ISSN 1840-4855
e-ISSN 2233-0046

Review article
http://dx.doi.org/10.70102/afts.2025.1833.215

EVALUATING THE EFFECTIVENESS OF PREDICTION TECHNIQUES FOR CYBERATTACKS: A COMPREHESIVE TAXONOMY

Azhar F. Al-zubidi[1*], Alaa Kadhim Farhan[2], Abeer Alsadoon[3]

[1]Computer Science Department, College of Sciences, AL Nahrain University, Jadriya, Baghdad, Iraq; Computer Sciences Department, University of Technology, Baghdad, Iraq.
e-mail: cs.21.07@grad.uotechnology.edu.iq, orcid: https://orcid.org/0000-0002-0395-9482
[2]Computer Sciences Department, University of Technology, Baghdad, Iraq.
e-mail: alaa.k.farhan@uotechnology.edu.iq, orcid: https://orcid.org/0000-0002-1036-9392
[3]School of Computing and Mathematics, Charles Sturt University, Sydney, Australia.
e-mail: alsadoon.abeer@gmail.com, orcid: https://orcid.org/0000-0002-2309-3540

Received: May 19, 2025; Revised: August 05, 2025; Accepted: September 03, 2025; Published: October 30, 2025

SUMMARY

The rising threat of cyberattacks in today's society emphasizes the urgent need for improved methods to both detect and prevent these incidents. This paper focuses on assessing the effectiveness of various techniques for predicting cyberattacks. The DTCF taxonomy was proposed for predicting these attacks, considering datasets, techniques, challenges, and future trends. This taxonomy includes four key stages. 1) data preprocessing, 2) feature selection, 3) development of prediction models, and 4) their subsequent validation and assessment. Our research reviews progress algorithms for each stage, analyzing their advantages and weaknesses. Consequently, the results of this study emphasize the critical role of precise detection and prediction in combating the increasingly complex threat of multiple cyberattacks, which are inherently more challenging to identify and predict than isolated incidents. Our examination of diverse learning methods reveals the essential role of data preprocessing in enhancing the efficacy of prediction systems. Effective preprocessing aids in reducing issues like noise, outliers, missing data, and extraneous features and, by doing so, refining the accuracy of predictions.

Key words: *cyberattack prediction, cyberattack datasets, machine learning models, dtcf taxonomy.*

INTRODUCTION

The advance of cyberattacks in recent years, has made their detection and prevention complex increasingly. Computer systems and networks are exposed to novel and sophisticated methods of attack made by hackers and cybercriminals causing substantial harm to individuals, organizations, and even governments. Consequently, machine learning (ML) emerged as a pivotal asset in combating these cyber threats. This article explores the application of ML for identifying and averting cyberattacks. We will

explore various ML algorithms including deep learning (DL) utilized in the area of cybersecurity and assess the strengths and limitations of each method. Moreover, we will shed light on some critical challenges in using ML effectively in cybersecurity focusing on aspects like data privacy and ensuring security. This paper postulate that ML holds considerable promise for enhancing our defensive capabilities against cyberattacks advocating for ongoing investment in this technology as a cornerstone in the future landscape of cybersecurity [1]. The use of machine learning (ML) for preventing cyberattacks aims to bolster our real-time threat detection and response capabilities [2]. Traditional methods in cybersecurity depend typically on manual monitoring and examination of the system logs. This approach can be both labor-intensive and slow hindering prompt threat identification and reduction. ML enhances our ability to rapidly detect and address threats potentially before they impose considerable harm [40]. To achieve this purpose cybersecurity professionals are employing a variety of ML algorithms such as DL, classification, and clustering, to analyze large volumes of data and identify patterns of suspicious behavior. These algorithms are trained by historical data and can continuously learn and adapt to new threats allowing them to improve their accuracy over time [3]. This achievement in cyberattack prediction depends on several elements: data preprocessing, feature engineering, intelligent model selection and implementation along with validation and evaluation. Data preprocessing is vital involving cleaning and formatting data to a form usable with prediction models like numerical, categorical, or textual. It also balances data distribution and reduces dimensionality and, in this way, boosts model efficiency and effectiveness. Since data is often noisy and incomplete preprocessing is crucial for eliminating irrelevant details and filling gaps. Acquiring current relevant data is challenging due to the dynamic and diverse nature of cyberattacks and limited data sources. Using diverse, up-to-date cyber datasets is crucial as outdated data fails to reflect the latest attack patterns and complexities leading to model biases and errors. Hence, employing current and relevant data is essential to effective accurate cyberattack prediction models [4]. The second step in cyberattack prevention is featuring engineering which includes extracting and selecting features. The third step involves choosing and implementing appropriate ML models such as supervised or unsupervised learning. These models in cyber-security, help identify patterns in network traffic or anomalies within system logs. For example, a supervised learning model trained on a labeled dataset of network traffic can identify cyberattack indicators within real-time traffic. The fourth step validation tests the ML model performance using methods like cross-validation. This ensures the model ac-curacy and reliability in cybersecurity. Lastly, the fifth step evaluation assesses the artificial intelligence (AI) system's overall effectiveness in thwarting cyberattacks. [5]. Measures like precision, recall, and accuracy are applicable to this objective. The assessment procedure in cybersecurity can uncover any system weaknesses or constraints enabling ongoing improvement and optimization. Machine learning (ML) is essential in addressing cyberattacks, due to its ability to rapidly detect and counteract increasingly complex threats [6]. ML algorithms effectively identify suspicious patterns in large data sets a task challenging for traditional methods [4]. ML enhances the efficiency and accuracy of cybersecurity efforts by replacing slower error-prone manual processes with AI automation. This is vital considering the significant financial and reputational risks posed by cyberattacks underscoring ML's essential role in cybersecurity [42]. Organizations can minimize the risk of these costly incidents protecting their assets and brand by using AI to prevent cyberattacks [8]. Finally, the use of AI in cybersecurity is essential to staying ahead of constantly evolving threats. Computer systems and networks are vulnerable to new methods of exploitation that hackers and cybercriminals constantly devise making it essential to have advanced technologies that can adapt and respond for error. We can automate the detection and response processes enabling us to address these threats by using ML technologies [7].

Research Motivation

The motivation behind this work is listed as follows:

1. As hackers exploit vulnerabilities in computer systems and networks, cyberattacks have grown more complex and harder to identify and stop over time.

2. Cyberattacks are increasing in frequency, complexity, and sophistication, posing serious threats to individuals, organizations, and governments.

3. ML models are able to learn from data and adapt to new situations, making them suitable for detecting and preventing cyberattacks.

4. ML models can also improve the efficiency and accuracy of security operations by reducing false positives, automating response actions, and providing insights and recommendations.

LITERATURE REVIEW

In this paper, we have examined several recent papers that explore the use of ML techniques for predicting cyberattacks. Unlike previous reviews that focused on specific aspects or applications of cyberattack prediction, we have offered a complete and comparative review of the state-of-the-art in this area at this time. Because it includes taxonomy with several components for each of the publications we reviewed, our study differs from others in terms of organization. The taxonomy is based on four dimensions: data sources, ML techniques, evaluation methods, and performance results. By using this taxonomy, we have highlighted the similarities and differences among the existing studies, as well as the gaps and challenges that need to be addressed. The taxonomy also helps to identify the best practices and recommendations for cyberattack prediction research and practice. We believe that our paper offers a novel and useful perspective on the potential and limitations of ML for cybersecurity. [1, 2] It is a survey paper that offers a thorough overview and a neutral comparison of the available DL methods for cyber security intrusion detection. We evaluate various methods such as the deep belief network, stacked AE, CNN, RNN, LSTM, and GRU network. We also discuss the challenges and limitations of DL methods and compare their performance with different datasets [3].

This document covers all high-dimensional big data anomaly detection methods. We examine statistical, distance-based, density-based, clustering-based, subspace-based, feature selection, and feature extraction methods. [41] is a detailed survey of fuzzy signature-based IDS for cyber security intrusion detection. We discuss fuzzy signature-based (FSB) anomaly, misuse, and hybrid detection systems. We compare fuzzy signature-based IDSs to others and evaluate their pros and cons. [5] gives a complete and systematic overview of the literature and the current state of secure data analytics using ML and DL models and techniques and offers some insights and recommendations for researchers and practitioners in this field. [42] provides an (SLR) of the (AIDS) in IoT using DL techniques. We analyze the existing published literature regarding AIDS using DL techniques in securing IoT environments. We discuss the advantages and disadvantages of these techniques, as well as the challenges and future directions for AIDS in the IoT. In [7], a thorough review of the application of (ANN) approaches for cybersecurity is provided. Various ANN methods, cyberattacks, datasets, and applications in cybersecurity are discussed. We also stress the importance of cybersecurity for IoT-driven healthcare systems. [43] A review of cybersecurity methods that were suggested and implemented recently for detecting and predicting attacks and a review of cybersecurity techniques for attack detection, prediction, and prevention are provided in this paper. We evaluate the techniques, benefits, and drawbacks of these strategies and provide possible areas of study for the future.

Contributions

In this paper, we conduct a comparative analysis of many papers on cyberattack prediction from various perspectives. Unlike previous surveys that focused on specific aspects or applications of cyberattack prediction, we aim to provide a holistic and systematic overview of the current state of the art in this field Table 1. The search flow chart of reviewed papers shows the criteria and process of selecting relevant papers for the literature review. It includes the year the paper was published, journal quarter, and other details, such as keywords A summary of contributions can be listed as follows:

1. We analyze the 30 papers on cyberattack prediction that were selected by summarizing their main findings, methods, datasets, and limitations.

2. We extract a table for the datasets used in the abovementioned papers and their properties, such as size, format, domain, features, and links.

3. Classify the ML techniques that were used to predict cyberattacks. In addition, we explain how they work and describe their advantages and disadvantages.

4. We draw taxonomy for all the prediction systems based on their data sources, ML techniques, evaluation methods, and performance results.

5. We list and compare all the datasets as in Table 2, focusing mainly on Network security like (NSL-KDD99, UNSW-NB15, KDD Cup), IOT datasets like (IoT-Botnet 2020, DS2OS, N-BaIoT), industrial control systems, such as (CIDDS-001, SCADA data (Gas Pipeline (GP), Secure Water Treatment (SWaT)), Modbus Dataset), and Web Security datasets like (ISCX-URL2016).

6. We recommend the best, second-best, and third-best models based on our analysis and explain why they are better for cyberattack prediction.

7. We suggest some future directions and open problems for cyberattack prediction research.

Table 1. Searching engines

Searching Engines	URL link
Springer	https://www.springer.com/gp
IEEE Xplore	https://ieeexplore.ieee.org/Xplore/home.jsp
Elsevier	https://www.elsevier.com/en-xm
MDPI	https://www.mdpi.com/
Hindawi	https://www.hindawi.com/journals/
Google Scholar	https://scholar.google.com/
ResearchGate	https://www.researchgate.net/
Semantic Scholar	https://www.semanticscholar.org/

Cyber Attack Datasets

One of the crucial components of cyberattack prediction is the method used for creating and testing the prediction models. Different datasets may have different properties, such as domain, format, size, attack type, and number of features, that can affect the performance and applicability of the prediction models. Therefore, it is important to compare and evaluate the datasets used in the literature on cyberattack prediction. In this section, we create a table Table 2 for the properties of the datasets used in the papers that we have reviewed and analyzed [5], we only used datasets that are openly shared and published by reputable sources, such as NSL-KDD, ToN_IoT, CIDDS-001, and ISCX-URL2016. Various domains like Network security, IOT, Industrial control systems, and Web Security, respectively which have multiple cyberattack types such as network intrusion, DDoS, phishing, malware, and botnet [7,8].

Table 2. Dataset information

No.	**Dataset name**	**Domain**	**Format**	**Size**	**Attack Type**	**no. of Features**	**Dataset URL**
1	NSL-KDD99	Network security	ARFF	4 million network connections	DoS, Probe, R2L, U2R	42 features	http://nsl.cs.unb.ca/NSL-KDD/
2	UNSW-NB15	Network security	CSV	2 million network packets	DoS, Probe, R2L, U2R	45 features	https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/

3	KDD Cup	Network security	CSV		DoS, Probe, R2L, U2R	41 features	http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
4	ToN_IoT	IOT	CSV	8 million IoT network flows	DoS, MITM (Man-In-The-Middle), Reconnaissance	17 features	https://research.unsw.edu.au/projects/toniot-datasets
5	CI-CIDS2017	Network security	CSV	16 million network flows	DoS, DDoS, Brute-Force, Web Attacks, Infiltration, Botnet	80 features	http://www.unb.ca/cic/datasets/ids-2017.html
6	CICDDoS 2019	Network security	CSV	2.2 million and 51,000 rows	DDoS	88 features	http://www.unb.ca/cic/datasets/ddos-2019.html
7	CIDDS-001	ICS	CSV	4.9 million net	DoS, DDoS, Botnet, Infiltration	80 features	https://www.kaggle.com/datasets/dhoogla/cidds001
8	IoT-Botnet 2020	IOT	CSV	2,000,064 instances	Botnet, Mirai	115 features	https://research.unsw.edu.au/projects/bot-iot-dataset
9	ISCX-URL2016	Web Security	CSV	1 million URLs	Malicious, Benign URLs	30 features	https://archive.ics.uci.edu/ml/datasets/phishing+websites#
10	Malware Dataset		MAT	9,150 binary executable files.	Viruses, Worms, Trojans, and other types of malware	56,102 features	https://www.kaggle.com/amauricio/pe-files-malwares
11	DS2OS	IOT	CSV	24,373 system call traces	Privilege Escalation, Remote Code Execution,	2,903 features	https://www.kaggle.com/datasets/francoisxa/ds2ostraffictraces
12	N-BaIoT	IOT	ARFF	50,000 network flows	DoS, MITM, Information Gathering,	11 columns	https://archive.ics.uci.edu/ml/datasets/detection_of_IoT_botnet_attacks_N_BaIoT
13	SCADA data (Gas Pipeline (GP), Secure Water Treatment (SWaT))	ICS security	CSV	varying numbers of instances.	Sensor Spoofing, Command Injection, Denial-of-Service	51 features	https://itrust.sutd.edu.sg/testbeds/secure-water-treatment-swat/
14	Modbus Dataset	ICS security	CSV	137,052 rows	Modbus Attacks, such as Modbus Command	17	https://ieee-dataport.org/documents/modbus-dataset-ics-anomaly-detection

15	IOT-23	IOT	CSV	42,880 rows	MITM, Information Gathering, Exploitation	115	https://www.stratosphereips.org/datasets-iot23
16	LITNET-2020	IOT	CSV	5,000,000 rows	DoS, DDoS, Botnet, Port Scanning, and more	83	https://dataset.litnet.lt
17	NetML-2020	IOT	CSV	7,077,175 rows	"DoS, DDoS, Botnet,		https://eva-lai.cloudcv.org/web/

Structure of This Survey

In Figure 1, we list the organization of this paper according to popular review structures. Here, 2 is the DTCF taxonomy components, 3 is the system classification based on the reviewed papers, 4 is the system evaluation list of all the evaluation metrics, and 5 is the conclusion.

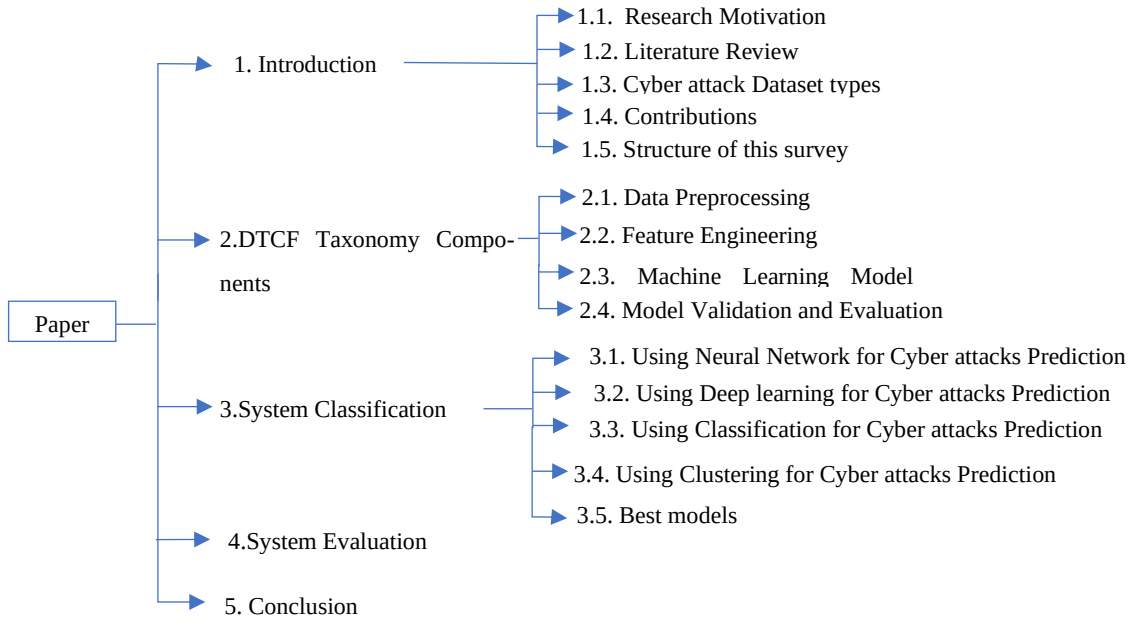

Figure 1. Paper structure

DTCF TAXONOMY COMPONENTS

Each component in the DTCF taxonomy plays a critical role in the design and implementation of an accurate and effective prediction system and is based on four main components: preprocessing, feature selection, prediction model construction, and validation and evaluation. The aim is to provide a clear and consistent structure for describing and comparing different ML and DL approaches and applications. This taxonomy is presented in Figure 2 and consists of the following components.

PREDICTION MODEL SYSTEM CLASSIFICATION

The objective of this work is to provide a comprehensive overview of the literature on cybersecurity approaches for attack prediction, detection, and prevention. To achieve this goal, we have conducted a

systematic literature review (SLR), and we will consider datasets, techniques, challenges, and future directions (DTCF taxonomy) for each of the analyzed papers in Table 3. We have analyzed these papers depending on various criteria, such as the research questions, the research methods, the data sources, the cyber security techniques, and the evaluation metrics. To present our findings in a clear and concise way, we have created a classification table of the review papers, which we refer to as the system classification.

Using Neural Network for Cyber Attacks Prediction

Several neural network models have been proposed to solve the problem of cyberattack detection. For example, [44] employed 10 well-known supervised and unsupervised ML techniques to locate useful and efficient ML-AIDS in computer networks. EM, k-means, and SOM are examples of unsupervised ML algorithms, while ANN, DT, k-NN, NB, RF, SVM, and CNN are examples of supervised ML techniques. To exploit the advantage of an MLP's propensity for learning complicated and nonlinear patterns from network traffic data and classifying them into normal or attack categories, ML-AIDS models have been evaluated using the CICIDS2017 dataset with network attacks in the real world. This dataset contains various types of network attacks, such as DDoS, brute force, and SQL injection, as well as benign traffic. The dataset suffers from the class imbalance problem, which poses a challenge in the classification task, as the model may be biased towards the majority class. MLP can overcome this challenge by using different techniques, suchas oversampling, undersampling, and weighted loss functions, to balance the classes and improve the performance [11].

Introduced a technique for selecting embedded features by using GIWRF in Intrusion Detection Systems (IDS). This method underwent evaluation by using UNSW-NB 15 and Network TON_IoT datasets for binary classification. The paper presents a comparative analysis of various ML models including Decision Trees (DT), AdaBoost, LSTM, Gradient Boosted Trees (GBT), Multi-Layer Perceptrons (MLP) and Gated Recurrent Units (GRU) focusing on specific task. The findings revealed that DT when coupled with specialized feature selection technique outperformed other models. This novel approach integrating DT with the technique proved to be a more effective than past methods. The validity of the method was demonstrated by testing it across two distinct datasets, TON_IoT and UNSW-NB 15 which feature a mix of realistic network attacks and benign traffic thus validate its efficacy within various contexts.

The datasets are imbalanced, which means that there is more normal traffic data than attack traffic data. This poses a challenge in the classification task. [45] presented CyberLearning for both classification and regression. CyberLearning contains neurons and is modelled after the brain. [13]

Using Deep Learning for Cyberattack Prediction

Detecting and preventing cyberattacks is a challenging task, as attackers constantly evolve their techniques and strategies. Therefore, there is a need for advanced methods that can learn from data and predict the occurrence and type of cyberattack [12]. One such method is DL, which is a branch of ML that can learn complex patterns and features from large amounts of data. [14] adapted a DL architecture to represent network traffic data to classify malicious and benign network packets using DL, deep feedforward neural network, feature selection, dimensionality reduction, and clustering. The model had the highest accuracy of 99.92% for warm attacks with UNSW_NB15, an accuracy of 99.99% for the CICIDS2017 dataset and an FPR of 0.00001 compared with the approach used by [15]. The method achieved an impressive accuracy of 99.9%, but it was only tested on three kinds of attacks in a single dataset, which limits its ability to handle the variety and complexity of evolving attack types. [16] developed an (HT-RLSTM) approach that can locate attacks. We fixed some problems with the data (KDD99, UNSW-NB15, NSL-KDD99, and CIDDS-001), such as missing values, scaling, imbalance, and overlap. This helped us deal with uncertain data and avoid false alarms. This framework can stop new kinds of attacks such as APT and zero-day attacks. Highly informative features were used to train the HT-RLSTM and gain deep insights. Our comparison showed that the HT-RLSTM outperformed other methods, such as SVM, KNN, ANFIS, and ANN, in terms of multiple metrics. It achieved scores between (94–97%) against those of other methods (82–95%).

However, the data used to train LSTM models for cyberattack prediction may not represent all attack scenarios [10]. If the training data are not sufficiently varied, the model may not generalize to new situations and detect novel attack patterns. [17] built a DeNNeS framework with derivative and deductive expert systems. Data are used to train a DNN, which gives the derivative expert system rules. It uses these rules in its knowledge base of input from the user and makes a decision by majority voting. It was compared against k-NN, JRip, SVM, DT, GNB, and RF on Android malware data. DeNNeS had 5:8% and 4:9% less FPR and 8:5% and 5:8% more ACC than JRip and RF, the rule learners, and RF, the best ML model. While [18] used TensorFlow, we constructed a deep ML model that can handle deep neural network training and inference techniques. TensorFlow helps computer science and other research and development. [19] proposed an approach to protect and detect DDoS attacks over a network by using multiple classification algorithms, assessing the DIDDOS's efficacy using naive Bayes and other traditional ML classifiers (NB), utilizing DL techniques such as (GRU), (RNN), (SMO) and (RNN), and comparing cutting-edge research and traditional methods such as (NB), (RNN), and (SMO).

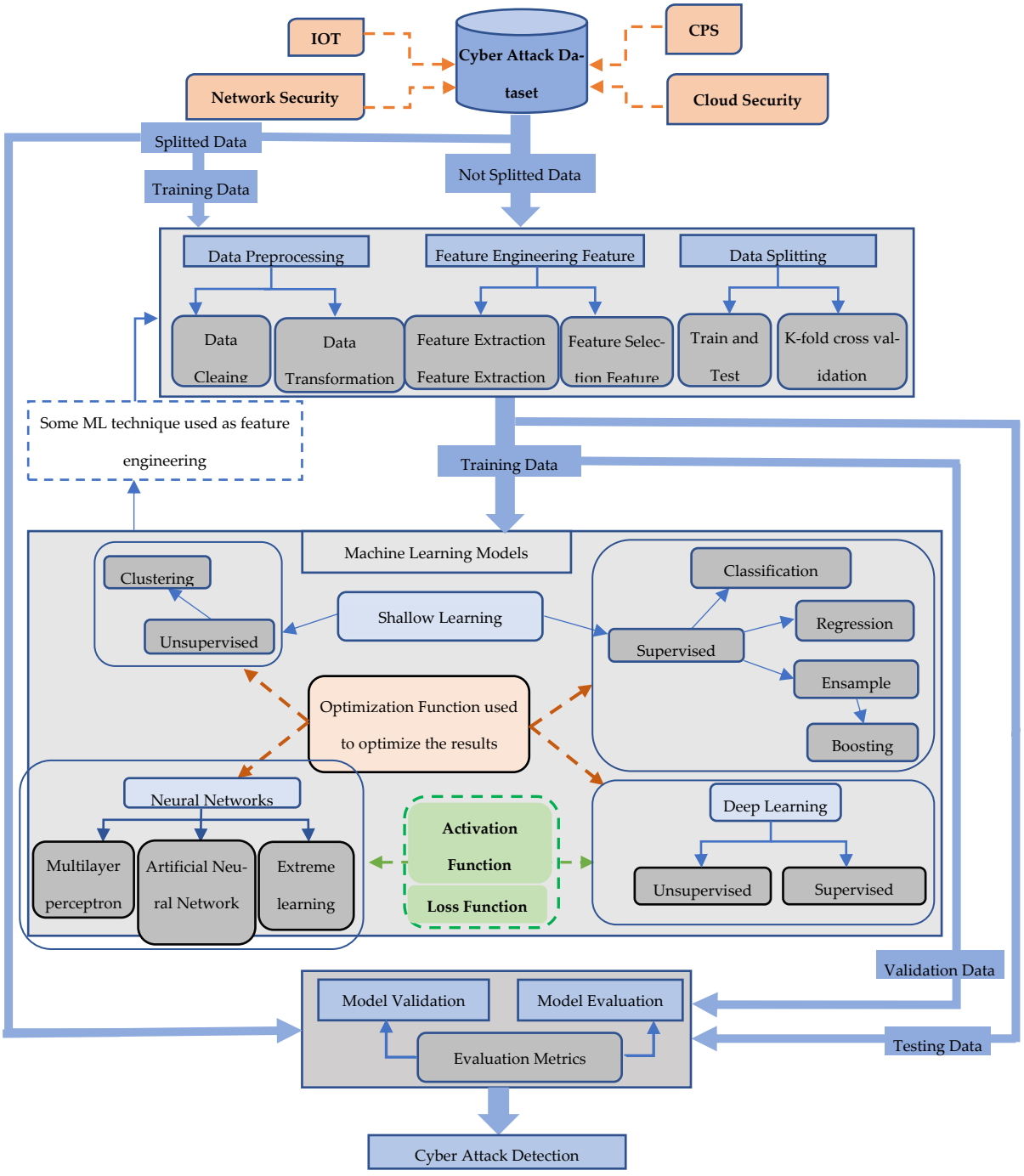

Figure 2. DTCF taxonomy

According to experimental findings, DDoS classification accuracy readings of 99.69% and 99.94% were obtained for the CICDDoS2019 dataset, which appears to be a complex dataset due to its variety of different DDoS attacks, reflection attacks and exploitation attacks, using GRUs. This makes it possible to capture more complex patterns, particularly when the data include a sequence of events or actions over time. They may require more resources and expertise to train and optimize the classification model. Proposed a hybrid optimization algorithm named HHO-PSO-DLNN [20].

To protect normal services from DDoS attacks and botnets, the hybrid HHO-PSO boosts the conventional LSTM model by choosing a few optimal features that increase the classification accuracy. The hybrid HHO-PSO-DLNN model outperformed the HHO-DLNN, DLNN, PSO-DLNN, and other models in finding intrusions in a cloud computing environment. However, predicting and finding DDoS attacks is hard and complex, and no one method or mix of methods can work for all cases. To fight DDoS attacks well, it is important to use a mix of different techniques and tools. it is third best model is the one that uses HHO-PSO-DLNN . This model can detect DDoS attacks and prevent botnets from disturbing network and server services by combining HHO, PSO, and DLNN. HHO-PSO is an optimization technique that can improve the typical LSTM model with greater classification accuracy. However, this model may not be able to handle other types of cyberattacks or complex network data.

In another study, [21] that introduced hybrid semantic deep learning (DL) architecture aimed to detecting intrusions in cloud environments. The research explores also the application of encryption algorithms to bolstering cloud storage security. Additionally, it explores the utilization of optimization algorithms to choosing the more effective encryption key further enhancing the security measures. From the evaluation and testing of the model by using real-time intrusion detection benchmark datasets, an accuracy of 98.47% was obtained for the UNSW-NB15, and an accuracy of 99.98% was obtained for the NSL-KDD dataset. Proposed a (TVCDNN) for the detection of (DDoS) attacks, as they flood a network with a large amount of traffic and make the services inaccessible to legitimate users [22]. The objective of that paper was to developing DDoS attack detection system capable for desicrimination the complex and nonlinear patterns in network traffic data and categorize them as either normal or attack-related patterns. Genetic-based optimization techniques were employed in order to refine structure and parameters of Deep Neural Network (DNN). The suggested TVCDNN model underwent testing by two public network traffic datasets and was evaluated against various other classifiers and optimization methods. The results demonstrated that TVCDNN accurately and efficiently detects DDoS attacks and also exceeding the leading attack detection systems currently available.

The study by [23] explored using Bayesian Neural Networks (BNNs) to detect unusual patterns in network traffic, potentially signaling cyber threats with a focus on the UNBS-NB-15 and KDD99 datasets to test. However, while BNNs offer a promising method to cyberattack protection within physical networks their suitability varies across organizations. It's essential to entities for weighing the costs and benefits of BNNs and assess their capacity to effectively deploy and manage such systems.

In the study conducted by [24] a novel deep learning model combining Kalman filtering (HDSCNN-KF) with Siamese convolutional neural networks (SCNNs) was proposed to deal the issue of scarce and unbalanced labeled data in cyber-physical systems (CPSs) focusing on advanced abnormality and threat detection. This approach significantly increase CPS security demonstrating a low false-negative rate (1.10%) and high detection rate (98.90%) on the Power System dataset. The model showed superior performance also with Gaussian mixture models (GMM) in attribute percentages of 75% and 100%. While thry effective in detecting diverse cyberattacks, enhancements in algorithm precision and speed could be achieved through dimensionality reduction technique like PCA and ICA. Separately, the authors of introduced TFDPM a framework extracting temporal and feature patterns from historical data followed by using a conditional diffusion probabilistic model for future value predictions.

Our method is a promising approach for attack detection in cyber-physical systems, but it may face some challenges such as data availability, complexity, assumptions, false-positives, scalability, and overfitting. Proposed CTP-DHGL, a dynamic heterogeneous graph learning-based end-to-end cyber threat prediction model that automatically predicts attacks from public security data [25]. For the CTP and AlienVault datasets, CTP-DHGL outperformed static-based techniques by 9.65%–23.16% and

12.70%–25.03%, respectively, in terms of precision. CTP-DHGL can record graph dynamics for predicting attacks on computers. However, the model must learn from past data and improve its predictions over time to become useful in predicting cyberattacks. Examples include using advanced techniques such as comparative learning and reinforcement learning. Applied a (CNN) to develop effective and efficient (ML-AIDS). with a recent and highly imbalanced CICIDS2017 dataset with real-world network attacks [26]. Proposed a DNN-based anomaly detection system for IoT network architecture that learns complicated network flows and classifies them as benign or anomalous [27]. Our model has a 99.01% detection accuracy [9]. The model detects anomalies at 99.9% and benign traffic at 96.2%. However, we acknowledge that DNNs have some limitations: they require large amounts of labelled data for training, the risk of overfitting to the training data, and the challenge of explaining the model's decisions. Proposed a hybrid model of a CNN and LSTM to detect botnet attacks on different kinds of IoT devices and tested the system with a dataset created from injecting ten attacks on nine commercial devices[28]. Proposed a (CNN-LSTM) hybrid deep-learning model for IoT botnet detection. The system detects Mirai and BASHLITE IoT attacks from types of four security camera. We evaluated our model based on evaluation metrics and demonstrated that it achieves optimal performance in detecting botnet attacks [56]. Used LSTM, CNN, and LSTM–CNN algorithms to identify phishing and authentic website URLs [29]. Our technology detected phishing websites well. The LSTM–CNN and LSTM algorithms had accuracies of 97.6% and 96.8%, respectively, while the CNN algorithm had an accuracy of 99.2%. Proposed a DL model to help standard IDSs identify ICS cyberattacks and balance skewed datasets [30]. These new representations were then used by an ensemble DL attack detection model that is tailored for an industrial control system (ICS) environment, which is an interesting and practical application of cyberattack prediction. The model employs DNN and DT classifiers to identify cyberattacks from the new representations. Presented an ensemble method that uses deep models such as LSTM and a DNN and a meta-classifier (logistic regression) to detect network anomalies [31]. The models were evaluated with heterogeneous datasets, including NetML-2020, IoT-23, and LITNET-2020, which are data collected in an IoT environment, it is second best model, it can capture both nonlinear and complicated patterns and sequential and temporal dependencies from network traffic data. It also employs a meta-classifier logistic regression to integrate DNN and LSTM predictions via stacked generalization, which can improve the accuracy and robustness of the model. LuNet, a DNN architecture, detects large-scale network breaches [32]. LuNet learns traffic data spatial features with a CNN and temporal features with LSTM. LuNetwas tested with UNSW-NB15 and NSL-KDD. LuNet surpassed other types of network intrusion detection methods in terms of validation accuracy and false-positive rate. However, it cannot classify backdoors and worms. proposed CAD to detect anomalies in cloud-based environments using ML models [33]. An ensemble ML (EML) model classifies binary anomalies, whereas a CNN-LSTM classifies multiclass anomalies. We evaluated our binary anomaly detection and multiclass anomaly categorization with a difficult UNSW dataset.

Table 3. Technique classification

Ref.	Machine Learning Techniques																							
	Neural Network		Deep learning						Classification													Clustering		
	MLP	ANN	DNN	CNN	RNN	LSTM	GRU	AE	HMM	BN	NB	GB	XG boost	RF	DT	SVM	LR	SGD	ELM	EL	k-NN	k-means	SOM	PCA
[8]			✓		✓	✓																		
[9]			✓																					
[10]											✓		✓	✓	✓	✓	✓	✓			✓			
[11]		✓		✓							✓			✓	✓	✓					✓	✓	✓	
[12]													✓	✓							✓			
[13]				✓	✓																			
[14]			✓																					
[15]				✓																				
[16]																			✓					✓
[17]				✓																✓				
[18]	✓					✓	✓					✓	✓		✓									
[19]				✓		✓						✓	✓	✓	✓									

[20]								✓		✓														
[21]																		✓					
[22]					✓		✓				✓												
[23]			✓	✓	✓	✓	✓																
[24]			✓																				
[25]				✓		✓																	
[26]				✓		✓																	
[27]			✓											✓									
[28]						✓								✓					✓				
[29]			✓			✓										✓							
[30]												✓											
[31]				✓				✓															
[32]											✓	✓	✓	✓									
[33]										✓										✓			
[34]						✓																	

Using Classification for Cyberattack Prediction

We can use classification techniques to predict the type or category of cyberattack based on network traffic data or other relevant features. Labelled data are used to train a model that can assign new data to predefined classes. For example, we can use classification to predict whether a network packet is normal or malicious or what kind of attack it is, such as DDoS, brute force, or SQL injection. Presented CyberLearning, which is an ML-based cybersecurity modelling approach with correlated feature selection. By analysing the effectiveness of various ML security models, this model uses a binary classification model to detect anomalies and a multiclass model for cyberattacks. This system can be a powerful approach for cyberattack detection because it allows the system to benefit from the strengths of each individual model and improves the overall performance. However, it also increases the complexity and performance of the system. discussed cybersecurity data science and related methodologies and highlighted data-driven intelligent decision-making for cyber defense [34]. It also discussed the problems and future goals in cybersecurity data science and provided an ML-based multilayered cybersecurity modelling framework. Employed metaheuristic cyber ant optimization to extract aberrant health features [35]. Then, an attack was detected with an ensemble crossover XGBoost classifier. Our method significantly improved the detection accuracy, true positive rate, and false-positive rate. It improved IoT malware detection, protecting patients and health care providers. However, it is important to note that the success of the model will depend on the quality and quantity of the data that are used to train it. It is also important to thoroughly test the model with a variety of real-world data in a health cloud environment to see how well it performs in real time. In [36], to determine the best and most suitable ML-AIDS for networks and computers, 10 well-known ML algorithms from both supervised and unsupervised learning were used. The supervised approaches were ANN, DT, k-NN, NB, RF, SVM, and CNN, while the unsupervised approaches were EM, k-means, and SOM. We tested the ML-AIDS models with a real-world and highly imbalanced CICIDS2017 dataset that has different types of network attacks. The results showed that the DTAIDS and NB-AIDS models are more effective in detecting web attacks than the other models that have inconsistent and lower performance. Suggested a hybrid method that combines ELM and Bayesian optimization and uses a cloud architecture to prevent cyberattacks in real-time IoMT settings [37]. It makes better predictions by taking into account the predictions from the individual ML methods. The proposed method outperformed the other methods in terms of precision, recall, F1 score, F2 score, Fbeta score, and AUC-ROC curve, with values of 0.990300, 0.990300, 0.990300, 0.989175, 0.986652, and 0.870034, respectively. The results showed that the hybrid method of ELM and Bayesian optimization is more accurate than using either method alone. However, there are other intelligent methods that can also protect the devices and network using relevant information, such as ML-based IDS, anomaly detection, and firewall rule-based systems. [38] proposed a unified learning framework for regression and multiclass classification problems using an extreme learning machine (ELM). ELM is a simple and efficient algorithm that works for generalized single-hidden-layer feedforward networks (SLFNs) with random hidden nodes. It was shown that ELM has better scalability, faster learning speed and similar or better generalizability than those of traditional SVMs and their variants. suggested predicting MTM and DoS assaults with RF, XGBoost, GB, and DT. From the two datasets, The following was found: All algorithms detect MTM and DoS attacks with

approximately 99% and 97% accuracy, respectively [39]. These algorithms detected MTM and DoS attacks as well.

These algorithms can produce highly accurate models, especially when the data are well suited to the algorithm and the features are well engineered. However, to apply pretrained models, more DL algorithms, and all state-of-the-art models to future datasets, Proposed six ML classification techniques to identify eleven DDoS assaults with distinct DDoS attack datasets. The Canadian Institute of Cyber Security's CICDDoS2019 dataset comprises eleven CSV DDoS attack files. We compared logistic regression, DT, RF, AdaBoost, KNN, and NB to identify the best detection classification algorithms. Presented ML detection and classification of DDoS attacks utilizing k-NN, QDA, GNB, and CART [34]. All algorithms can classify and detect such attacks, but CART surpasses the others in terms of prediction accuracy, stability, prediction speed, and training time. This work has not adjusted or optimized the methods' hyperparameters, which could be investigated in future studies along with the predictors' and designed features' utility. In [31], two anomaly based ML models were introduced to establish that they provide greater security than misuse-based methods. This CNN-ensemble learning model uses NB, KNN, logistic regression, and SVM. The ensemble model outperformed the CNN model in terms of "explainability" as they can help understand how the prediction models work and why they make certain decisions, and computing efficiency.

Using Clustering for Cyberattack Prediction

As shown in [30], KPCA-DEGSA-HKELM, powerful IDS, can detect malicious assaults. DEGSA-HKELM increases the mean F score by 2.09% and 1.25% for the KDD99 dataset. The current work's mean F score was 21.01%, 28.34%, 1.44%, 15.93%, which are higher than those of CSVAC, KDDwinner, Dendron, and CPSO-SVM, respectively. The current approach has a lower FAR than those of other issued methods for the UNSW-NB15 dataset. The accuracy of a model for the industrial intrusion TE dataset improves by 3.45% with the suggested technique. KPCA-DEGSA-HKELM tests KDD99, UNSW-NB15, and intrusion TE datasets faster than CPSO-SVM, specifically by 60.57%, 82.21%, and 49.09%, respectively. Proposed a hybrid feature selection-based intelligent cyber threat detection system for IoT networks using ML [26]. kNN, RF, and XGBoost help IoT networks make quick and effective decisions. Proposed an ML-based two-tier network anomaly detection model for network malware detection [39]. BiLSTM reduces the feature space and selects the best features in the dimension reduction step. NB, certainty factor voting KNN classifiers, and DT classify network traffic as abnormal or normal. This technique was compared with others with the NSL-KDD dataset.

Best Models

Hybrid models are models that combine two or more ML techniques to improve the prediction performance. For example, some papers used DNN and CNN, RF and SVM, or DNN and RF hybrid models. Hybrid models can leverage the advantages and weaknesses of the individual techniques. However, hybrid models also have some drawbacks, such as increased complexity, computational cost, and difficulty of interpretation. Therefore, hybrid models should be carefully designed and evaluated to ensure their effectiveness and efficiency for cyberattack prediction. In [31], CNN, LSTM, and LSTM-CNN were used. LSTM-CNN was used to detect phishing URLs by analysing spatial and temporal aspects from URL characters and sequences to gather complementary information. Gradient boosting techniques such as XGBoost and AdaBoost were used to combine the predictions of numerous weak learners into a strong learner to improve the model classification, so it considers to be best model The second best model is the one that uses CNN, LSTM, and LSTM-CNN. This model can detect phishing URLs by analysing spatial and temporal aspects from URL characters and sequences to gather complementary information. It also uses gradient boosting techniques such as XGBoost and AdaBoost to combine the predictions of numerous weak learners into a strong learner, which can enhance the model performance. In [33], a novel model that uses a DNN, LSTM, and logistic regression was proposed to detect network anomalies and cyberattacks using network traffic data. A DNN captures nonlinear and complicated patterns, while LSTM captures sequential and temporal dependencies. It employs meta-classifier logistic regression to integrate DNN and LSTM predictions via stacked generalization. In [12], a hybrid HHO-PSO-DLNN model was used to detect DDoS attacks and prevent botnets from disturbing network and

server services by combining (HHO), (PSO), and (DLNN). HHO-PSO, optimizing the typical LSTM model with greater classification accuracy.

SYSTEM EVALUATION

Evaluation measurements are essentially in the development of cyberattack prediction models offering a numerical assessment of the model effectiveness and aiding in decision-making throughout the process. These measurements are consistently applied to evaluate model performance. They serve to compare various models and adjust their hyperparameters in the validation stage. Typical measurements for classification models include accuracy as illustrated in Table 4, F1 score, recall, precision, and ROC curve, as mentioned in [34]. Validation and evaluation processes are important in identifying the advantages and limitations of various cyberattack prediction models along with determination of their appropriateness of different attack types. These steps also shed light upon the inherent challenges and constraints with predicting cyberattacks including factors like data quality, scalability, interpretability and ethical considerations [35]. The finalized model submitted for testing with an independent dataset employing evaluation measurements to provide an impartial assessment of its performance. This essential step helps for deciding whether the model is prepared to practical application, as noted in. Common machine learning performance measurements include accuracy, F-measure, precision, ROC-AUC and recall. These measurements are widely used because their comprehensive nature in evaluating model performance. These measurements give comprehensive assessment for the model's performance through taking into account TP, TN, FP and FN cases. These measurements enable researchers to assess the model performance and shortcomings and work on its improvement for addressing the issue [36].

1. Accuracy (ACC): The model's accuracy represents the percentage of correct predictions, calculated by the ratio of correct forecasts to total predictions. This measurement is used in classification problems with balanced classes and relatively equal sample sizes. In cases of imbalanced classes, a model that consistently predicts the majority class may have high accuracy even if it is not beneficial, eq (1) [37].

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

2. Precision: Precision is useful in high-cost false-positives. False-positives can result in unneeded medical testing and treatments. Precision estimates how many positive predictions are correct, offering a more precise model performance rating, eq (2).

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

3. Recall: Recall is useful when false-negatives are expensive. False-negatives in disease detection can delay treatment, worsening outcomes. Recall evaluates how many positive examples were properly detected, offering a more detailed model performance evaluation, eq (3).

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

4. F-Measure: The F-Measure is useful. A weighted harmonic mean of precision and recall gives a fairer assessment of the model's performance, eq (4) [38].

$$F1 = \frac{2*(PR*RR)}{PR+RR} \quad (4)$$

5. ROC-AUC: The ROC-AUC helps to identify positive and negative examples. It displays the trade-off between the TPR and FPR at different classification criteria. The area under the curve (AUC) shows the model's performance regardless of the threshold, eq (5).

$$ROC - AUC = Area\ under\ the\ ROC\ curve \quad (5)$$

The false-positive rate (FPR), false-negative rate (FNR), true positive rate (TPR), true negative rate (TNR), cross entropy €, geometric mean (G-mean), and Matthews correlation coefficient (MCC) are also used in machine learning, but they differ from the five previously mentioned metrics

(accuracy, precision, recall, F-measure, ROC-AUC) in what they target and how they evaluate the model's performance. [39].

6. False-Positive Rate (FPR): assesses the model's classification of negative cases. It helps in spam and fraud detection when false-positives are costly, eq (6).

$$FPR = \frac{FP}{FP+TN} \quad (6)$$

7. False-Negative Rate (FNR): measures the model's false-negative rate. Medical diagnosis and detection of diseases benefit from it, as false-negatives are expensive, eq (7).

$$FNR = \frac{FN}{TP+FN} \quad (7)$$

8. True Positive Rate (TPR) or Detection Rate (DR): evaluates the model's positive instance accuracy. In medical diagnosis and disease detection, it helps identify true positives, eq (8).

$$TPR = \frac{TP}{TP+FN} \quad (8)$$

9. True Negative Rate (TNR): measures the model's negative instance accuracy. In credit risk assessment and fraud detection, it helps identify actual negatives, eq (9).

$$TNR = \frac{TN}{FP+TN} \quad (9)$$

10. Cross Entropy €: measures the difference between predicted and actual probability distributions. Multiclass classification problems employ it to optimize model performance, eq (10).

$$€ = -\frac{1}{n}\sum(y * \log(p) + (1-y) * \log(1-p)) \quad (10)$$

11. Geometric Mean (G-Mean): G-Mean measures the model's balance between positive and negative identification. Disease diagnostics and credit risk assessment benefit from it, eq (11).

$$G - Mean = \sqrt{TPR * TNR} \quad (11)$$

12. Matthews: Correlation coefficient (MCC) is a set of data correlation coefficients. It balances the model performance when the dataset is inconsistent, eq (12).

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}} \quad (12)$$

Table 4. Evaluation metrics

Ref.	Accuracy Evaluation Metrics												
	ACC	PR	RR	F-Score or F1	FPR	FNR	TPRor DR	TNR	Cross Entropy €	Fbeta score	ROC-AUC	G-mean	MCC
[6]	✓	✓	✓	✓	✓								
[7]	✓	✓	✓	✓	✓	✓	✓						
[8]	✓	✓	✓	✓	✓								
[9]	✓	✓	✓	✓									
[10]	✓	✓	✓	✓	✓								
[11]	✓	✓	✓	✓	✓								
[12]	✓	✓	✓		✓								
[13]	✓	✓				✓		✓					
[14]	✓	✓	✓	✓	✓	✓				✓			

[15]							✓						
[16]	✓	✓	✓	✓	✓	✓							
[17]					✓							✓	
[18]	✓	✓	✓	✓	✓	✓							
[19]	✓	✓	✓	✓		✓						✓	
[20]			✓	✓	✓						✓	✓	
[21]	✓	✓	✓	✓	✓								
[22]	✓	✓	✓	✓	✓	✓	✓		✓				
[23]	✓	✓	✓	✓	✓	✓	✓						
[24]	✓	✓	✓	✓	✓								
[25]	✓	✓	✓	✓	✓								
[26]	✓	✓	✓	✓	✓								
[27]	✓	✓	✓	✓	✓	✓							✓
[28]	✓	✓	✓		✓								
[29]			✓	✓	✓								
[30]	✓	✓	✓	✓	✓								
[31]	✓	✓		✓		✓			✓			✓	
[32]	✓	✓											
[33]			✓		✓							✓	

Each metric evaluates a different model performance facet. The problem and researcher or user's needs determine the measure used. using a combination of evaluation metrics that can capture different aspects and dimensions of prediction models, such as accuracy, precision, recall, f1-score, roc-auc, confusion matrix, etc. We also propose to use some additional metrics that can complement or enhance the existing metrics, such as false positive rate, false negative rate, true negative rate, cross entropy, geometric mean, and Matthews correlation coefficient, the authors used (ACC, PR, F-Score, FPR) to evaluate their phishing email detection model based on hybrid semantic deep learning (HSDL). They achieved an accuracy of (99.98 99.45 99.01 99.21) respectively on their dataset. And in [35], They achieved (accuracy of 99.69% for reflection attacks, 99.94% of exploitation attacks) by using this combination of metrics. in [37], the authors used (ACC, PR, RR, F-Score or F1, FPR, FNR, TPR) to evaluate their malware detection model based on deep embedded neural network expert system (DeNNeS). They achieved a ((ACC of 99% and 96.7% and FPR of 0.8% and 1.8%, for each model) for phishing, (99:7% and 90:6% ACC, and FPR of 0:3% and 8:8%, for each model) for Android malware. in [35], the authors used recall to evaluate their fake news detection model based on gated recurrent units (GRU), recurrent neural networks (RNN), naive Bayes (NB), and sequential minimal optimization (SMO). They achieved (accuracy of 99.69% for reflection attacks, 99.94% of exploitation attacks). in [22], the authors used f-measure to evaluate their cyberbullying detection model based on hybrid optimization algorithm HHO-PSO-DLNN (Harris Hawks Optimization, Particle Swarm Optimization and Deep Learning Neural Network). They achieved an f-measure of 98.9% on their dataset. in [23], the authors used roc-auc to evaluate their network intrusion detection model based on tuned vector convolutional deep neural network (TVCDNN). They achieved a roc-auc of 99.8% on their dataset.

CONCLUSION

Cyberattack prediction is a crucial and challenging task for ensuring the security and reliability of various systems and networks. We have presented a systematic literature review and a taxonomy of the existing cyberattack prediction systems. The outcome of the structured evaluation of techniques for predicting cyberattacks, is that we recommend the best, second-best and third-best models based on our analysis and explain why they are better for cyberattack prediction, that are used in the reviewed papers based on analysis. This taxonomy concerned with predicting cyberattacks using AI-based techniques,

such as deep learning, natural language processing, and graph neural networks. These techniques can accurately detect and predict multiple types of cyberattacks that can affect various domains and systems. Therefore, it is essential to develop prediction models that can handle multiple types of cyberattacks and provide timely and accurate responses. We also hope that our paper can inspire new ideas and directions for future research in this field. Some of the possible future research topics are as follows:

- Propose a novel taxonomy for machine learning techniques for predicting cyberattacks, based on four dimensions: datasets, techniques, challenges, and future directions (DTCF taxonomy), as well as identify the gaps and opportunities for future research in this domain.

- Analysing diverse cyber datasets that can capture the dynamic and complex nature of cyberattacks, as well as address the issues of privacy, ethics, and availability.

- Using more advanced evaluation methods that can capture the trade-off between different metrics or the cost of different types of errors.

- Exploring the applications of cyberattack prediction in various domains, such as IoT, network, industry, etc., as well as the challenges and opportunities in these domains.

REFERENCES

[1] Barik K, Misra S, Konar K, Fernandez-Sanz L, Koyuncu M. Cybersecurity deep: approaches, attacks dataset, and comparative study. Applied Artificial Intelligence. 2022 Dec 31;36(1):2055399. https://doi.org/10.1080/08839514.2022.2055399

[2] Asl MR, Naderi H. Filter Spamming in Computer Networks by Text Mining and Machine Learning Method. International Academic Journal of Science and Engineering. 2016;3(2):146-60.

[3] Hussain Ali Y, Sabu Chooralil V, Balasubramanian K, Manyam RR, Kidambi Raju S, T. Sadiq A, Farhan AK. Optimization system based on convolutional neural network and internet of medical things for early diagnosis of lung cancer. Bioengineering. 2023 Mar 2;10(3):320. https://doi.org/10.3390/bioengineering10030320

[4] Leema AA, Balakrishnan P, Jothiaruna N. Harnessing the power of web scraping and machine learning to uncover customer empathy from online reviews. Indian Journal of Information Sources and Services. 2024;14(3):52-63. https://doi.org/10.51983/ijiss-2024.14.3.08

[5] Dixit P, Kohli R, Acevedo-Duque A, Gonzalez-Diaz RR, Jhaveri RH. Comparing and analyzing applications of intelligent techniques in cyberattack detection. Security and Communication Networks. 2021;2021(1):5561816. https://doi.org/10.1155/2021/5561816

[6] Bamal S, Singh L. Detecting Conjunctival Hyperemia Using an Effective Machine Learning based Method. J. Internet Serv. Inf. Secur. 2024;14(4):499-510. https://doi.org/10.58346/JISIS.2024.I4.031

[7] Ahsan M, Nygard KE, Gomes R, Chowdhury MM, Rifat N, Connolly JF. Cybersecurity threats and their mitigation approaches using Machine Learning—A Review. Journal of Cybersecurity and Privacy. 2022 Jul 10;2(3):527-55. https://doi.org/10.3390/jcp2030027

[8] Dasari, D. R., & Bindu, G. H. Feature Selection Model-Based Intrusion Detection System for Cyberattacks on the Internet of Vehicles Using Cat and Mouse Optimizer. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 2024;15(2):251-269. https://doi.org/10.58346/JOWUA.2024.I2.017

[9] Prabhakaran V, Kulandasamy A. Hybrid semantic deep learning architecture and optimal advanced encryption standard key management scheme for secure cloud storage and intrusion detection. Neural Computing and Applications. 2021 Nov;33(21):14459-79.

[10] Ramprasath J, Ramya P, Rathnapriya T. Malicious attack detection in software defined networking using machine learning approach. International Journal of Advances in Engineering and Emerging Technology. 2020 Jul 31;11(1):22-7.

[11] Al-zubidi AF, Farhan AK, Towfek SM. Predicting DoS and DDoS attacks in network security scenarios using a hybrid deep learning model. Journal of Intelligent Systems. 2024 Apr 23;33(1):20230195. https://doi.org/10.1515/jisys-2023-0195

[12] Alnumay WS. Use of machine learning for the detection, identification, and mitigation of cyber-attacks. International Journal of Communication and Computer Technologies. 2024;12(1):38-44.

[13] Maseer ZK, Yusof R, Bahaman N, Mostafa SA, Foozy CF. Benchmarking of machine learning for anomaly based intrusion detection systems in the CICIDS2017 dataset. IEEE access. 2021 Feb 3;9:22351-70.

[14] Kumar, P.; et al. Toward design of an intelligent cyberattack detection system using hybrid feature reduced approach for iot networks. Arab. J. Sci. Eng. 2021, 46, 3749-3778. https://doi.org/10.1109/ACCESS.2021.3056614

[15] Wu P, Guo H. LuNET: a deep neural network for network intrusion detection. In2019 IEEE symposium series on computational intelligence (SSCI) 2019 Dec 6 (pp. 617-624). IEEE.

[16] Nagarajan SM, Deverajan GG, Bashir AK, Mahapatra RP, Al-Numay MS. IADF-CPS: Intelligent anomaly detection framework towards cyber physical systems. Computer Communications. 2022 Apr 15;188:81-9. https://doi.org/10.1016/j.comcom.2022.02.022

[17] Lv L, Wang W, Zhang Z, Liu X. A novel intrusion detection system based on an optimal hybrid kernel extreme learning machine. Knowledge-based systems. 2020 May 11;195:105648. https://doi.org/10.1016/j.knosys.2020.105648

[18] Tufan E, Tezcan C, Acartürk C. Anomaly-based intrusion detection by machine learning: A case study on probing attacks to an institutional network. IEEE Access. 2021 Mar 26;9:50078-92. https://doi.org/10.1109/ACCESS.2021.3068961

[19] Disha RA, Waheed S. Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. Cybersecurity. 2022 Jan 4;5(1):1.

[20] Shahzad F, Mannan A, Javed AR, Almadhor AS, Baker T, Al-Jumeily OBE D. Cloud-based multiclass anomaly detection and categorization using ensemble learning. Journal of Cloud Computing. 2022 Nov 3;11(1):74.

[21] Mouti S, Shukla SK, Althubiti SA, Ahmed MA, Alenezi F, Arumugam M. Cyber Security Risk management with attack detection frameworks using multi connect variational auto-encoder with probabilistic Bayesian networks. Computers and Electrical Engineering. 2022 Oct 1;103:108308. https://doi.org/10.1016/j.compeleceng.2022.108308

[22] Nayak J, Meher SK, Souri A, Naik B, Vimal S. Extreme learning machine and bayesian optimization-driven intelligent framework for IoMT cyber-attack detection. The Journal of Supercomputing. 2022 Sep;78(13):14866-91.

[23] Ur Rehman S, Khaliq M, Imtiaz SI, Rasool A, Shafiq M, Javed AR, Jalil Z, Bashir AK. DIDDOS: An approach for detection and identification of Distributed Denial of Service (DDoS) cyberattacks using Gated Recurrent Units (GRU). Future Generation Computer Systems. 2021 May 1;118:453-66. https://doi.org/10.1016/j.future.2021.01.022

[24] Ahmad Z, Shahid Khan A, Nisar K, Haider I, Hassan R, Haque MR, Tarmizi S, Rodrigues JJ. Anomaly detection using deep neural network for IoT architecture. Applied Sciences. 2021 Jul 30;11(15):7050. https://doi.org/10.3390/app11157050

[25] Mahdavifar S, Ghorbani AA. DeNNeS: deep embedded neural network expert system for detecting cyber attacks. Neural Computing and Applications. 2020 Sep;32(18):14753-80. https://doi.org/10.1007/s00521-020-04830-w

[26] Alshingiti Z, Alaqel R, Al-Muhtadi J, Haq QE, Saleem K, Faheem MH. A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN. Electronics. 2023 Jan 3;12(1):232. https://doi.org/10.3390/electronics12010232

[27] Al-Abassi A, Karimipour H, Dehghantanha A, Parizi RM. An ensemble deep learning-based cyber-attack detection in industrial control system. Ieee Access. 2020 May 4;8:83965-73. https://doi.org/10.1109/ACCESS.2020.2992249

[28] Saharkhizan M, Azmoodeh A, Dehghantanha A, Choo KK, Parizi RM. An ensemble of deep recurrent neural networks for detecting IoT cyber attacks using network traffic. IEEE Internet of Things Journal. 2020 May 21;7(9):8852-9. https://doi.org/10.1109/JIOT.2020.2996425

[29] Dutta V, Choraś M, Pawlicki M, Kozik R. A deep learning ensemble for network anomaly and cyber-attack detection. Sensors. 2020 Aug 15;20(16):4583. https://doi.org/10.3390/s20164583

[30] Aljabri M, Aljameel SS, Mohammad RM, Almotiri SH, Mirza S, Anis FM, Aboulnour M, Alomari DM, Alhamed DH, Altamimi HS. Intelligent techniques for detecting network attacks: review and research directions. Sensors. 2021 Oct 25;21(21):7070. https://doi.org/10.3390/s21217070

[31] Yan T, Zhou T, Zhan Y, Xia Y. TFDPM: Attack detection for cyber–physical systems with diffusion probabilistic models. Knowledge-Based Systems. 2022 Nov 14;255:109743. https://doi.org/10.1016/j.knosys.2022.109743

[32] Al-Juboori SA, Hazzaa F, Jabbar ZS, Salih S, Gheni HM. Man-in-the-middle and denial of service attacks detection using machine learning algorithms. Bulletin of Electrical Engineering and Informatics. 2023 Feb 1;12(1):418-26. https://doi.org/10.11591/eei.v12i1.4555

[33] Sangodoyin AO, Akinsolu MO, Pillai P, Grout V. Detection and classification of DDoS flooding attacks on software-defined networks: A case study for the application of machine learning. IEEE Access. 2021 Aug 31;9:122495-508. https://doi.org/10.1109/ACCESS.2021.3109490

[34] Alsaedi EM, Farhan AK. RCAE_BFV: Retrieve Encrypted Images Using Convolution AutoEncoder and BFV. Iraqi Journal of Computers, Communications, Control & Systems Engineering. 2022;22(3):48–61. https://doi.org/10.33103/uot.ijccce.22.3.5

[35] Mohammed, A. A., Al-Ghrairi, A. H. T., Al-zubidi, A. F., & Saeed, H. M. (2023, February). Unsupervised classification and analysis of Istanbul-Turkey satellite image utilizing the remote sensing. In *AIP Conference Proceedings* (Vol. 2457, No. 1, p. 040007). AIP Publishing LLC. https://doi.org/10.1063/5.0118339

[36] Ameen ZH, AL-Bakri NF, Al-zubidi AF, Hashim SH, Jaaz ZA. A New COVID-19 Patient Detection Strategy Based on Hidden Naïve Bayes Classifier. Iraqi Journal of Science. 2024 Nov 30:6705-24. https://orcid.org/0000-0001-6870-8572

[37] Saadi ZM, Sadiq AT, Akif OZ, Farhan AK. A survey: Security vulnerabilities and protective strategies for graphical passwords. Electronics. 2024 Aug 1;13(15):3042. https://doi.org/10.3390/electronics13153042

[38] Al-zubidi AF, Farhan AK, El-Kenawy ES. Surveying Machine Learning in Cyberattack Datasets: A Comprehensive Analysis. Journal of Soft Computing and Computer Applications. 2024;1(1):1. https://doi.org/10.70403/3008-1084.1000

[39] Al-zubidi AF, Farhan AK. Multi-Class Anomaly Detection in Network Intrusion Detection Using Variational Autoencoder. International Journal of Safety & Security Engineering. 2025 Jun 1;15(6).

[40] Alsaedi EM, Farhan AK, Falah MW, Oleiwi BK. Classification of encrypted data using deep learning and Legendre polynomials. InThe International Conference on Innovations in Computing Research 2022 Aug 11 (pp. 331-345). Cham: Springer International Publishing.

[41] Abdalrdha, Z. K., Al-Bakry, A. M., & Farhan, A. K. (2023, December). Improving the CNN Model for Arabic Crime Tweet Detection Based on an Intelligent Dictionary. In 2023 16th International Conference on Developments in eSystems Engineering (DeSE) (pp. 748-753). IEEE.

[42] Al-zubidi, A. F., Farhan. Multi-Class Semi-Supervised Anomaly Detection for Network Intrusion Detection: A Novel Variational Autoencoder Approach. International Journal of Safety & Security Engineering. 2025;15(6).

[43] Gupta R, Tanwar S, Tyagi S, Kumar N. Machine learning models for secure data analytics: A taxonomy and threat model. Computer Communications. 2020 Mar 1;153:406-40. https://doi.org/10.1016/j.comcom.2020.02.008

[44] Dahiya M, Nitin N, Dahiya D. Intelligent cyber security framework based on SC-AJSO feature selection and HT-RLSTM attack detection. Applied Sciences. 2022 Jun 21;12(13):6314. https://doi.org/10.3390/app12136314

[45] Sarker IH. CyberLearning: Effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks. Internet of Things. 2021 Jun 1;14:100393. https://doi.org/10.1016/j.iot.2021.100393