Journal of Applied Engineering and Technological Science
Vol 7(1) 2025:460-482

ENHANCING CYBERSECURITY THREAT DETECTION USING MACHINE LEARNING: A COMPREHENSIVE REVIEW

P. Somasundari[1*], V. Kavitha[2]
Assistant Professor, Department of Computer Science and Engineering, Rajalakshmi Institute of Technology, Chennai, Tamil Nadu, India[1]
Professor, Department of Computer Science and Engineering, University College of Engineering, Kancheepuram, Tamil Nadu, India[2]
plsomasundari@gmail.com[1], kavinayav@gmail.com[2]

Received: 24 July 2025, Revised: 06 November 2025, Accepted: 13 November 2025
**Corresponding Author*

ABSTRACT
Cybersecurity forms the backbone of digital infrastructure that protects overstretched payment systems, governmental operations, and business continuity today. With machine learning (ML) techniques, it can help analyze a large amount of data and improve cyber-security. It's tough to quantify how effective the ML-based cybersecurity system is, especially when we theorize it. This review paper talks about the significant role of ML in security, threat detection and security measures. Using machine learning algorithms helps in cybersecurity as they make the system automatic and fast. We can implement a threat detection security model using widely used ML algorithms. For classification purposes, we have Support Vector Machines (SVM), Decision Trees (DT), Random forests (RF), and Adaptive and Extreme gradient boosting (XGBoost). This review paper proposes ML algorithms for the implementation of cybersecurity with some practical application demonstrations. Machine learning algorithms can provide valuable analytics to help bolster security and reduce threats. We assess the accuracy of threat detection in network security by utilizing a set of formulas based on confusion, recall, F1-score, time complexity, accuracy and precision. This review synthesizes algorithmic performance across benchmark datasets (CICIDS2017 NSL-KDD UNSW-NB15) to identify significant gaps in previous ML-based cybersecurity frameworks. The results demonstrate the superior precision (90. 8 percent) and scalability of XGBoost.

Keywords: *Cyber security, threat detection, machine learning, Adaptive Boosting, XGBoost, SVM, RF and accuracy.*

1. Introduction

Distributed energy resources (DERs) refer to various data analytics related to energy use and performance (Okoli et al., 2024). We must make sure that the network connections are in order and scattered. Artificial intelligence (AI) and machine learning (ML) are being used more to enhance cybersecurity through automated threat detection and response. Algorithms like SVM Random Forest and XGBoost are better at identifying anomalies and intrusion patterns than traditional systems. However, problems with current ML-based approaches still exist such as high computational cost limited adaptability to changing attacks and lack of interpretability (Katiyar et al., 2024). Cyber-physical security ensures a secure level of performance, maintaining high runtime standards. It prioritizes reliability and fosters trust, enhancing efficiency in energy practices. The cyber threat is significantly affecting financial issues for individuals, and more economic aspects of the process. To resolve the technical issues, early detection of the trends online is essential.

Explainable Artificial Intelligence (XAI) methods detect various attacks in cybersecurity issues, analyze more data, and compare the previous day's data with current performance (Al-Shehari et al., 2024). They identify unknown data to detect unpredictable values and facilitate decision-making regarding network connections and performance (Yeboah-Ofori et al., 2021). The main detection in a cybersecurity process is anomalies; malware is initially verifying the detection, and testing validation of the performance is more accurate for the outcomes. The prevalent method used for final data relates to the prevention of the techniques. Additionally, digital technology faces different kinds of threats, such as cyber-attacks and spam, which impact the methods used for prevention in this process.

The network security procedure provides assurance of performance efficiency and communication. Wu and his collaborators demonstrate that every technique for gathering information on the connection network earns a high measure of trust. It also identifies misclassifications only to change the channel of the method. In the communication process, it acts as a strong sender and receiver. The suggested ML correctly classifies by balancing the stability of the data with the objective of accuracy when network problems arise (Wazid et al., 2022). It is useful for a specific analysis in data processing work, and multitasking process helps to enhance performance.

The cyber threat mainly impacts individuals and eventually creates problems in society. Unprofessional threats are increasingly happening in the field of digital technology. Moreover, threats of cyberattacks on companies and others are on the rise (Keserwani et al., 2022). Proper security must be maintained because technicians incur low reliability, and the protection processes do not benefit adequately. The implementation, recovery from the problems, and maintenance of the procedures incur high costs (Nassif et al., 2021). The management of threats and addressing the complexities and the lack of viability of the performance is crucial.

The cybersecurity of recent attacks is often uncertain, characterized by increased data mismanagement, changing channels, and misclassification of information (Hossain & Islam, 2023). Attacks and threats, which frequently occur within the span of time, impact various types of properties, resulting in losses to economic and financial processes. The presented techniques are not suitable for any kind of situation, and code implementation makes secure processes a difficult task (Al Razib et al., 2022). Power consumption is high, there are slow levels of runtime performance, and the response of the performance is not predictable. The extent of attacks is not determined, leading to financial losses for every industry, which now takes no steps in the processes suitable for the technologies.

The absence of an integrated ML-based cybersecurity review has been identified as the primary research gap. Thus, the objectives of this work are to: (1) critically analyse machine learning algorithms for cyber threat detection; (2) assess comparative performance; and (3) suggest future research directions that prioritize explainability and scalability.

2. Literature Review

Cyber threat detection is a significant concern for individuals, and most threats originate from digital technology. To address these issues, an automatic system is expected to provide value through early detection; however, it consumes more power (Ferrag et al., 2021). The proposed method is a ML approach that aims for greater accuracy, lower power consumption, and high reliability in performance. However, the process incurs high computational costs and is always data-dependent for performance.

Cyber threat detection is most often found in the people, companies, and education within the field to address Bayesian classification, as increased data analysis reduces noise filtering performance (Saheed & Arowolo, 2021). However, this technique faces a zero-frequency problem and sensitivity distribution issues. The methodology utilizing Support Vector Machine (SVM), which offers better data classification over time and enhances the reliability of the techniques. However, the process requires more time when dealing with large data sets it is costly, also requires high computation and is sensitive to kernel selection.

The Density-Based Local Outlier Factor (DBLOF) algorithm identifies data locations and sizes, focusing on the current location to produce results efficiently and enhance the detection process (Ozkan-Okay et al., 2024). However, these techniques suffer from misclassification and high dimensionality. The proposed method, a conventional detection algorithm, enables early detection, reduces misclassification, and enhances the efficiency of the outcomes. However, the process has slow processing time and high computation costs.

Cyber Supply Chain (CSC) is identifying the unknown data, and its analysis in the unpredictable variables improves testing and validation of the process, but there are reputational damages and malware attacks (Mukesh, 2025). The proposed method is Cyber Threat Intelligence (CTI), which reduces integrability, removes reputation errors, and addresses malware attacks. Though, the process is overloaded with data, raising the reliability of the techniques.

Distributed Denial of Service (DDoS) is a security against harmful data threats during the process. Also, it is usable for trouble of network traffic. Many solutions can provide the technology, but with limitations on the response time (Yaseen, 2023). The method which is democratic will bring out a better response to performance issues and reduce risk. The process happens often but with a broad range of methods.

Cyber Security uses attack tactics to identify and prevent data breaches. It uses ML for increased reliability and process stability. But it must pay exorbitant computational costs (Haider et al., 2021). The method proposed is Boosting and Bagging Algorithms which will compare the performance of the Both Current and Provided Data in terms of Computational Costs. Yet, the process is a complicated coding implementation that requires more strength.

The cyber threat detection and early workload process is a maintenance task in the secure level utilized for network security techniques. Its focus on stability helps secure performance, but the complex task involves network connections (Mohammed et al., 2024). The resolution for the methods using ML is a bit of a network connection issue, which may take multiple attempts to complete the multitasking operations. However, the process is more complicated and may require some time to perform effectively.

Cloud computing is a more data-collecting high level of security, rapidly performing its focus on the continuous monitoring of each type of data, maintaining the process. However, the method frequently encounters signal issues (Shaukat et al., 2020). The proposed method is ML that automatically detects threats and removes the unproductive value, reducing the signal issue. However, the process is more network connection loss and has greater time complexity.

The random forest's multiple data are collected and maintained securely to ensure the program's reliability yet demands significant processing time and memory and the process is computationally complex and time-consuming (Siddiqi & Pak, 2021). The proposed method is an ensemble model that is highly reliable and meets standard stability requirements. However, the process involves regression problems, and the code implementation is complex.

Software Defined Networking (SDN) is a marker in the framework of data structures and secures the main software of the network connection, centralized data, and high-level programmability (Injadat et al., 2020). However, this method presents additional potential security risks. The proposed method is a Boosting and Bagging Algorithms model, which analyzes data efficiently, achieving a high accuracy level. However, the process incurs more expense and higher power consumption.

Edge computing networks are collecting more data, focusing on each type of data protection and measures in collection, testing, and validation, but the process is more time-consuming and low reliability (Li & Yan, 2022). The proposed method is ML, which gathers network connections more accurately and provides high scalability, allowing critical situations to be handled easily. However, the process is more steps to include during the level of performance.

Information and communication technology (ICT) is different communication of the server, and receives performance, and primarily maintains secure and clear data in the threat and prevention of performance, low trustworthy range (Ye et al., 2021). The proposed method is a ML for a more data identity changing channel, and unknown data produces protection. However, the process is a more expensive, high level of complex implementation of the techniques.

Table 1 demonstrates the data secured based on ML technology using SVM and Gaussian Algorithms., the author's previous structure techniques, classification, accuracy and performance evaluation. Table 1 lists deep learning and ensemble approaches that have been documented in recent research. While CNN and LSTM successfully capture intricate attack patterns algorithms like AdaBoost and XGBoost achieve greater precision and less overfitting when compared to conventional models. Though, the deployment of these techniques in real-time cybersecurity environments is limited because they typically require large datasets and higher computational resources. High computational complexity, limited interpretability and poor adaptability are persistent challenges faced by existing ML-based cybersecurity methods despite numerous advancements. These limitations highlight the need for a unified comparative review combining algorithmic analysis and empirical benchmarking.

The revolutionary significance of machine learning, specifically, convolutional neural networks in enhancing cyber security measures is examined in this essay. The study will consider how these technologies can protect and possibly expose sensitive information, including the impact on data privacy and information protection (Chukwunweike et al., 2024). In this novel study, the ML techniques with the multi factor authentication (MFA) shall be designed to strengthen the network security. Additionally, the focus is on the network intrusion detection in the present study. It is important that energy saving and reduced environmental impact are integrated into the security system, in addition to conventional encryption and biometric techniques. It also talks about how centralized systems have some shortcomings. These include flaws in security and breakdown of the system (Mahmood et al., 2024).

Table 1 - Data Security Based on ML Technology Using SVM and Boosting and Bagging Algorithms.

Author/year	Classification	Techniques	Accuracy	Performance Evaluation
Kuppa & Le-Khac (2021)	Decision Trees	Reinforcement Learning (RL)	90%	Prediction, Sensitivity.
Arshad et al. (2022)	Boosting and Bagging Algorithms	decentralized networks	92%	Recall, FN score
Guo et al. (2021)	Adaptive ML models	Ensemble techniques	95%	Precision, validation process
Ahn et al. (2023)	SVM	ML techniques	93%	dataset evaluations
Ejiofor (2023)	Boosting and Bagging Algorithms	ML techniques	96%	economic losses
Kravchik & Shabtai, (2021)	spam classification	ML techniques	93%	Prediction level, accuracy
Ige et al. (2024)	Gaussian Classifications	LSTM	94%	cross-validation, Improves classification
Le et al. (2020)	Convolutions Neural Network (CNN)	Auto-Encoders (AE)	98%	Prediction, Sensitivity.

The model adapts with expected attack patterns using online learning in this paper. There's improved performance against various types of attacks as the dynamic feature selection function overcomes the usual limitations. The researchers tested the model against commercial models used in industry. This effort lays a strong foundation upon which proactive threat identification and mitigation in environments can be established through the reinforcing of network security on evolving cyber risks. Such strategies are useful in detecting and stopping cyber warfare that can wreak havoc for individuals, enterprises, and even entire countries.

Security experts can find threatening indicators that have not been discovered before using a machine learning algorithm that uses statistics to identify patterns and deviations in large datasets. It also discusses the shortcomings and challenges of different approaches such as adversarial attacks, interpretability problems, and data quality (Ozkan-Okay et al., 2024).

We aim to tackle high dimensionality challenges in intrusion detection and enhance classifier classification performance, which will lead to better and more effective intrusion detection ultimately. The NSL-KDD data set, a popular benchmark in this field, is employed in our research to do this. The J48 tree has the greatest reported accuracy of 79.1% of the classifiers examined (Nabi & Zhou, 2024). The author proposed creating AI-driven compliance frameworks to solve ethical issues, enhancing interpretability in deep learning models, and using self-supervised learning for fraud detection. This work offers a unique method for protecting Bitcoin transactions by combining ML and Reinforcement Learning (RL), providing researchers, financial institutions, and policymakers with useful information (Olutimehin, 2025).

The author concentrates on applying random forest, decision tree classifier, ensemble, long short-term memory, and convolutional neural network models on the Iot23 dataset within the framework of a collaborative threat intelligence framework for Iot security. The study examines privacy concerns, implementation specifics, and the smooth incorporation of machine learning-based methods for ongoing model enhancement. Tests conducted on the Iot23 dataset show how well the suggested solution works to improve Iot security and reduce possible risks (Nazir et al., 2024). The difficulties presented by adversarial ML in the context of network security along with potential solutions. The ever-changing nature of network environments and security systems that may have limited resources. According to Khan and Ghafoor (2024), this supports efforts to strengthen security systems based on machine learning against threats.

The author suggested that ML is necessary for better cloud security. Using AI-driven techniques, security systems can notice patterns, anomalies and threats in huge datasets. Machine learning algorithms can predict an attack that is yet to occur and prepare better defenses for it by learning from previous attack data. Artificial intelligence improves safeguards for identity management through authentication and access control solutions to minimize unauthorized access and data breaches. (Mamidi, 2024). The author proposed that ML are essential. Looking at previously seen attack data, the ML algorithms can anticipate future attacks and design better countermeasures. AI-supported access control and authentication systems enhance identity management by decreasing chances of data breaches and unauthorized access (Ekundayo et al. 2024).

The author investigated how Adversarial ML and Artificial Intelligence (AI) techniques can be effective against growing cyber-attacks. This method is maintaining cyber security tools a step ahead of any attackers by adapting to the ever-changing digital landscape. Safeguarding one’s digital property and ensuring the integrity of networks will call for the application of AI and adversarial ML in cyber security techniques. This is because of changing complex cyber threats (Ijiga et al. 2024). We seek to develop reliable network intrusion detection tools by means of ML approaches. In our work, feature selection, data normalization, standardization, hyper parameter tweaking are used as model optimization techniques. The results highlight the efficacy of the Random Forest Classifier (0.97) and commend the use of a variety of datasets and “modern optimization procedures” (Tendikov et al., 2024).

Table 2 - presents a comparative overview of recent studies applying machine learning.

Author Name	Methods	Methods Used	Drawbacks
Almotairi et al. (2024)	ML	K-Best algorithm	Despite its high performance, the proposed model may face scalability challenges with real-time intrusion detection in resource-constrained IoT environments.
Vashishth et al. (2024)	ML and AI	KNN	AI and ML-based solutions that work in the cloud and secure cloud data can be resource-hungry or introduce latency in detecting threats that happen in real-time.
Vaddadi et al. (2023)	AI and ML	SVM	Even if the AI and ML models perform well, they may not be able to detect zero-day attacks and adversarial inputs.
Ahsan et al. (2021)	ML	Bidirectional Long Short-Term Memory	Additionally, advancements made on benchmark datasets might not apply to tougher situations with noisier or evolving data.
Dhaiya et al. (2021)	ML	Genetic Algorithm	Integrating complex systems increases complex overheads and needs specialized knowledge to deploy them effectively and maintain their performance.

Tulli, (2023)	ML	ML Model	Though useful, ML in marketing and finance might inaccurately predict or unfairly segment us due to biased or incomplete data.
Agarwal et al. (2021)	ML	SVM, KNN, and NB	Using traditional machine learning models such as NB, SVM, KNN, etc. can restrict scalability and elasticity to complex and evolving cyber threats. Moreover, the effectiveness of the method largely relies on the features' quality. Further, it does not generalize well to real network traffic variability.

Table 2 compares various recent studies related to using ML and artificial intelligence in cyber security and other technological fields. A summary of title, methods used and drawbacks identified for each study. While using machine learning increase accuracy and efficiency, most importantly it cannot monitor and address the challenges of scalability, evolving threats, and real-time adaptability. This table aims to balance performance and practicality through a wide set of options. The traditional machine learning methods utilized in cybersecurity are listed in Table 2. While models such as SVM and KNN provide respectable accuracy they require a significant amount of computation time and are sensitive to parameter changes. Though they provide greater stability and interpretability, Random Forest and Decision Tree perform worse on large or unbalanced datasets. When all is said and done, classical models are not very flexible or scalable to changing threats.

3. Advance Techniques in Cybersecurity

Utilizing Machine Learning techniques can predict and detect security threats at an early stage. It facilitates quick responses to problems and prevents them from growing. We must first estimate the threat to effectively deal with the attacks using these techniques. But the timing of the various attacks is known.

3.1 Anomaly Detection

Unsupervised learning is effectively used to detect anomalies, identify unknown data, and match values. The classification employed in the SVM is an early detection method in the misclassification process, aimed at analyzing the change channel and decoding the techniques of the attacks.

3.2 Intrusion Detection

The data is collected, classified, and analyzed for training purposes to measure performance during testing. Its use of Random Forest is a measurement of the team data, and testing validates the performance.

3.2 Malware Detection

The coding serves as a pattern for the code signature, focusing on behavior analysis of its data regarding performance and identifying malicious activities. The signature analysis of the code implementation is detected in the attacks during the verification and testing validation of the process.

The Figure 1 is an evaluation in the performance, and the accuracy of the process, and focusing the behavior analysis. The collected data is used in the more secure network connection of the techniques.

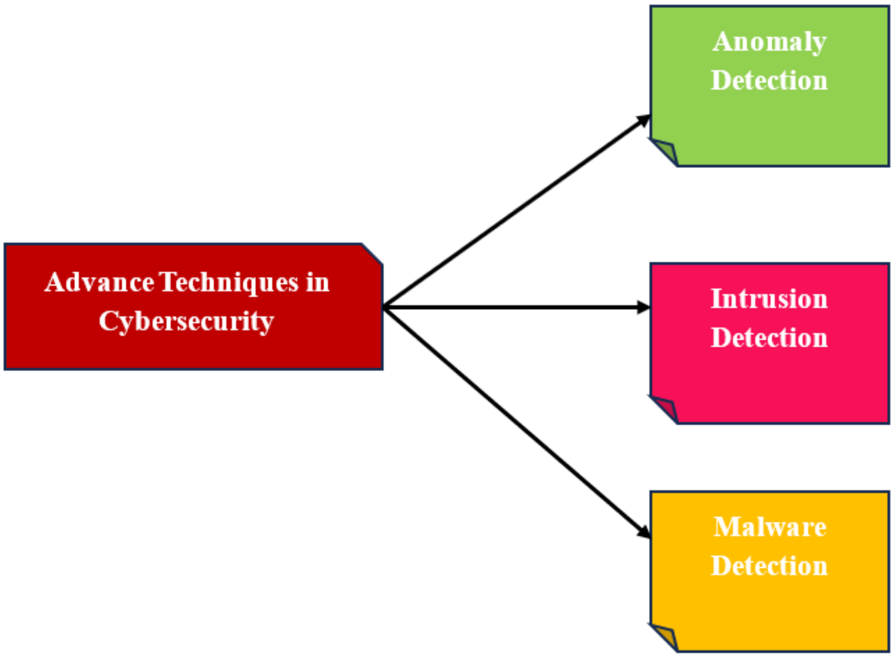


Fig. 1. Advance Techniques in Cybersecurity.

4. Cybersecurity

Cybersecurity is a critical issue in the current landscape of digital technology, impacting individuals, companies, industries, and other organizations. It enhances security by improving performance and preventing threats and attacks while also predicting potential issues in the process.

4.1 Confidentiality

The aim of confidentiality is to stop unauthorized individuals, organizations or systems from accessing and/or disclosing information. It focuses on maintaining security, reliability, and high level of efficiency.

4.2 Integrity

Integrity is used to prevent any unauthorized modification or destruction of information, ensuring greater scalability and reducing the lack of transparency in the process.

5. Testing method in cyber-Security

The cybersecurity team is testing security audits, risk assessments, and ethical hacking, focusing on protection testing and performance validation. The evaluation in the performance, and the accuracy of the process, and focusing on the behavior analysis.

5.1 Penetration Testing

Identify the vulnerabilities, and real-world cyber is a measure, and assess the security level of the process, as it is one of the testing and validation methods for the process.

5.2 Security Audits

The testing process is a maintenance task focused on ensuring security and reliability, standard scalability, and assessing performance range, while concentrating on the high level of data collected from the techniques.

5.3 Ethical Hacking

The testing method is employed for the papers, possibly to alter the secondary stage in the coding implementation process. It offers greater security within the high level of performance range, protecting the ethical concerns associated with the techniques.

5.4 Network Security Testing

The network connection of the process is used for each type to maintain the reliability and trustworthiness of the processes. The collected data is used in the more secure network connection of the techniques.

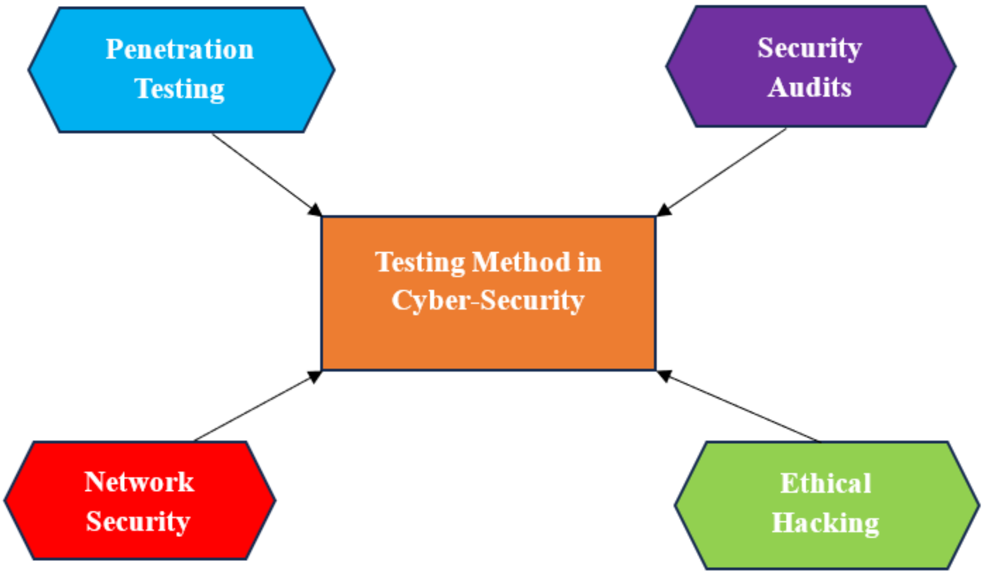


Fig. 2. Testing method in cyber-Security.

The figure 2 is a enhances security by improving performance and preventing threats and attacks while also predicting potential issues in the process. Performance is automatically predicted, allowing for early detection of malware attacks

6. Security Techniques in Cyber Threat Detection

Cyber threat detection is a low for the security of the process because of the various attacks and different threat detection methods, as well as the low range of the detection and different communication of the performance.

6.1 Signature-Based Detection

The signature identifies an unknown attacker that disrupts network and traffic performance and detects malicious activity during this process. Continuous measurement is performed in the background to monitor the techniques constantly.

6.2 Anomaly-Based Detection

The more effectively organizations detect novel or zero-day threats, the more secure their network infrastructure becomes, and the more effectively they identify anomalies. Performance is automatically predicted, allowing for early detection of malware attacks.

7. Cyber Security in Machine Learning

The cyber threat detection level is enhancing performance and elevating the ninth security level while consistently maintaining the reliability of the techniques. It's an early prediction in the initial stage of the cyber threat detection process.

7.1 Endpoint Protection

The prediction of unknown data accurately forecasts potential risks and is resolved for the near end of the performance. The calculations involve testing and validation of the process, as well as prevention of each type of data network secure connection for a performance.

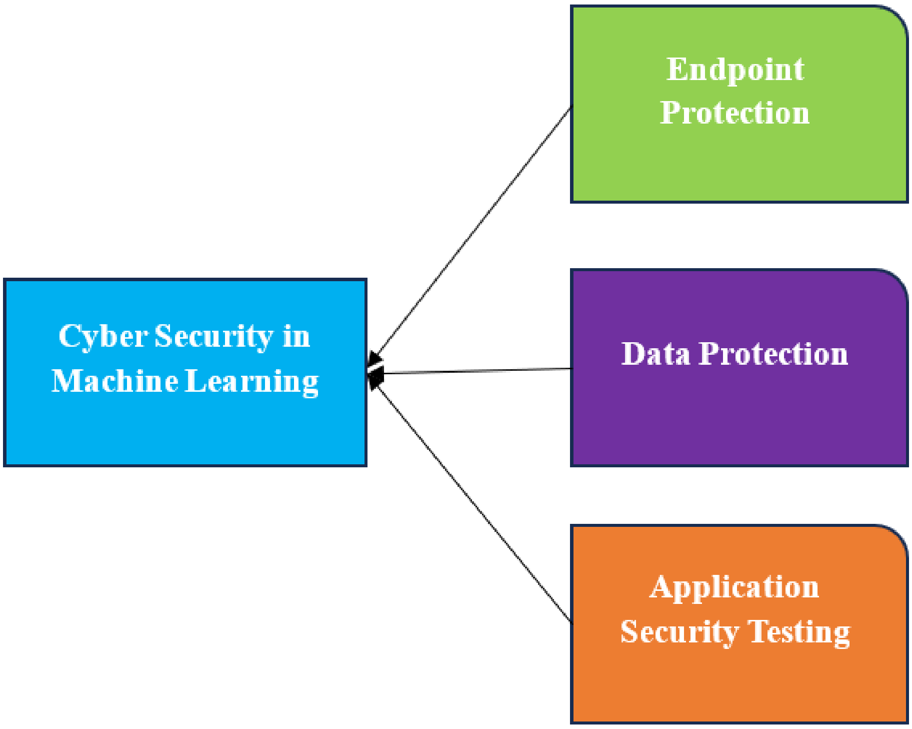

Endpoint Protection
Cyber Security in Machine Learning
Data Protection
Application Security Testing

Fig. 3. Cyber Security in ML.

Figure 3 is a cyber-attack and threats in live operations, every individual's behavior during the process is analyzed. Continuous measurement is performed in the background to monitor the techniques constantly.

7.2 Data Protection

The data is collected for the network connection of the process and product-to-cloud data on performance, as well as for conducting assessments. This enables organizations to defend against cyberattacks preemptively, as well as threats such as insecure storage, data leaks, and weak authentication.

7.3 Application Security Testing (AST)

The information assesses the security investigation and forecasts diverse attacks and cyberthreats affecting performance. For counteracting cyber-attacks and threats during live operations, analysis of individualistic behaviour in the course is taken up.

8. Methodology

This presentation constitutes an effective cybersecurity threat detection framework improvement implementation. Our architecture aims at better identification of threats than existing architecture. This study makes use of advanced machine learning techniques, particularly XGBoost technology, for better detection performance of these threats. XGBoost helps systems more accurately and timely detect attacks from more complex data patterns and anomalies. The research gives ways which will help in maintaining detection of the sophisticated cyber-attack.

As shown in Figure 4, detection threat security model which use already established ML algorithms. The working of this model has been derived from various popular ML techniques. SVM is a classification method and DT is another classification method that gives you a clear decision. Furthermore, the model includes Random Forest (RF), which improves accuracy and robustness through ensemble learning, as well as Adaptive Boosting, a technique that continuously adds weak learners, as well as XGBoost, a very efficient and scalable gradient boosting algorithm. This paper describes the use of ML algorithms to improve cybersecurity protection and outlines how these algorithms can be effectively put into practice. Notably, the paper also highlights the specific circumstances in which ML techniques are applied to combat cybersecurity attacks. ML Algorithms can help in getting meaningful analysis of data security. To strengthen security and reduce vulnerabilities across their infrastructure, organizations and companies can use this proposed approach.

8.1 Research Design Overview

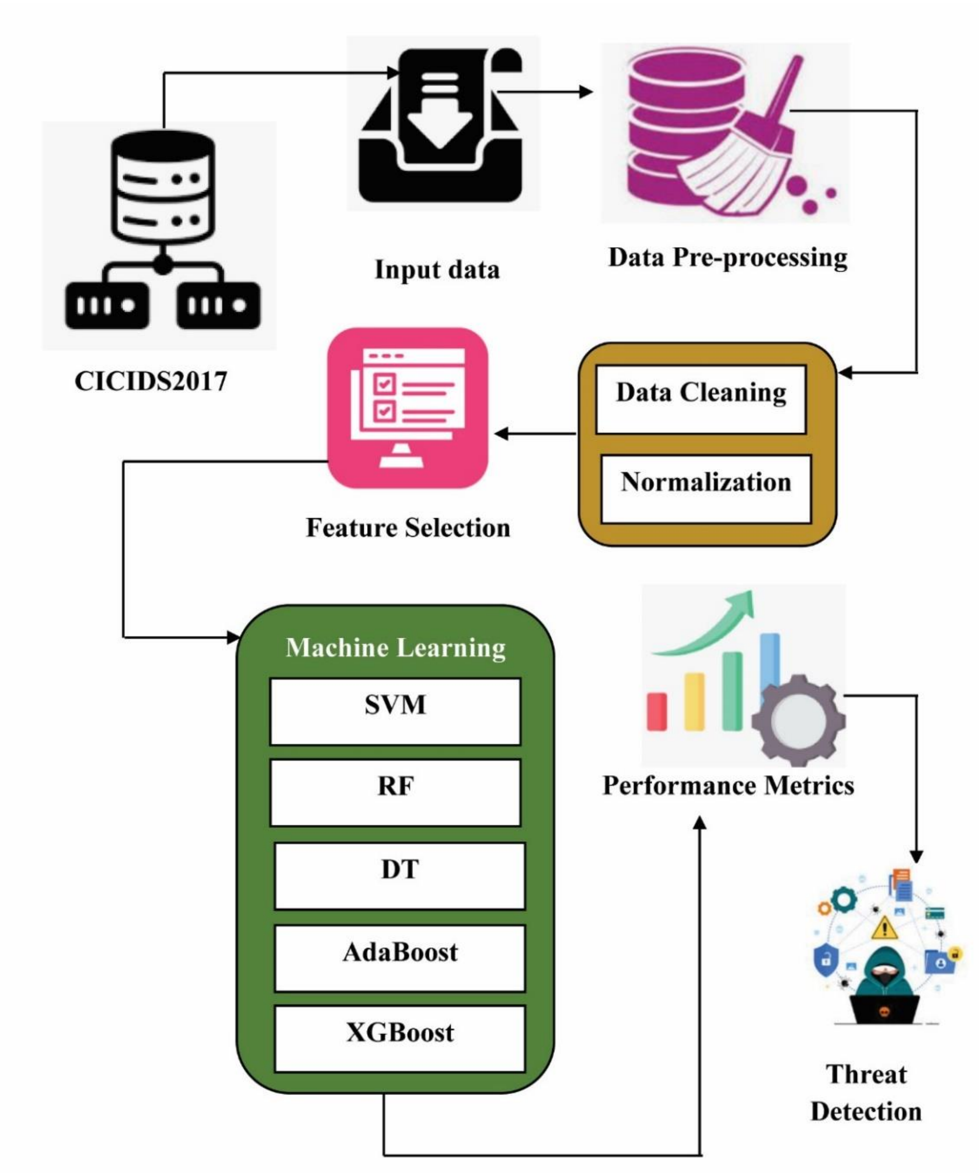


Fig. 4. The Proposed Architecture Diagram based on Cyber Security.

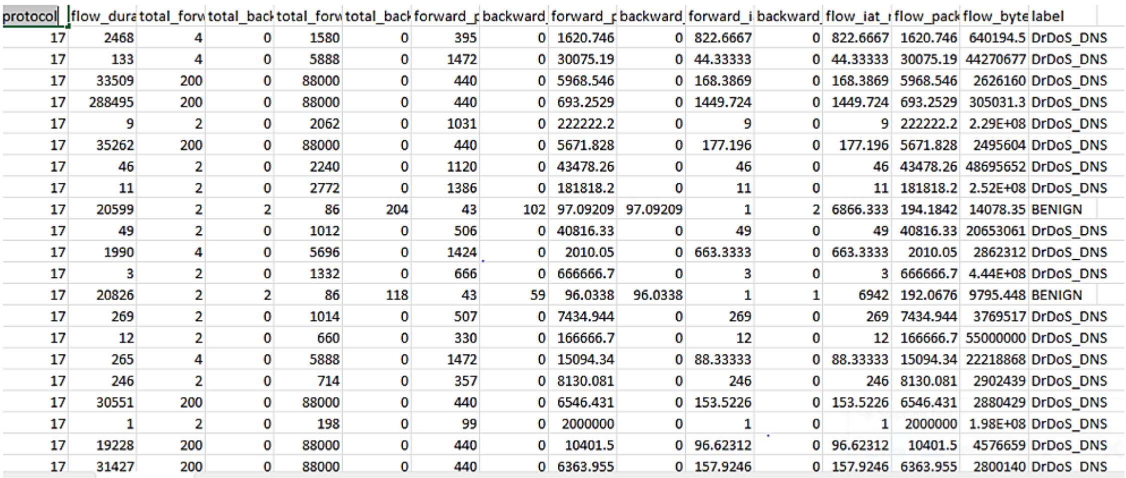
A hybrid research design that combines systematic review and experimental evaluation is used in this study. During the review phase studies from 2019 to 2025 were gathered and examined from databases like IEEE Xplore Scopus and SpringerLink. Peer-reviewed publications pertaining to ML-based cybersecurity threat detection were the focus of the selection process which adhered to established inclusion and exclusion criteria. Using the CICIDS2017 dataset five popular algorithms Decision Tree (DT) Random Forest (RF) Support Vector Machine (SVM)

AdaBoost and XGBoost were implemented during the experimental phase. Accuracy precision recall F1-score and time complexity were used as evaluation metrics to compare performance through systematic data preprocessing feature selection training and testing. The hybrid design allows for both an empirical validation of the relative efficacy of current approaches and a thorough understanding of them.

8.2 Dataset Collection

We examine intrusion detection CICIDS2017 dataset from a cybersecurity company in this section. The records in the dataset are labelled for research of network intrusion detection. In the same way and for other days, data flow and activity are normal allowing for the calculation of data points from different attacks with benign data. The CICIDS2017 dataset is used in this work. Approximately 2.8 m data is available for this dataset. It contains eighty-five attributes and they are labelled.

As illustrated in Figure 5, different types of attacks are manually classified and analyzed to identify and categorize various system hacking attacks. The methodology used in this study was a hybrid of structured review and experimentation. The experimental design was informed by the screening of fifty pertinent papers. The dataset was separated into subsets for testing (30 percent) and training (70 percent). To optimize the model the following hyperparameters were used: n_estimators = 200 max_depth = 6 and learning rate = 0. 1. Five-fold cross-validation was used to assess the model's performance to guarantee its generalizability and robustness.



protocol	flow_dura	total_forw	total_back	total_forw	total_back	forward_p	backward_	forward_p	backward_	forward_i	backward_	flow_iat_	flow_pack	flow_byte	label
17	2468	4	0	1580	0	395	0	1620.746	0	822.6667	0	822.6667	1620.746	640194.5	DrDoS_DNS
17	133	4	0	5888	0	1472	0	30075.19	0	44.33333	0	44.33333	30075.19	44270677	DrDoS_DNS
17	33509	200	0	88000	0	440	0	5968.546	0	168.3869	0	168.3869	5968.546	2626160	DrDoS_DNS
17	288495	200	0	88000	0	440	0	693.2529	0	1449.724	0	1449.724	693.2529	305031.3	DrDoS_DNS
17	9	2	0	2062	0	1031	0	222222.2	0	9	0	9	222222.2	2.29E+08	DrDoS_DNS
17	35262	200	0	88000	0	440	0	5671.828	0	177.196	0	177.196	5671.828	2495604	DrDoS_DNS
17	46	2	0	2240	0	1120	0	43478.26	0	46	0	46	43478.26	48695652	DrDoS_DNS
17	11	2	0	2772	0	1386	0	181818.2	0	11	0	11	181818.2	2.52E+08	DrDoS_DNS
17	20599	2	2	86	204	43	102	97.09209	97.09209	1	2	6866.333	194.1842	14078.35	BENIGN
17	49	2	0	1012	0	506	0	40816.33	0	49	0	49	40816.33	20653061	DrDoS_DNS
17	1990	4	0	5696	0	1424	0	2010.05	0	663.3333	0	663.3333	2010.05	2862312	DrDoS_DNS
17	3	2	0	1332	0	666	0	666666.7	0	3	0	3	666666.7	4.44E+08	DrDoS_DNS
17	20826	2	2	86	118	43	59	96.0338	96.0338	1	1	6942	192.0676	9795.448	BENIGN
17	269	2	0	1014	0	507	0	7434.944	0	269	0	269	7434.944	3769517	DrDoS_DNS
17	12	2	0	660	0	330	0	166666.7	0	12	0	12	166666.7	55000000	DrDoS_DNS
17	265	4	0	5888	0	1472	0	15094.34	0	88.33333	0	88.33333	15094.34	22218868	DrDoS_DNS
17	246	2	0	714	0	357	0	8130.081	0	246	0	246	8130.081	2902439	DrDoS_DNS
17	30551	200	0	88000	0	440	0	6546.431	0	153.5226	0	153.5226	6546.431	2880429	DrDoS_DNS
17	1	2	0	198	0	99	0	2000000	0	1	0	1	2000000	1.98E+08	DrDoS_DNS
17	19228	200	0	88000	0	440	0	10401.5	0	96.62312	0	96.62312	10401.5	4576659	DrDoS_DNS
17	31427	200	0	88000	0	440	0	6363.955	0	157.9246	0	157.9246	6363.955	2800140	DrDoS DNS

Fig. 5. Dataset Feature Selection.

8.3 Data Pre-Processing

This section describes the data in the dataset, cleans it using preprocessing, detects and resolves missing values, analyses and imputes missing values using statistical techniques, or removes cases and features with missing values. Normalization or standardization converts data into a standard scale that is sensitive to the size of the variables, allowing for the comparison of variables of varied sizes. Data normalization methods like min-max scaling and Z-score normalization standardize numerical features to a consistent scale, preventing any single feature from distorting your analysis. Techniques such as mean subtraction and scaling to unit variance ensure features have a zero mean and unit variance, which benefits certain ML algorithms.

As shown in Equation 1, normalization can reduce the training time because all the data collected from the dataset and used in training are of the same size. Calculate the maximum and minimum values of each feature with a normal range of 0 and 1. Let's assume X – input value, X_{Nor} –normalization data, X_{min} –minimize data, X_{max} –maximize data.

$$X_{Nor} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Values of constant data attributes are defined based on network security during data preprocessing, with the goal of minimizing information loss in the data.

8.4 Linear Correlation Algorithm

The equation calculates various systems for hacking attacks and measures the prediction level of attacks, let's assume the p, q- input variable, s- linear correlation.

$$-\sum_{x\in X}\sum_{x\in X}(p,q)\log\frac{s(p,q)}{s(p)s(q)} \tag{2}$$

The equation for Backdoor and Trojan attacks is an example of scanning attacks, and it assesses the security of performance and the prediction level in the detection process. Let assume TA= Trojan Attacks, FP=Functional prediction

$$\frac{TA}{TA+FP} \tag{3}$$

The different types of attacks include cross-site scripting, DDoS, brute force, and injection attacks, which affect performance, security, and reliability of the process. Let assume FQ=Functional Quality.

$$\frac{TA}{TA+FQ} \tag{4}$$

The equation determines normal activity and harmless data, while the logs from the other days also display data points from various attacks and harmless data. Let assume FQ=Functional Quality.

$$\frac{TA}{TA+FP+TQ+FQ} \tag{5}$$

8.5 Machine Learning Algorithms and Parameters

This section uses various ML classification techniques and an XGBoost-based model to detect and analyse cybersecurity threats.

8.5.1 Decision Tree (DT)

This section discusses that the advanced classification DT technique is commonly implemented in various application areas. The DT algorithm is a nonlinear supervised learning method that decomposes the security data into smaller subsets and increases the relevant branches of the tree as below equation 6.

$$E: H(x) = -\sum_{i-1}^{n} p(x_i)\log_2 p(x_i)$$

(6)

Common metrics among them are "Gini" (Gini coefficients) and "Entropy" (information gain). Let's assume, E – entropy, G –gini, pi^2 _ probability of classifying an element as a specific anomaly.

8.5.2 Random Forest (RF)

The decision classifier forest model comprises multiple decision trees, utilizing the RF algorithm, which is widely employed in various applications. Furthermore, it combines clustering (packing) and random selection operations to create a set of controlled, distributed decision trees. The results were measured using majority voting of the decision trees generated by the forest model, as below equation 7.

$$G(E) = 1 - \sum_{i=1}^{c} pi^2$$

(7)

The quality of tree splitting is measured by the "Gini" of a decision tree constructed using a stochastic forest security model.

8.5.3 Support Vector Machine (SVM)

This section utilized ML technology based on the SVM algorithm to provide network protection by creating hyperplanes between data spaces. The security dataset is evaluated using a kernel function, which is categorized as either normal or abnormal. There are several types, such as linear, non-linear, RBF, and sigmoid. where λ –parameter that sets the spread, K – kernel.

$$K(x,y) = exp\left(-\lambda\|x-y\|^{2}\right)$$

(8)

As defined in Equation 8 above, the SVM technique based on the RBF kernel function was used to estimate the given safety data and achieve the target.

8.5.4 Adaptive Boosting (AdaBoost)

Improving the model reduces bias and variance in the dataset and turns weak learner into a strong learner. The AdaBoost framework implements an adaptive classifier that improves performance. A maximum depth decision tree classifier gives accurate results by calculating the optimum value for the estimators.

8.5.5 Extreme Gradient Boosting (XGBoost)

We can optimize the weights of the neural network with the help of the XGBoost algorithm to minimize the loss function for the set of individual samples the model cycle. Furthermore, certain slope-increasing methods has studied for the most accurate approximation of the XGBoost model. Machine learning algorithm, through advanced regularization and second order gradient calculations, minimizes loss, and thus boosts generalization. This means we use XGBoost Machine learning algorithm to improve cybersecurity and a threat mitigation just like other domains where it is used.

Calculate the eigenvalue vector formed by K decision trees as shown in Equation 9. Let's assume Where $\hat{y}$ –ground truth vector.

$$\hat{y} = \sum_{k=1}^{k} f_k\,(x_i), f_k \in \mathcal{F}$$

(9)

As shown in Equation 10, the score vector on the corresponding leaf is estimated based on the number of leaves and the data points. Let's assume conventional training loss function, Ω –Regularization prevents.

$$L^{(t)} = \sum_{i=1}^{n} l(y_i - \hat{y}_i) + \sum_{k+1}^{k} \Omega\,(f_k)$$

(10)

Calculates the gain split on the left and right at a given leaf node, as illustrated in equation 11. Let's assume L, R –left and right, G –gain

$$G = \frac{1}{2}\left(\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} + \frac{G_L^2+G_R^2}{H_L+H_R+\lambda}\right) - \lambda$$

(11)

This section utilized the XGBoost algorithm to mitigate cybersecurity threats with a positive impact.

9. Result and Discussion

This evaluates the performance of various ML models in terms of accuracy, precision, recall, F1-score and time complexity on different sizes of datasets to detect the cyber security threat. Implemented Models are Naïve Bayes, HHO-SVM, KNN, and XGBoost. The research applies 33,926 points of data from a cybersecurity attack dataset for robust and reliable threat detection.

Table 3 - Simulation Parameter.

Simulation	Value
Dataset Name	CICIDS2017
Number of Records	33926
Language	Python
Tool	Jupyter
Training	24576
Testing	9,350

The simulation parameters were significantly improved (shown in Table 3). The evaluation was done using test-and-train manner. Kaggle is a well-known open-source platform where data scientists collect any amount of data to train their Artificial Intelligence programs. The CICIDS2017 dataset is hosted at Kaggle. So, the XGBoost based ML methods were used with the dataset to see their performance.

These are some of the metrics that are used to evaluate the performance of the model, including precision, accuracy, recall, and F1 score. The proposed approach allows determination of this set of metrics using performance metrics ML methods. To be specific, Table 4 shows how these methods have been applied in relation to confusion measures.

Table 4 - Confusion Metrics.

Metrics	Formula
Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$
Precision (Pre)	$\frac{TP}{TP+FP}$
Recall (Rec)	$\frac{TP}{TP+FN}$
F1-Score	$2*\frac{Pre*Rec}{Pre+Rec}$

Table 5 - Comparison and Proposed Methods of Accuracy Prediction.

Methods	UNSW-NB15		NSL-KDD		CICIDS2017	
	Train	Test	Train	Test	Train	Test
Naive Bayes (Sarker, 2021)	128552	46789	15683	6906	19,070	14856
Harris Hawks Optimization-SVM (HHO-SVM) (Kaur et al., 2024)	139694	35647	20149	2398	21,470	12456
Logistic regression (LR) (Babagana et al., 2024)	160552	14789	19785	2759	23,470	10456
K-Nearest Neighbours (KNN) (Wang et al., 2024)	99520	75821	17541	5003	18,294	15632
Particle Swarm Optimization and SVM (PSO + SVM) (Yadav et al., 2024)	113686	58973	14789	7755	22,668	11258
SVM (Musa et al., 2024)	10456	44789	17896	5647	21486	10456

For measuring and comparing prediction accuracy along with the proposed methods and data sets (Table 5).

Table 6 - Performance of Precision.

Number of Records	Naive Bayes	HHO-SVM	KNN	XGBoost

8491	53.9	66.4	69.4	80.5
16972	54.6	65.3	70.4	83.6
25463	56.4	63.3	73.7	88.7
33926	58.4	70.5	75.3	90.8

The precision levels of the four ML models Naïve Bayes, HHO-SVM, KNN and XGBoost are compared as shown in Figure 6 and in Table 6. The models are evaluated over four datasets, whose sizes are 8491, 16972, 25463 and 33926. XGBoost always performs better in precision than other models, across all dataset sizes. Like, at 8491 records, it reaches a precision of about 81% while MI records about 54%. XGBoost is good at classifying a certain malignancy as the dataset grows, with a precision score as high as 90.8%. It can achieve this score at 33,926 records. The model performs these tasks well utilizing a large volume of input data which indicates its robustness and scalability. XGBoost's accuracy on different scales shows its suitability for detecting cybersecurity threats and reducing false positives, making it useful for protecting real-world systems.

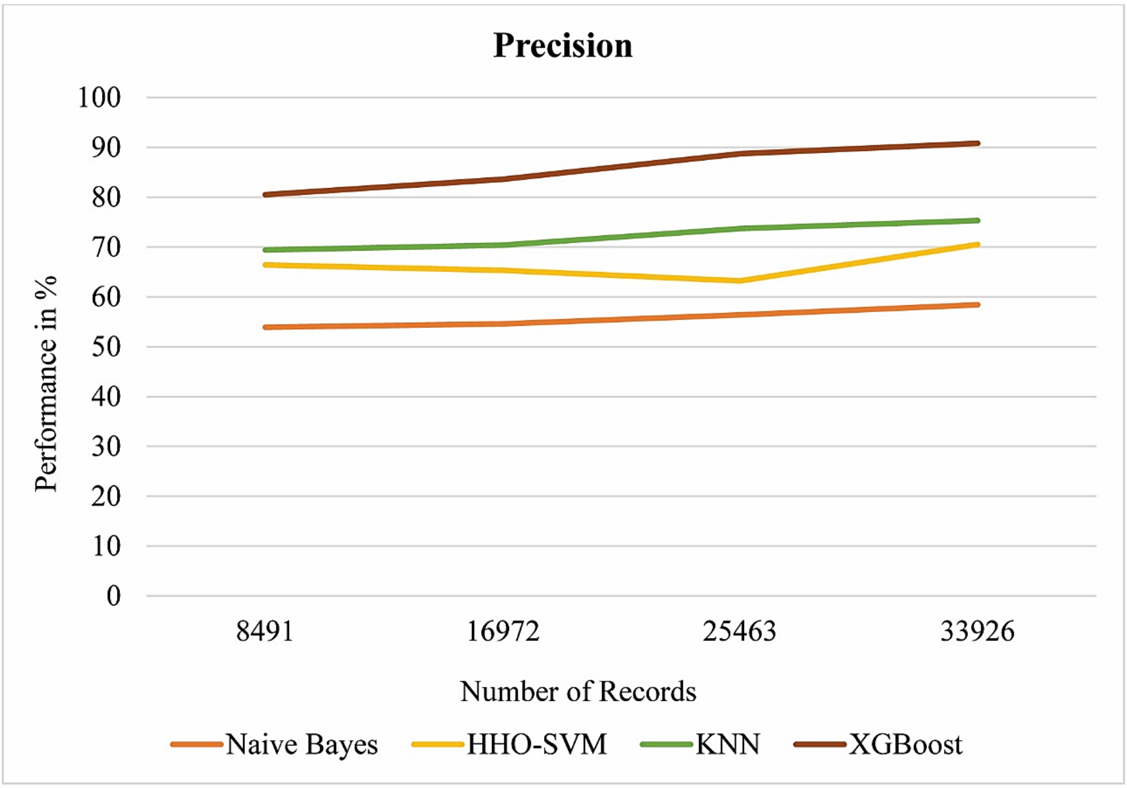


Fig. 6. Analysis of Precision.

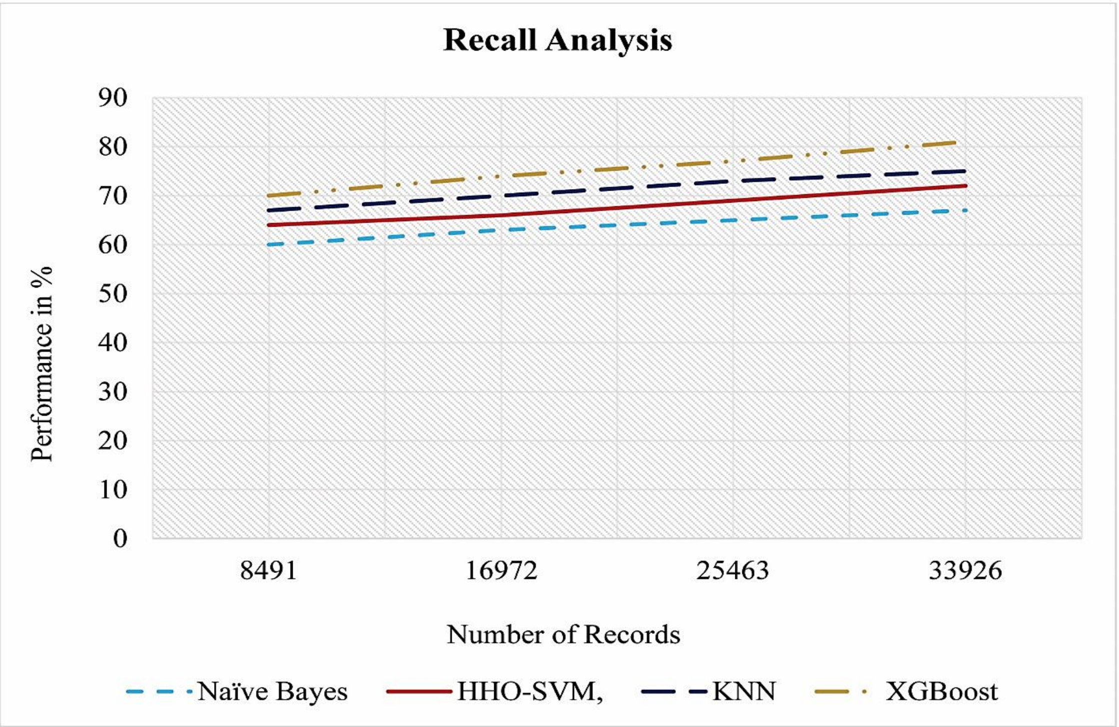


Fig. 7. Recall Performance.

The performance of recall of four ML models (Naïve Bayes, HHO-SVM, and KNN and the proposed XGBoost) against different datasets ranging from 8491 to 33926 in records is shown by the Figure7 and Table 7. Among these models, XGBoost consistently shows the maximum recall across all data scales. XGBoost achieves around 81% recall for the smallest dataset size at 8491, much higher than the value of SVM at 55%.

Table 7 - Performance of Recall.

Number of Records	Naïve Bayes	HHO-SVM,	KNN	XGBoost
8491	60	64	67	70
16972	63	66	70	74
25463	65	69	73	77
33926	67	72	75	81

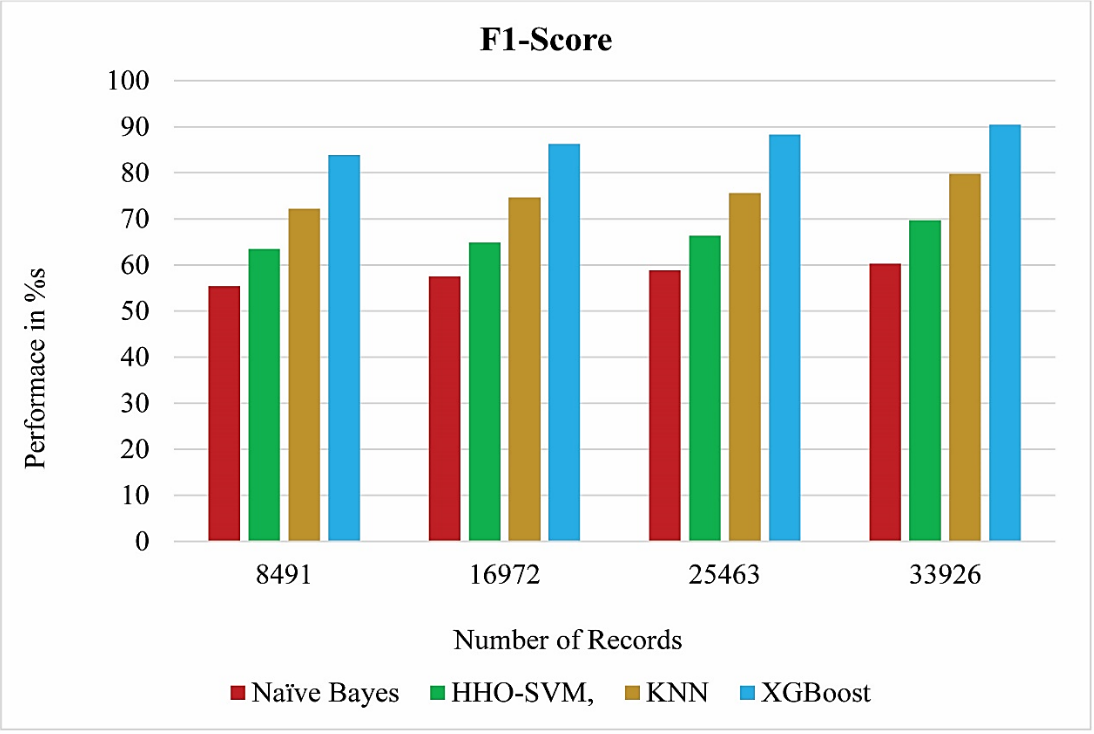


Fig. 8. Analysis of F1-score.

Table 8 - Performance of F1-Score.

Number of Records	Naïve Bayes	HHO-SVM,	KNN	XGBoost
8491	55.4	63.5	72.2	83.9
16972	57.5	64.9	74.7	86.3
25463	58.8	66.4	75.6	88.3
33926	60.3	69.69	79.8	90.5

Figure 8 and Table 8 show the F1-score performances of Naïve Bayes, HHO-SVM, and KNN, and the proposed XGBoost on various dataset records. The F1-score is used to summarize the performance of a classifier, especially when class distribution is uneven. In cyber threat detection, both false positives and false negatives must be immediately minimized. At all data points, the highest F1-score sustains by XGBoost starting at a fair 85% for 8491 records. Further the score grows steadily, touching 90.5% at 33926 records.

Figure 9 and Table 9 show the accuracy of different ML models Naïve Bayes, HHO-SVM, and KNN and the proposed XGBoost on datasets ranging from 8491 to 33926 records. Accuracy is a basic performance measure in threat detection systems, representing the proportion of correctly classified events among the total samples. The XGBoost model shows a strong and consistent trend of increasing accuracy 93.4% with a larger dataset record. The steady improvement here demonstrates the ability of XGBoost to scale effectively with larger cybersecurity datasets.

Table 9 - Performance of Accuracy.

Number of Records	Naïve Bayes	HHO-SVM,	KNN	XGBoost
8491	53.5	62.6	72.4	84.2
16972	56.7	64.8	75.9	86.1
25463	58.5	68.4	77.5	88.9
33926	50.5	70.3	80.7	93.4

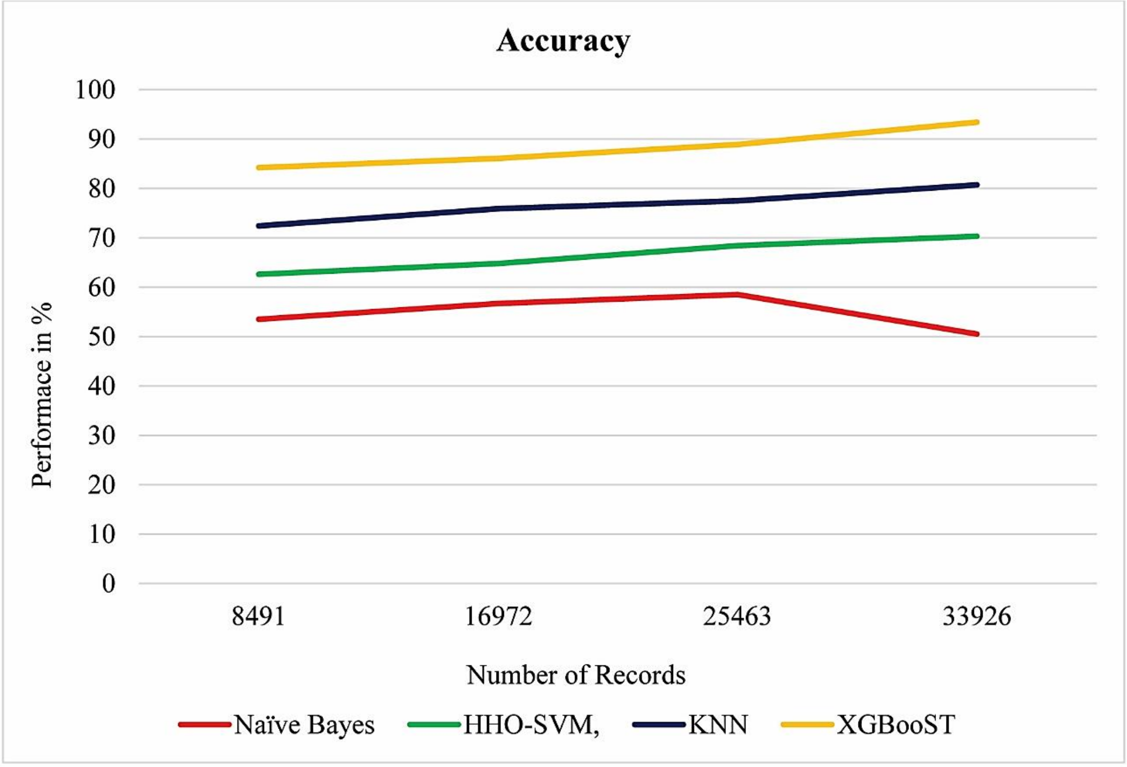


Fig. 9. Analysis of Accuracy.

In Figure 10 and Table 10, time complexity results, XGBoost performed the best, improving from 10.3ms to 98% as the dataset records. The previous Naïve Bayes, HHO-SVM, and KNN method had the lowest time complexity, at approximately 23.4 ms, indicating that it managed to detect more threats. Furthermore, the time complexity, which evaluates the balance, identified XGBoost as the top-performing model.

Table 10 - Performance of Time Complexity.

Number of Records	Naïve Bayes	HHO-SVM,	KNN	XGBoost
8491	35.3	30.5	27.5	23.4
16972	31.3	27.5	24.9	20.3
25463	28.4	22.4	20.4	16.4
33926	23.4	18.3	15.9	10.3

On the CICIDS2017 dataset XGBoost outperformed Naïve Bayes HHO-SVM KNN and other models in terms of accuracy (93.4%) precision (90.8%) recall (81%) and F1-score (90.5%) exhibiting strong generalization scalability and decreased time complexity. Its regularization and gradient boosting techniques reduced overfitting and accelerated convergence as demonstrated by a one-way ANOVA (p 0. 05). This work supports XGBoosts effectiveness and scalability for real-time intrusion detection while maintaining a lower computational cost than deep learning models in comparison to earlier research. The study’s goals of creating a reliable accurate and effective detection framework are in line with XGBoosts performance which shows practical suitability for SIEM IoT and edge-based cybersecurity applications despite minor misclassifications in uncommon attack types caused by class imbalance.

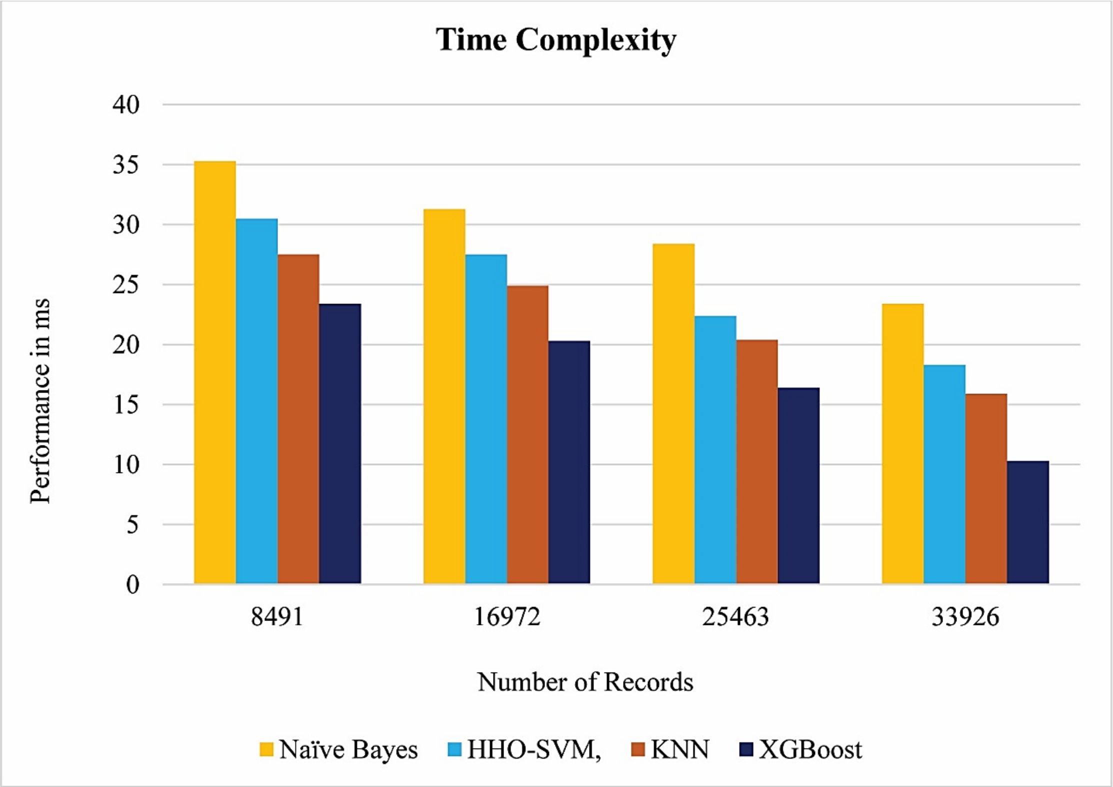


Fig. 10. Analysis of Time Complexity.

10. Conclusion

This review paper introduces different preprocessing and feature selection techniques to improve the performance and results of ML models designed for cybersecurity datasets. First, some data issues are discussed. Data pre-processing methods play a vital role in creating ML models, as the accuracy of the models heavily relies on data quality. As a result, the techniques of data pre-processing are important to analyze the cybersecurity dataset and obtain useful knowledge and information from the dataset. This paper reviews the use of ML algorithms to enhance cybersecurity, detailing their implementation and practical applications. ML provides valuable insights for security data analysis. The biggest dataset yields the most accurate XGBoost model predictions. The analysis shows the performance of the system is steadily improving and reaches 93.4% accuracy when trained on many examples. The model performs better with data then it can generalize and learn latent patterns of what is in data. Despite the reviews demonstration of XGBoosts efficacy its shortcomings include imbalanced datasets and a lack of real-time validation. Adversarial robustness federated learning frameworks and XAI integration should be the main topics of future research. It may be possible to use additional data points by investigating their characteristics to improve model training.

References

Agarwal, A., Sharma, P., Alshehri, M., Mohamed, A. A., & Alfarraj, O. (2021). Classification model for accuracy and intrusion detection using machine learning approach. *PeerJ Computer Science*, *7*, e437. https://doi.org/10.7717/peerj-cs.437

Ahn, B., Kim, T., Ahmad, S., Mazumder, S. K., Johnson, J., Mantooth, H. A., & Farnell, C. (2023). An overview of cyber-resilient smart inverters based on practical attack models. *IEEE Transactions on Power Electronics*, *39*(4), 4657-4673. https://doi.org/10.1109/TPEL.2023.3342842

Ahsan, M., Nygard, K. E., Gomes, R., Chowdhury, M. M., Rifat, N., & Connolly, J. F. (2022). Cybersecurity threats and their mitigation approaches using Machine Learning—A Review. *Journal of Cybersecurity and Privacy*, *2*(3), 527-555. https://doi.org/10.3390/jcp2030027

Ahsan, M., Gomes, R., Chowdhury, M. M., & Nygard, K. E. (2021). Enhancing machine learning prediction in cybersecurity using dynamic feature selector. *Journal of Cybersecurity and Privacy, 1*(1), 199–218. https://doi.org/10.3390/jcp1010011

Al Razib, M., Javeed, D., Khan, M. T., Alkanhel, R., & Muthanna, M. S. A. (2022). Cyber threats detection in smart environments using SDN-enabled DNN-LSTM hybrid framework. *IEEe Access*, *10*, 53015-53026. https://doi.org/10.1109/ACCESS.2022.3172304

Alashhab, A. A., Zahid, M. S., Isyaku, B., Elnour, A. A., Nagmeldin, W., Abdelmaboud, A., Abdullah, T. A. A., & Maiwada, U. D. (2024). Enhancing DDoS attack detection and mitigation in SDN using an ensemble online machine learning model. *IEEE Access, 12,* 51630–51649. https://doi.org/10.1109/ACCESS.2024.3384398

Almotairi, A., Atawneh, S., Khashan, O. A., & Khafajah, N. M. (2024). Enhancing intrusion detection in IoT networks using machine learning-based feature selection and ensemble models. *Systems Science & Control Engineering, 12*(1), 2321381. https://doi.org/10.1080/21642583.2024.2321381

Al-Shehari, T. A., Alshamrani, A., & Alsabaan, M. (2024). Enhancing insider threat detection in imbalanced cybersecurity settings using the density-based local outlier factor algorithm. *IEEE Access, 12,* 34820–34834. https://doi.org/10.1109/ACCESS.2024.3373694

Arshad, K., Khan, W., Khan, M. A., & Gumaei, A. (2022). Deep reinforcement learning for anomaly detection: A systematic review. *IEEE Access, 10,* 124017–124035. https://doi.org/10.1109/ACCESS.2022.3224023

Babagana, A., Adewale, A. A., & Idris, Y. A. (2024). The role of artificial intelligence in cybersecurity: A review of AI techniques and applications. *Journal of Artificial Intelligence and Security, 12*(2), 56–70.

Chukwunweike, J. N., Praise, A., & Bashirat, B. A. (2024). *Harnessing Machine Learning for Cybersecurity: How Convolutional Neural Networks are Revolutionizing Threat Detection and Data Privacy*. https://doi.org/10.55248/gengpi.5.0824.2402

Dhaiya, S., Pandey, B. K., Adusumilli, S. B. K., & Avacharmal, R. (2021). Optimizing API Security in FinTech Through Genetic Algorithm based Machine Learning Model. *International Journal of Computer Network and Information Security*, *13*(3), 24.

Ejiofor, O. E. (2023). A comprehensive framework for strengthening USA financial cybersecurity: Integrating machine learning and AI in fraud detection systems. *European Journal of Computer Science and Information Technology, 11*(6), 62–83. https://doi.org/10.37745/ejcsit.2013/vol11n66283

Ekundayo, F., Atoyebi, I., Soyele, A., & Ogunwobi, E. (2024). Predictive analytics for cyber threat intelligence in fintech using big data and machine learning. *International Journal of Research Publication and Reviews,* *5*(11), 1–15. https://doi.org/10.55248/gengpi.5.1124.3352

Ferrag, M. A., Maglaras, L. A., & Janicke, H. (2021). Cyber security intrusion detection for agriculture 4.0: Machine learning-based solutions, datasets, and future directions. *IEEE/CAA Journal of Automatica Sinica,* *9*(3), 407–436. https://doi.org/10.1109/JAS.2021.1004344

Guo, H., Li, J., Liu, J., Tian, N., & Kato, N. (2021). A survey on space-air-ground-sea integrated network security in 6G. *IEEE Communications Surveys & Tutorials, 24*(1), 53-87. https://doi.org/10.1109/COMST.2021.3131332

Haider, A., Aslam, M., Shah, M. A., Khan, A., & Baig, M. A. (2021). A real-time sequential deep extreme learning machine cybersecurity intrusion detection system. *Computers, Materials & Continua, 66*(2), 1785–1798. https://doi.org/10.32604/cmc.2020.013910

Hossain, M. A., & Islam, M. S. (2023). Ensuring network security with a robust intrusion detection system using ensemble-based machine learning. *Array, 19,* 100306. https://doi.org/10.1016/j.array.2023.100306

Ige, A. B., Kupa, E., & Ilori, O. (2024). Analyzing defense strategies against cyber risks in the energy sector: Enhancing the security of renewable energy sources. *International Journal of Science and Research Archive,* *12*(1), 2978–2995. https://doi.org/10.30574/ijsra.2024.12.1.1186

Ijiga, O. M., Idoko, I. P., Ebiega, G. I., Olajide, F. I., Olatunde, T. I., & Ukaegbu, C. (2024). Harnessing adversarial machine learning for advanced threat detection: AI-driven strategies in cybersecurity risk assessment and fraud prevention. *J. Sci. Technol, 11*, 001-024. https://doi.org/10.53022/oarjst.2024.11.1.0060

Injadat, M., Moubayed, A., Shami, A., & Lutfiyya, H. (2020). Multi-stage optimized machine learning framework for network intrusion detection. *IEEE Transactions on Network and Service Management, 18*(2), 1803–1816. https://doi.org/10.1109/TNSM.2020.3014929

Katiyar, N., Tripathi, M. S., Kumar, M. P., Verma, M. S., Sahu, A. K., & Saxena, S. (2024). AI and Cyber-Security: Enhancing threat detection and response with machine learning. *Educational Administration: Theory and Practice, 30*(4), 6273-6282. https://doi.org/10.53555/kuey.v30i4.2377

Kaur, S., Mahajan, S., Yousuf, M., & Koul, A. (2024). Security issues and challenges in cybersecurity: A comprehensive review. *Journal of Information Technology and Software Engineering, 14*(1), 1–6.

Keserwani, H., Ali, S., Bhatnagar, A., & Gupta, M. (2022). Security enhancement by identifying attacks using machine learning for 5G network. *International Journal of Communication Networks and Information Security, 14*(2), 124–141.

Khan, M., & Ghafoor, L. (2024). Adversarial machine learning in the context of network security: Challenges and solutions. *Journal of Computational Intelligence and Robotics, 4*(1), 51–63.

Kravchik, M., & Shabtai, A. (2021). Efficient cyber-attack detection in industrial control systems using lightweight neural networks and PCA. *IEEE Transactions on Dependable and Secure Computing, 19*(4), 2179–2197. https://doi.org/10.1109/TDSC.2021.30

Kuppa, A., & Le-Khac, N.-A. (2021). Adversarial XAI methods in cybersecurity. *IEEE Transactions on Information Forensics and Security, 16*, 4924–4938. https://doi.org/10.1109/TIFS.2021.3117075

Le, D. C., Zincir-Heywood, N., & Heywood, M. I. (2020). Analyzing data granularity levels for insider threat detection using machine learning. *IEEE Transactions on Network and Service Management*, *17*(1), 30–44. https://doi.org/10.1109/TNSM.2020.2967721

Li, Y., & Yan, J. (2022). Cybersecurity of smart inverters in the smart grid: A survey. *IEEE Transactions on Power Electronics, 38*(2), 2364–2383. https://doi.org/10.1109/TPEL.2022.3206239

Mahmood, R. K., Mahameed, A. I., Lateef, N. Q., Jasim, H. M., Radhi, A. D., Ahmed, S. R., & Tupe-Waghmare, P. (2024). Optimizing network security with machine learning and multi-factor authentication for enhanced intrusion detection. *Journal of Robotics and Control (JRC), 5*(5), 1502–1524. https://doi.org/10.18196/jrc.v5i5.22508

Mamidi, S. R. (2024). The role of AI and machine learning in enhancing cloud security. *Journal of Artificial Intelligence General Science (JAIGS), 3*(1), 403–417. https://doi.org/10.60087/jaigs.v3i1.161

Mohammed, S. H., Shabut, A. M., Ali, M. H., & Al-Zubaidi, S. (2024). A review on the evaluation of feature selection using machine learning for cyber-attack detection in smart grid. *IEEE Access, 12*, 44023–44042. https://doi.org/10.1109/ACCESS.2024.3370911

Mukesh, V. (2025). A Comprehensive Review of Advanced Machine Learning Techniques for Enhancing Cybersecurity in Blockchain Networks. *Journal ID, 8736*, 2145.

Musa, M. A., Adeyemi, A. O., & Okon, I. E. (2024). AI-driven approaches for improving cybersecurity in healthcare systems. *Journal of Medical Systems and Cybersecurity, 3*(1), 17–30.

Nabi, F., & Zhou, X. (2024). Enhancing intrusion detection systems through dimensionality reduction: A comparative study of machine learning techniques for cyber security. *Cyber Security and Applications, 2*, 100033. https://doi.org/10.1016/j.csa.2023.100033

Nassif, A. B., Shahin, T. M., Talib, M. A., & Azad, M. A. K. (2021). Machine learning for cloud security: A systematic review. *IEEE Access, 9*, 20717–20735. https://doi.org/10.1109/ACCESS.2021.3054129

Nazir, A., He, J., Zhu, N., Wajahat, A., Ullah, F., Qureshi, S., ... & Pathan, M. S. (2024). Collaborative threat intelligence: Enhancing IoT security through blockchain and machine learning integration. *Journal of King Saud University-Computer and Information Sciences*, *36*(2), 101939. https://doi.org/10.1016/j.jksuci.2024.101939

Okoli, U. I., Obi, O. C., Adewusi, A. O., & Abrahams, T. O. (2024). Machine learning in cybersecurity: A review of threat detection and defense mechanisms. *World Journal of Advanced Research and Reviews, 21*(1), 2286–2295. https://doi.org/10.30574/wjarr.2024.21.1.0315

Olutimehin, A. T. (2025). The Synergistic Role of Machine Learning, Deep Learning, and Reinforcement Learning in Strengthening Cyber Security Measures for Crypto Currency Platforms. *Deep Learning, and Reinforcement Learning in Strengthening Cyber Security Measures for Crypto Currency Platforms (February 11, 2025)*. https://dx.doi.org/10.2139/ssrn.5138889

Ozkan-Okay, M., Akin, E., Aslan, Ö., Kosunalp, S., Iliev, T., Stoyanov, I., & Beloev, I. (2024). A comprehensive survey: Evaluating the efficiency of artificial intelligence and machine learning techniques on cyber security solutions. *IEEe Access*, *12*, 12229-12256. https://doi.org/10.1109/ACCESS.2024.3355547

Saheed, Y. K., & Arowolo, M. O. (2021). Efficient cyber-attack detection on the internet of medical things-smart environment based on deep recurrent neural network and machine learning algorithms. *IEEE Access, 9*, 161546–161554. https://doi.org/10.1109/ACCESS.2021.3128837

Sarker, I. H. (2021). CyberLearning: Effectiveness analysis of machine learning security modelling to detect cyber-anomalies and multi-attacks. *Internet of Things, 14*, 100393. https://doi.org/10.1016/j.iot.2021.100393

Shaukat, K., Luo, S., Varadharajan, V., & Chen, S. (2020). Performance comparison and current challenges of using machine learning techniques in cybersecurity. *Energies, 13*(10), 2509. https://doi.org/10.3390/en13102509

Siddiqi, M. A., & Pak, W. (2021). An agile approach to identify single and hybrid normalization for enhancing machine learning-based network intrusion detection. *IEEE Access*, *9*, 137494–137513. https://doi.org/10.1109/ACCESS.2021.3118361

Tendikov, N., Rzayeva, L., Saoud, B., Shayea, I., Hadri Azmi, M., Myrzatay, A., & Alnakhli, M. (2024). Security information event management data acquisition and analysis methods with machine learning principles. *Results in Engineering, 22*, 102254. https://doi.org/10.1016/j.rineng.2024.102254

Tulli, S. K. C. (2023). Enhancing Marketing, Sales, Innovation, and Financial Management Through Machine Learning. *International Journal of Modern Computing*, *6*(1), 41-52.

Vaddadi, S. A., Vallabhaneni, R., & Whig, P. (2023). Utilizing AI and machine learning in cybersecurity for sustainable development through enhanced threat detection and mitigation. *International Journal of Sustainable Development through AI, ML and IoT, 2*(2), 1–8.

Vashishth, T. K., Sharma, V., Sharma, K. K., Kumar, B., Chaudhary, S., & Panwar, R. (2024). Enhancing cloud security: The role of artificial intelligence and machine learning. In *Improving security, privacy, and trust in cloud computing* (pp. 85-112). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-1431-9.ch004

Wang, L., Zhang, Y., Li, X., & Chen, W. (2024). A deep learning-based approach for network intrusion detection. *Journal of Cybersecurity Research, 6*(3), 112–126.

Wazid, M., Das, A. K., & Rodrigues, J. J. P. C. (2022). Uniting cybersecurity and machine learning: Advantages, challenges and future research. *ICT Express, 8*(3), 313–321. https://doi.org/10.1016/j.icte.2022.04.007

Wu, H., Ding, S., Wang, H., Yang, S., & Deng, Q. (2020). Research on artificial intelligence enhancing internet of things security: A survey. *IEEE Access, 8,* 153826–153848. https://doi.org/10.1109/ACCESS.2020.3018170

Yadav, A., Singh, R., Verma, R., & Sharma, D. (2024). Application of AI in cybersecurity: A survey. *International Journal of Advanced Computer Science and Applications, 15*(1), 89–99.

Yaseen, A. (2023). The role of machine learning in network anomaly detection for cybersecurity. *SAGE Science Review of Applied Machine Learning, 6*(8), 16–34.

Ye, J., Chen, B., Yang, Y., & Mu, Y. (2021). A review of cyber–physical security for photovoltaic systems. *IEEE Journal of Emerging and Selected Topics in Power Electronics, 10*(4), 4879–4901. https://doi.org/10.1109/JESTPE.2021.3111728

Yeboah-Ofori, A., Imamverdiyev, Y., & Epiphaniou, G. (2021). Cyber threat predictive analytics for improving cyber supply chain security. *IEEE Access*, *9*, 94318–94337. https://doi.org/10.1109/ACCESS.2021.3087109