Transportation Systems,* 24 (12), 15687-15701. https://doi.org/:10.1109/TITS.2022.3202869

Alzahrani, A., and Aldhyani, T. H. H. (2022). Artificial Intelligence Algorithms for Detecting and Classifying MQTT Protocol Internet of Things Attacks. *Electronics*, 11 (22). https://doi.org/:10.3390/electronics11223837.

Bhusal, D., Shin, R., Shewale, A. A., Veerabhadran, M. K. M., Clifford, M., Rampazzi, S., and Rastogi, N. (2023). SoK: Modeling Explainability in Security Analytics for Interpretability, Trustworthiness, and Usability. *Proceedings of the 18th International Conference on Availability, Reliability and Security*, Benevento, Italy. https://doi.org/:10.1145/3600160.3600193

Dong, M., Yao, L., Wang, X., Benatallah, B., Zhang, S., and Sheng, Q. Z. (2023). Gradient Boosted Neural Decision Forest. *IEEE Transactions on Services Computing,* 16 (1), 330-342. https://doi.org/:10.1109/TSC.2021.3133673

Et al., N. K. (2023). AI in Cybersecurity: Threat Detection and Response with Machine Learning. *Tuijin Jishu/Journal of Propulsion Technology,* 44, 38–46. https://doi.org/:10.52783/tjjpt.v44.i3.237

He, N., Zhang, Z., Wang, X., and Gao, T. (2023). Efficient Privacy-Preserving Federated Deep Learning for Network Intrusion of Industrial IoT. *International Journal of Intelligent Systems,* 2023 (1), 2956990. https://doi.org/:10.1155/2023/2956990

Lee, H. W., Han, T. H., and Lee, T. J. (2023). Reference-Based AI Decision Support for Cybersecurity. *IEEE Access,* 11, 143324-143339. https://doi.org/:10.1109/ACCESS.2023.3342868

Li, H., Wu, J., Xu, H., Li, G., and Guizani, M. (2022). Explainable Intelligence-Driven Defense Mechanism Against Advanced Persistent Threats: A Joint Edge Game and AI Approach. *IEEE Transactions on Dependable and Secure Computing,* 19 (2), 757-775. https://doi.org/:10.1109/TDSC.2021.3130944

Li, S., Chai, G., Wang, Y., Zhou, G., Li, Z., Yu, D., and Gao, R. (2023). CRSF: An Intrusion Detection Framework for Industrial Internet of Things Based on Pretrained CNN2D-RNN and SVM. *IEEE Access,* 11, 92041-92054. https://doi.org/:10.1109/ACCESS.2023.3307429

Li, Y., Wang, J., Fujiwara, T., and Ma, K.-L. (2023). Visual Analytics of Neuron Vulnerability to Adversarial Attacks on Convolutional Neural Networks. *ACM Trans. Interact. Intell. Syst.,* 13 (4), Article 20. https://doi.org/:10.1145/3587470

Nameer, Q., Aqeel, J., and Muthana, M. (2023). The Usages of Cybersecurity in Marine Communications. *Transport Development,* 3 (18). https://doi.org/:10.33082/td.2023.3-18.05

Park, C., Lee, J., Kim, Y., Park, J. G., Kim, H., and Hong, D. (2023). An Enhanced AI-Based Network Intrusion Detection System Using Generative Adversarial Networks. *IEEE Internet of Things Journal,* 10 (3), 2330-2345. https://doi.org/:10.1109/JIOT.2022.3211346

Qasim, N., Shevchenko, Y.P., and Pyliavskyi, V. (2019). Analysis of methods to improve energy efficiency of digital broadcasting. *Telecommunications and Radio Engineering,* 78 (16), 1457-1469. https://doi.org/:10.1615/TelecomRadEng.v78.i16.40

Rao Sangarsu, R. (2023). Enhancing Cyber Security Using Artificial Intelligence: A Comprehensive Approach. *International Journal of Science and Research (IJSR),* 12 (11), 8-13. https://doi.org/:10.21275/SR231029092527

Rjoub, G., Bentahar, J., Wahab, O. A., Mizouni, R., Song, A., Cohen, R., Otrok, H., et al. (2023). A Survey on Explainable Artificial Intelligence for Cybersecurity. *IEEE Transactions on Network and Service Management,* 20 (4), 5115-5140. https://doi.org/:10.1109/TNSM.2023.3282740

Sagu, A., Gill, N. S., Gulia, P., Singh, P. K., and Hong, W.-C. (2023). Design of Metaheuristic Optimization Algorithms for Deep Learning Model for Secure IoT Environment. *Sustainability*, 15 (3). https://doi.org/:10.3390/su15032204.

Salloum, S., Gaber, T., Vadera, S., and Shaalan, K. (2022). A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques. *IEEE Access,* 10, 65703-65727. https://doi.org/:10.1109/ACCESS.2022.3183083

Sauka, K., Shin, G.-Y., Kim, D.-W., and Han, M.-M. (2022). Adversarial Robust and Explainable Network Intrusion Detection Systems Based on Deep Learning. *Applied Sciences*, 12 (13). https://doi.org/:10.3390/app12136451.

Shanthi, R. R., Sasi, N. K., and Gouthaman, P. (2023). A New Era of Cybersecurity: The Influence of Artificial Intelligence. *2023 International Conference on Networking and Communications (ICNWC),* 5-6 April. https://doi.org/:10.1109/ICNWC57852.2023.10127453.

Shashkov, A., Hemberg, E., Tulla, M., and O'Reilly, U.-M. (2023). Adversarial agent-learning for cybersecurity: a comparison of algorithms. *The Knowledge Engineering Review,* 38, e3. https://doi.org/:10.1017/S0269888923000012

Sugumaran, D., John, Y. M. M., C, J. S. M., Joshi, K., Manikandan, G., and Jakka, G. (2023). Cyber Defence Based on Artificial Intelligence and Neural Network Model in Cybersecurity. *2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM),* 6-7 April. https://doi.org/:10.1109/ICONSTEM56934.2023.10142590.

Yousif, O., Dawood, M., Jassem, F. T., and Qasim, N. H. (2024). Curbing crypto deception: evaluating risks, mitigating practices and regulatory measures for preventing fraudulent transactions in the middle east. *Encuentros: Revista de Ciencias Humanas, Teoría Social y Pensamiento Crítico,* (22), 311-334. https://doi.org/:10.5281/zenodo.13732337