Performance Analysis of Various Machine Learning based Algorithms on Cyber security Approaches

[1]Mr.VS.Pavan Kumar, [2]Dr.S.Arivalagan, [3]Dr.M.Murugesan, [4]Dr.P.Sudhakar

[1]Research Scholar, [2]Assistant professor, [3]Professor, [4]Associate Professor,
Department of CSE,
[1,2,4,]Annamalai University, [3] Anurag Engineering College

Abstract- **Pervasive use and development of the Internet and its mobile applications extended cyberspace. Cyberspace is prone to prolong and automated cyber-attacks. Cyber security methods render advancements in security measures to find cyberattacks. Conventional security systems are ineffective as cybercriminals were smart enough to avoidclassical security systems. Traditional security system is ineffective in identifying polymorphic security attacks. Machine learning (ML) approaches had a significant contributiontovarious applications of cybersecurity. In spite of the success, there exist certain difficultiesin assuringthe reliability of the ML mechanism. There were incentivized malicious adversariespresented in cyberspace that are ready to use these ML vulnerabilities. This study offers a detailed examination of various ML models to detect cyberattacks and accomplish cybersecurity. This study presents a detailed discussion of existing ML models for cyber security comprising intrusion detection, spam detection, and malware detection in recent days. In addition, the basic concepts of cybersecurity and cyberattacks are elaborated in detail. In addition, we have discussed the existing ML models for cybersecurity along with their aim, methodology, and experimental data. At the end of the study, a detailed overview of cybersecurity, cyberattacks, and recent cyberattack detection models are elaborated briefly.**

Keywords: **Cybersecurity; Machine learning; Cyberattacks; Data driven models; Security; Artificial Intelligence**

1. INTRODUCTION

In the modern era, the Internet is becoming a crucial one and needed in everybody's life making this interlinked network prone to variousmenaces[1]. There are many security threats in cyber-space such as jail-breaking, two-faced malware intrusion, and network intrusion. Such menaces would affect the security of networks or devices [2]. Most security companies across the world were focused on devising novel technologies for protecting software applications, computer devices, and networks from malware infections and network intrusion attacks. Cyber-attacks were less risky, cheaper, and convenient compared to physical attacks [3]. Cyber criminals just have some expenses beyond an Internet connection and a computer. They were unconstrained by distance and geography. Owing to the anonymous nature of Internet, it is tough to prosecute and find [4]. It is noted that assaults against information technology will be very attractive, it is anticipated that complicated cyber-attacks will keep on increasing. Fig. 1 represents the infrastructure of cybersecurity.

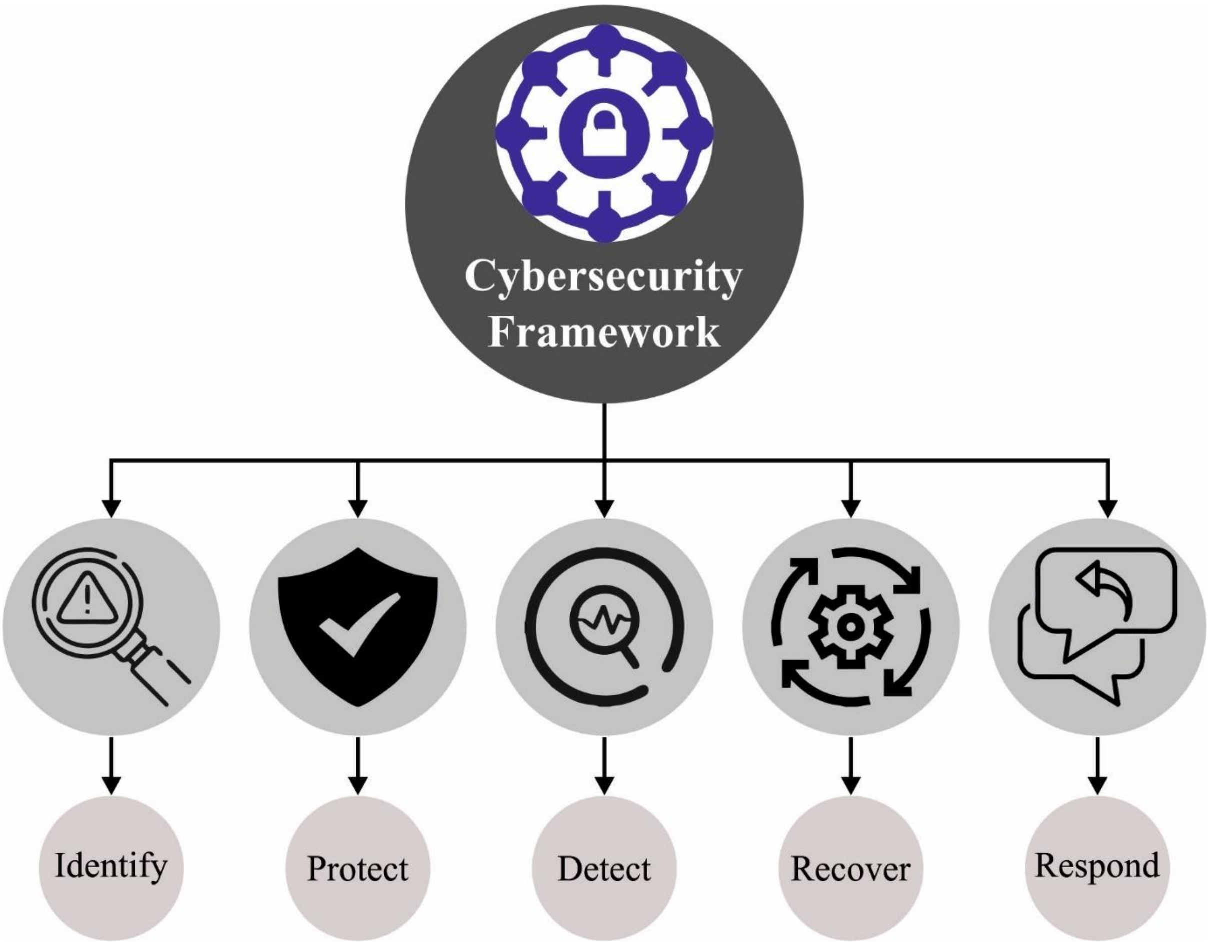

Fig. 1. Architecture of cybersecurity

A stable and secure computer mechanism should assure the integrity, confidentiality, and availability of data [5]. The security and integrity of a computer system were compromised if unauthorized program or individual enters a network or computer intends to disrupt or harm the normal action flow [6]. Cybersecurity can be defined as the security measures that are considered for protecting user assets and cyberspace against unlawful attacks and access. [7]. Internal and Inherent weakness in the implementation and configuration of network and computer forms vulnerability that is prone to threats and cyberattacks. Some instances of vulnerabilities in framing anetwork system are amateur or untrained personnel, incorrect configuration, and lack adequate process [8]. Such susceptibilities will mount the chances of attacks and threats from outside or within a network.

Several people from various domains are increasingly reliant on cyber networks. In simple terms, threat can be defined as an agent that causes undesirable and harmful effects on behaviour and actions of a network or computer using a specific penetration method [9]. Cybersecurity was to defend integrity of programs, data, and networks from cyberthreats to cyberspace.

Cybersecurity considers nearby problems of different cyber-attacks and modelling defense methods (countermeasures) that protect integrity, availability, and confidentiality of information and digital technology [10].

- The term Integrity was employed to thwart any deletion or modification in an unauthorized way.
- Confidentiality is exploited to prevent the disclosure of data to illegal systems or personnel.
- Availability is leveraged to ensure that systems accountable for processing, distributing, and saving data is accessible if required and by those who want them.

Several cyber-security specialists trust that malware will be the main tool for executing malevolent plans to breach cybersecurity efforts in cyberspace.

This study offers a detailed examination of various ML models to detect cyberattacks and accomplish cybersecurity. This study presents a detailed discussion of existing ML models for cybersecurity comprising intrusion, spam, and malware detections in recent days. In addition, the basic concepts of cybersecurity and cyberattacks are elaborated in detail. In addition, we have discussed the existing ML models for cybersecurity along with its aim, methodology, and experimental data. At the end of the study, a detailed overview of cybersecurity, cyberattacks, and recent cyberattack detection models are elaborated briefly.

2. Background Information

In this section, the relevant technology of cybersecurity data science involving different kinds of defense strategies and cybersecurity incidents is discussed.

Cybersecurity

In recent times, the information and communication technologies (ICTs) has changed dramatically, which is pervasive and intrinsically connected to the modern world. Therefore, defending ICT applications and systems from cyberattacks has been very much concerned for the security policymaker over the last few years. The act of defending ICT system from different cyberattacks or threats is called cybersecurity. Various aspects are related to cybersecurity: measures to secure ICT; the raw information and data it comprises and their transmitting and processing; related physical and virtual components of the system; the degree of security originating from applications of those measures; and finally, the related domain of expert endeavour. Generally, cybersecurity is concerned with understanding different cyberattacks and developing equivalent security systems that preserve numerous properties as follows.

- Confidentiality is leveragedto thwart the disclosure and access of data towards unauthorized entity, system, or individual.
- Integrity can be employed to preserve any destruction or modification of data by unauthorized means.
- Availability was utilizedto guarantee reliable and prompt access of systems and data assets to authorized entities.

Cybersecurity is used in different contexts, from commercialpurposes to mobile computing, and is split into different classes. This network security focuses primarily on protecting a network from intruders or cyberattackers; application privacy that keeps the devices and the software free from cyber-threats or risks; data privacy considers the privacy and security of pertinent information; operational security that involves the process of protecting and handling data resources. Traditionalcybersecurity system is made up of computer and network security system encompassing antivirus software, intrusion detections, or firewall system.

Cyberattacks and security risks

Typically, the risk related to any attack considers three privacy factors like impact include what the attack does, threats, viz., who is attacking, and vulnerability includes the weaknesses they are attacking. A security incident was an act which threaten the CIA of systems and data assets. Different kinds of cybersecurity incidents cause privacy risksto the individual or systems and networks of the organization.

They includes:

- Unauthorized access describes the act of accessing data to systems, network, or information without authorization which causes violation of privacy policy;
- Malware named malicious software, is software or program that is intended on purpose to create damage to the server, computer network, client, or computer, for example, botnets. Instances of distinct kinds of malware involving Trojan horses, computer worms, ransomware, viruses, adware, malicious bots, spyware, and so on; Ransomware, or Ransom malware, is a novel form of malware that thwarts userfrom accessing devices, personal files or systems, then demand an anonymous online payment for restoring the access.
- Denial-of-Service (DoS) can be referred to an attack intended to shut-down a network or machine, which makes it unreachable to its intended use by flooding target with traffic which causes a crash. Normally, the DoS assault employs one computer with an Internet connection, whereas distributed denial-of-service (DDoS) attacks use more than one computer and Internet connection to flood targetresources;
- Phishing is a kind of social engineering, leveragedfor a wide-ranging malevolent activity attained via human interaction, where fake attempts take part to achieve delicate data namely login credentials, personally identifiable information, credit card and banking details disguising oneself as a trusted entity, or individual through an electronic communications namely instant message, email, text, and so on.
- Zero-day attack is employedto define the menace of unknown security vulnerabilities where the patch hasn't been released or the application developer was not aware.

Cybersecurity defense strategies

Defense strategy was essential to secure information or data, networks, and information systems from intrusions or cyber-attacks. They take the responsibility to prevent security incidents or data breaches and reacting and monitoring to intrusions, that is determined by any sort of unauthorized activities which deteriorates the information system. Typically, intrusion detection systems (IDS) can be denoted by "software application or device that monitor systems or computer network for policy violations or malicious activity". The most common security solutions include user authentication, antivirus, firewalls, cryptography systems, access control, and data encryption however ineffective based on requirements in the cyber field. At the same time, IDS overcomes the problem by examining security information from numerous key points in a system or computer networks. Furthermore, IDS is utilized to find internal and external attacks. For example, a network IDS (NIDS), and host-based IDS (HIDS) are the renowned kinds related to scope of single computer to larger network. In HIDS, the system monitors essential files on single system, where it monitors and analyses network connection for suspicious traffic in NIDS. Likewise, anomaly-orientedIDS and signature-oriented IDS, are the two commonest variants.

3. Analysis of Various ML Models for Cybersecurity

In this study, we have investigated the performance of different ML models to find cyberattacks and accomplish cybersecurity. This study presented a comprehensive discussion of existing ML models for cybersecurity encompassing intrusion detection, spam detection, and malware detection in recent days.

Cui et al. [11] formulate a flexible ML detection algorithm for cyber-attacks in distribution systems taking spatiotemporal patterns into account. By the graph Laplacian related to system wide measurements, the abovementioned patterns can be detected. In addition, to train spatiotemporal patterns, a flexible Bayes classifier (BC) was employed which is violated if cyberattacks occur. Cyber-attacks will be identified by making use of flexible BC online. An et al. [12] suggest using unsupervised ensemble AE linked to the Gaussian mixture model (GMM) for adaption to many fields irrespective of the skewness of all domains. The attention-oriented latent representations and reconstructed attributes of the minimal error were used in the hidden space of the ensemble AE. To predict the sample density in the GMM, the expectation maximization (EM) approach was employed.

Almalaq et al. [13] proposed an attack detection technique based on DL for energy systems to solve this problem, which is trained through logs and information collected by phasor measurement unit (PMU). Specification or Property making was employed for constituting features, and data can be forwarded to several ML approaches, out of these RF was chosen as the fundamental technique of AdaBoost.

Avatefipour et al. [14] devise an innovative and effective anomaly detection (AD) method dependent on a modified one class SVM in the CAN traffic. This presented technique uses an enhanced method called the modified bat technique for identifying the structure with maximum accuracy level in offline training. The authors in [15] introduced a precise secured framework to find and halt data integrity assaults in WSNs in microgrids. An intellectual AD technique relevant to predictive intervals (PIs) is presented for differentiating malicious assaults with distinct severities at the rime of a secured operation. The devised AD technique was framed depending on the upper and lower bound prediction approach for offering best practicable PIs on the smart meter readings at electric customers. It even uses the combinatorial idea of PIs for solving the instability problems occur from the NNs.

Saheed and Arowolo [16] aim to illustrate how supervised ML approaches (ridge classifier, RF, KNN, and DT) and deep RNN are used to formulate an effective and efficient IDS in the IoMT platform for forecasting and classifying unexpected cyberthreats. Normalization and preprocessing of network data will be executed. Then, the researchers optimized features by making use of a bio-inspired PSO. In [17], the authors modelled an AD-IoT mechanism, which is an intellectual AD related to RF-ML method for solving the IoT cybersecurity threats in smart cities. The modelled solution will find compromised IoT gadgets at dispersed fog nodes effectually.

Kalech [18] devised cyber-attack detection approaches related to temporal pattern detection. This pattern detection approaches do not only search for anomalies in data sent by the SCADA elements on the network however searchers for anomalies that arise by exploiting legitimate commands so that incorrect and unauthorized time intervals amongst them can cripple the mechanism. In particular, the authors devise 2 approaches relevant to ANN and Hidden Markov Model (HMM).Wang et al. [19] devise an attack detection method for power systems related to ML that is trained through information and logs gathered by PMUs. The researchers execute feature construction engineering and transfer the information to various ML methods, where RF was selected as the fundamental technique of AdaBoost.

In [20], a new method has been offered for diagnosing possible false data injection attacks (FDIA) in DC-MGs for improving the cybersecurity of electrical systems. So, to find cyberattacks in DC-MG and to find the FDIA to distributed energy resources (DERs) unit, a novel singular value decomposition (SVD) and wavelet transform (WT) procedure related to deep ML has been modelled. Furthermore, this study renders a devised selective ensemble DL method by utilizing the GWO algorithm for finding the FDIA in DC-MG. In [21], an effective and efficient security control technique was modelled for identifying cyber-attacks on smart grids. This modelled technique will combine feature detection and reduction methods to minimize the more features and attain an enhanced detection rate. To eliminate irrelevant features and to improve detection performance, a correlation-oriented feature selection (CFS) algorithm was employed. An instance-based learning (IBL) method will classify cyberattacks and normal events through the chosen optimal features. Elkhadir et al. [22] present a novel variant of PCA called QR-OMPCA. Initially, this technique will integrate the mean calculation into feature extraction operation, thereby the optimal mean is acquired to improve the intrusion detection accurateness. Then, it includes a rapid QR decomposition.

Chen et al. [23] devise resilient function methods for non-linear processes that are prone to target cyber-attacks, along with that detection and managing standard types of cyberattacks. Cui et al. [24] formulated an ML-oriented AD (MLAD) method. Initially, load predictions offered by NN were employed for reconstructing the scaling and benchmark dataset by making use of k-means clustering. Then, the cyberattack template was predicted by NB classification related to the statistical features and cumulative distribution operation of scaling data. Lastly, dynamic programming was employed for computing the parameter and occurrence of single cyberattack on load prediction datasets.

In [25], the researchers devise a physics-guided ML for detecting cyberattacks on intrusion detection EVs taking changing driving scenarios into account. To reflect the transient physical features of EV, the researchersgather device-level (for example voltage and current in the motor drive) and vehicle level signals. After that, new data features regarding physical dynamics of the vehicle and critical performance of a system were devised, with which, the authors use data-driven approach with high-fidelity vehicular methods and physical power electronics.

Kravchik and Shabtai [26] introduces work on finding cyberattacks on industrial control system (ICS) utilizing CNNs. The authors propose a technique for AD related to measurement of the statistical deviation of estimated value from the monitored value. The researchers implemented the devised approach by utilizing various DNN structure which includes distinct variants of CNN and RNN. In [27], SVM is considered to be an ML approach that could complement efficiency of this IDS, offering a second line of recognition to minimize the false alarm count, or as an alternative detection method. The authors evaluate the efficiency of this IDS against two-class and one-class SVM, utilizing non- linear and linear forms.

4. Results and Discussion

This section examines the cybersecurity performance of different ML models.Table 1 shows the overall performance of different models available in the literature.Fig. 2 examines a comparative $prec_n$ and $reca_l$ examination of different cybersecurity approaches. The results implied that these methods have offered reasonable outcomes. Based on $prec_n$, the PSO-KNN model has gained higher $prec_n$ of 98.89% while the PSO-RC, CNN, LCNN, HaRM, CAN-ML, and flexible ML-CDSP models have obtained lower $prec_n$ of 97.60%, 93.54%, 93.68%, 91.99%, 92.39%, and 94.24% respectively. Also, in terms of $reca_l$, the PSO-KNN system has reached superior $reca_l$ of 94.56% while the PSO-RC, CNN, LCNN, HaRM, CAN-ML, and flexible ML-CDSP methods have acquired lesser $reca_l$ of 95.32%, 94%, 93.68%, 92.17%, 94.83%, and 94.09% correspondingly.

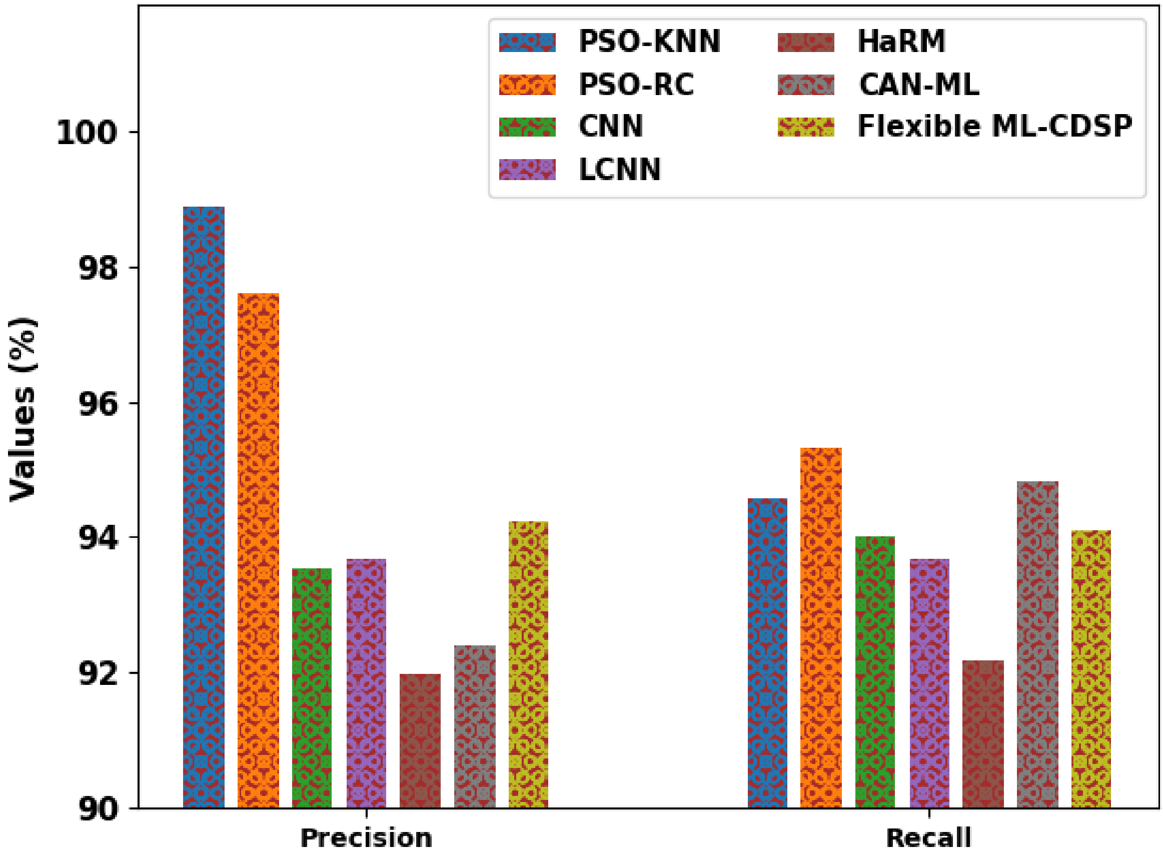

Fig. 2.$Prec_n$ and $reca_l$ analysis of various methodologies

Table 1 Comparative analysis of various approaches and measures

Methods	**Accuracy**	**Precision**	**Recall**	**F-Score**	**MCC**
PSO-KNN	98.90	98.89	94.56	92.33	97.77
PSO-RC	97.61	97.60	95.32	91.06	95.14
CNN	92.00	93.54	94.00	93.65	93.88
LCNN	94.00	93.68	93.68	92.68	93.34
HaRM	92.21	91.99	92.17	92.23	92.57
CAN-ML	94.89	92.39	94.83	92.02	92.46
Flexible ML-CDSP	95.85	94.24	94.09	94.15	94.17

Fig. 3 scrutinizes a comparative $accu_y$,F_{score} and MCC investigation of distinct cybersecurity approaches. The outcomes referred that these approaches have offered reasonable outcomes. With respect to $accu_y$, the PSO-KNN algorithm has achieved maximum $accu_y$ of 98.90% while the PSO-RC, CNN, LCNN, HaRM, CAN-ML, and flexible ML-CDSP models have attained lower $accu_y$ of 97.61%, 92%, 94%, 92.21%, 94.89%, and 95.85% correspondingly. Besides, interms of F_{score}, the PSO-KNN algorithm has gained higher F_{score} of 92.33% while the PSO-RC, CNN, LCNN, HaRM, CAN-ML, and flexible ML-CDSP systems have gained decreased F_{score} of 91.06%, 93.65%, 92.68%, 92.23%, 92.02%, and 94.15% correspondingly. Finally, based on MCC, the PSO-KNN model has gained superior MCC of 97.77% while the PSO-RC, CNN, LCNN, HaRM, CAN-ML, and flexible ML-CDSP methodologies have obtained lower MCC of 95.14%, 93.88%, 93.34%, 92.57%, 92.46%, and 94.17% correspondingly.

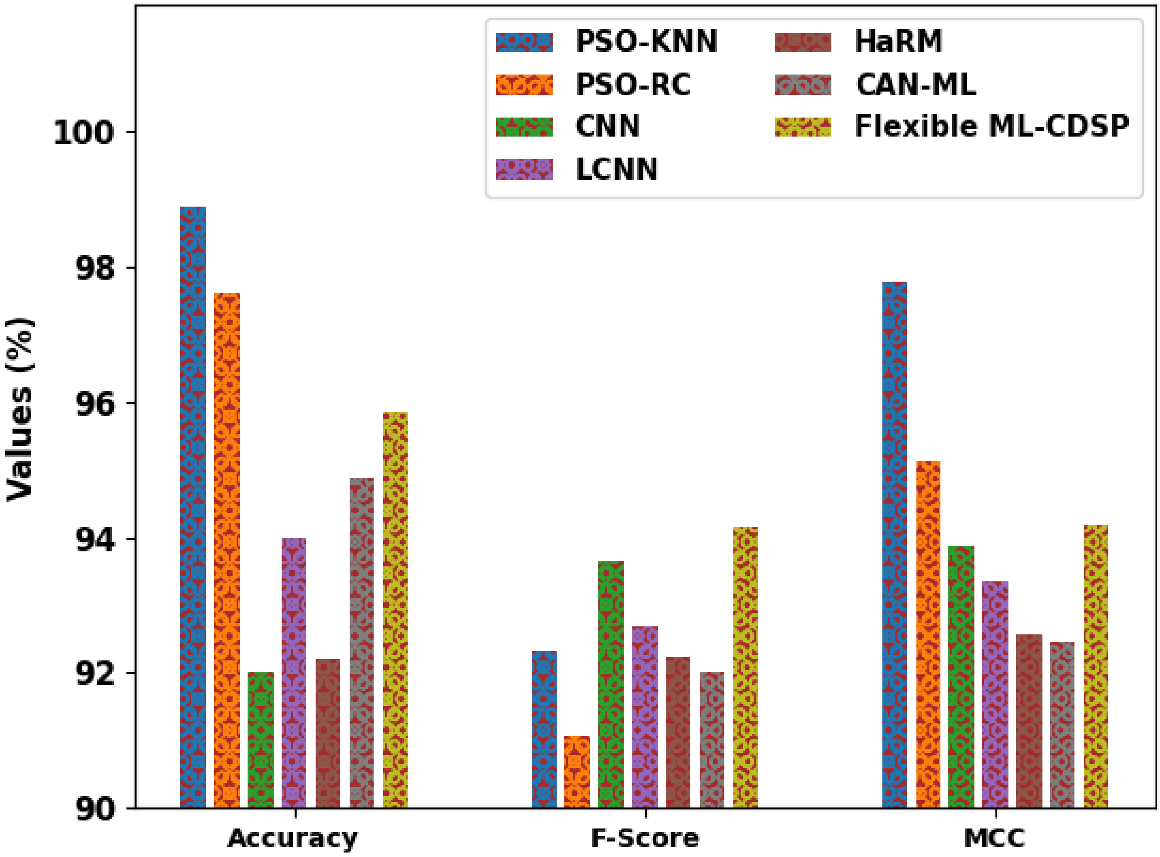

Fig. 3.$Accu_y$, F_{Score}, and MCC analysis of various methodologies

5. Conclusion

In this study, we have evaluatedthe performance of different ML models to detect cyberattacks and accomplish cybersecurity. This study presented a comprehensive discussion of existing ML models for cyber security encompassing intrusion detection, spam detection, and malware detection in recent days. Moreover, the basic concepts of cybersecurity and cyberattacks are elaborated in detail. In addition, we have deliberatedon the present ML models for cybersecurity along with their aim, methodology, and experimental data. At the end of the study, a detailed overview of cybersecurity, cyberattacks, and recent cyberattack detection models are elaborated briefly.

REFERENCES

1. Parizad, A. and Hatziadoniu, C., 2022. Cyber-Attack Detection Using Principal Component Analysis and Noisy Clustering Algorithms: A Collaborative Machine Learning-Based Framework. *IEEE Transactions on Smart Grid*.
2. Rashid, M.M., Kamruzzaman, J., Hassan, M.M., Imam, T. and Gordon, S., 2020. Cyberattacks detection in iot-based smart city applications using machine learning techniques. *International journal of environmental research and public health*, *17*(24), p.9347.
3. Alsamiri, J. and Alsubhi, K., 2019. Internet of things cyber attacks detection using machine learning. *International Journal of Advanced Computer Science and Applications*, *10*(12).
4. Zheng, H., Wang, Y., Han, C., Le, F., He, R. and Lu, J., 2018, August. Learning and applying ontology for machine learning in cyber attack detection. In *2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE)* (pp. 1309-1315). IEEE.
5. Alshehri, A., Khan, N., Alowayr, A. and Alghamdi, M.Y., 2023. Cyberattack Detection Framework Using Machine Learning and User Behavior Analytics. *COMPUTER SYSTEMS SCIENCE AND ENGINEERING*, *44*(2), pp.1679-1689.
6. Delplace, A., Hermoso, S. and Anandita, K., 2020. Cyber Attack Detection thanks to Machine Learning Algorithms. *arXiv preprint arXiv:2001.06309*.
7. Alsamiri, J. and Alsubhi, K., 2019. Internet of things cyber attacks detection using machine learning. *International Journal of Advanced Computer Science and Applications*, *10*(12).
8. Miao, Y., Chen, C., Pan, L., Han, Q.L., Zhang, J. and Xiang, Y., 2021. Machine learning–based cyber attacks targeting on controlled information: A survey. *ACM Computing Surveys (CSUR)*, *54*(7), pp.1-36.
9. Dutta, V., Choraś, M., Pawlicki, M. and Kozik, R., 2020. A deep learning ensemble for network anomaly and cyber-attack detection. *Sensors*, *20*(16), p.4583.

10. Komisarek, M., Pawlicki, M., Kozik, R. and Choras, M., 2021. Machine Learning Based Approach to Anomaly and Cyberattack Detection in Streamed Network Traffic Data. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, *12*(1), pp.3-19.
11. Cui, M., Wang, J. and Chen, B., 2020. Flexible machine learning-based cyberattack detection using spatiotemporal patterns for distribution systems. *IEEE Transactions on Smart Grid*, *11*(2), pp.1805-1808.
12. An, P., Wang, Z. and Zhang, C., 2022. Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection. *Information Processing & Management*, *59*(2), p.102844.
13. Almalaq, A., Albadran, S. and Mohamed, M.A., 2022. Deep machine learning model-based cyber-attacks detection in smart power systems. *Mathematics*, *10*(15), p.2574.
14. Avatefipour, O., Al-Sumaiti, A.S., El-Sherbeeny, A.M., Awwad, E.M., Elmeligy, M.A., Mohamed, M.A. and Malik, H., 2019. An intelligent secured framework for cyberattack detection in electric vehicles' CAN bus using machine learning. *IEEE Access*, *7*, pp.127580-127592.
15. Kavousi-Fard, A., Su, W. and Jin, T., 2020. A machine-learning-based cyber attack detection model for wireless sensor networks in microgrids. *IEEE Transactions on Industrial Informatics*, *17*(1), pp.650-658.
16. Saheed, Y.K. and 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*, pp.161546-161554.
17. Alrashdi, I., Alqazzaz, A., Aloufi, E., Alharthi, R., Zohdy, M. and Ming, H., 2019, January. Ad-iot: Anomaly detection of iot cyberattacks in smart city using machine learning. In *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0305-0310). IEEE.
18. Kalech, M., 2019. Cyber-attack detection in SCADA systems using temporal pattern recognition techniques. *Computers & Security*, *84*, pp.225-238.
19. Wang, D., Wang, X., Zhang, Y. and Jin, L., 2019. Detection of power grid disturbances and cyber-attacks based on machine learning. *Journal of information security and applications*, *46*, pp.42-52.
20. Dehghani, M., Niknam, T., Ghiasi, M., Bayati, N. and Savaghebi, M., 2021. Cyber-attack detection in dc microgrids based on deep machine learning and wavelet singular values approach. *Electronics*, *10*(16), p.1914.
21. Gumaei, A., Hassan, M.M., Huda, S., Hassan, M.R., Camacho, D., Del Ser, J. and Fortino, G., 2020. A robust cyberattack detection approach using optimal features of SCADA power systems in smart grids. *Applied Soft Computing*, *96*, p.106658.
22. Elkhadir, Z., Chougdali, K. and Benattou, M., 2017, November. An effective cyber attack detection system based on an improved OMPCA. In *2017 International Conference on Wireless Networks and Mobile Communications (WINCOM)* (pp. 1-6). IEEE.
23. Chen, S., Wu, Z. and Christofides, P.D., 2020. Cyber-attack detection and resilient operation of nonlinear processes under economic model predictive control. *Computers & Chemical Engineering*, *136*, p.106806.
24. Cui, M., Wang, J. and Yue, M., 2019. Machine learning-based anomaly detection for load forecasting under cyberattacks. *IEEE Transactions on Smart Grid*, *10*(5), pp.5724-5734.
25. Guo, L., Ye, J. and Yang, B., 2020. Cyberattack Detection for Electric Vehicles Using Physics-Guided Machine Learning. *IEEE Transactions on Transportation Electrification*, *7*(3), pp.2010-2022.
26. Kravchik, M. and Shabtai, A., 2018, January. Detecting cyber attacks in industrial control systems using convolutional neural networks. In *Proceedings of the 2018 workshop on cyber-physical systems security and privacy* (pp. 72-83).
27. Ghanem, K., Aparicio-Navarro, F.J., Kyriakopoulos, K.G., Lambotharan, S. and Chambers, J.A., 2017, December. Support vector machine for network intrusion and cyber-attack detection. In *2017 sensor signal processing for defence conference (SSPD)* (pp. 1-5). IEEE.