JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Machine Learning Based Cyber Security Intrusion Detection in Cloud Computing

Jasmeet Kaur Ghai[1], Susheel Tiwari[2]

[1]Department of Computer Science Madhyanchal Professional University, Bhopal, India.

[2]Department of Computer Science Madhyanchal Professional University, Bhopal, India.

Abstract

Cyberattack detection in cloud computing is a rising concern in the domain of distributed computing. With the increased use of cloud computing in every area of service, threats and attacks in these areas are also growing commensurately. Cyber-attacks in cloud computing have different variants, such as denial of service, probing, and many more flooding behaviours of network traffic. This paper proposes clustering and a classification-based algorithm for the detection of cyberattacks. The proposed algorithm is very efficient in terms of detecting attacks that are normal or abnormal. The machine learning (ML) algorithms that have been used here are support vector machines (SVM) and neural networks (NN). The evaluation metrics used in the comparison of performance are accuracy, recall, f1 score, and area under the Receiver Operating Characteristic Curve. The system obtained 99.4% test accuracy for the proposed algorithm. Though these techniques have the same accuracy, other metrics prove that Random Forest performs comparatively better.

Keywords:- Cloud Computing, intrusion, machine learning, classification

Introduction

Cloud computing security is a major challenge in the current scenario of information technology. The multiple integrations of different paradigms for service security in cloud computing are vulnerable. The emergence of cloud computing has raised interest in it as a versatile technology that can accommodate a wide range of uses[1,2,3]. It became apparent as a breakaway in Internet usage. As a result, cloud computing is currently a hot topic and has proven to be beneficial for small businesses in a world that is developing quickly. It is a blueprint for using the Internet to deliver a range of advantageous services[4]. Additional appealing benefits of cloud computing include lower hardware costs (as users do not require a powerful processor or any other hardware resources), rapid and continuous service upgrades, large storage capacity, worldwide cloud computing for documents, parallel processing, resource sharing, cloud computing excitement, and time savings. Nonetheless, efficiency, security, privacy and trust, control and ownership, availability, fault tolerance and recovery, and connection bandwidth costs are some of the major obstacles facing cloud computing. The main obstacles to cloud computing and its widespread adoption by businesses and organisations are data security and privacy, as cloud computing services are provided over the internet[5,6]. Furthermore, cloud computing's decentralised and open architecture has given rise to a class of computing that is vulnerable to intrusions and cyberattacks. According to NIST, an intrusion is an attempt to circumvent computer and network security measures or jeopardise security policies. Since the network is the foundation of the cloud and vulnerabilities in the network directly impact the security of the cloud, detecting and preventing network intrusions is one of the main security issues facing the cloud. Despite the relative novelty of machine learning applications in the IS field, they have already shown relevance in this matter: for finding vulnerabilities in source [7] and machine code [8,9], attack detection in networks [10], predicting balances on components of large distributed systems [11,12], anomaly detection in real-time data [13], etc. Since a direct search for vulnerabilities is only one of cloud computing analysis' subtasks, it is necessary to consider an application of ML on the full cycle of system investigation, not just for executable files. According to the reported survey of security analysis in cloud computing, machine learning algorithms boost the process of detection and prevention of intrusions. Intrusion detection in cloud computing faces a large number of features related to the data payload. The process of feature optimization and selection of features enables the detection of intrusion in cloud computing. This paper proposes a machine learning-based algorithm for the detection of cyberattacks in cloud computing. The rest of the paper organized as follows: in Section II, related work in the area of intrusion detection; in Section III, proposed methodology; in Section IV, experimental analysis; and finally, conclusion in Section V.

II. Related Work

Recently, machine learning algorithms have improved the detection and prevention of cyber-attacks in cloud computing. A machine learning-based algorithm is employed as a content filter that reduces traffic load and prevents cyber threats. This section describes

the recently proposed methodology for cyber-attack detection in cloud computing. In [2], the authors proposed a high-performance, artificial intelligence-based solution that defends the vehicle network against cyberattacks. The accuracy of the suggested system was quite good, at 97.30%. [3] methodology is meant to act as a manual for upcoming studies in this field. Better accuracy (100%) was attained by the KNN- and DT-algorithms in binary and multiclass classification. In [4], the proposed method assigns a virtual IP address to the host while masking its actual IP address. The proposed method achieves transparency by mapping a virtual IP to the real IP while keeping it intact. The suggested solution is implemented using Mininet and the RYU-Controller. In [5], the strategies and tactics employed to counter DDoS attacks in cloud environments differ from those employed in traditional networks, and sometimes they do not. This research aims to examine the difficulties and countermeasures for denial-of-service (DDoS) attacks in cloud environments and to compare them. In [6], authors design attributes based on the parameters used in the construction of the IDS-based DNN (IDSDNN) or to improve its performance. This method uses a hybrid framework called „IGASAA," which combines machine-learning techniques, namely the Improved Genetic Algorithm (IGA) and Simulated Annealing Algorithm (SAA). In [7], current developments and trends in ML and DL-based NIDS are discussed in relation to the suggested technique, assessment criteria, and dataset choice. We emphasised some research issues and presented the future scope of the research on enhancing ML and DL-based NIDS by using the weaknesses of the given methods. in [8] These data sets also underwent max-min normalisation, and support vector machine (SVM), K-Nearest Neighbour (KNN), and decision tree (DT) algorithms—three of the traditional machine-learning techniques—were used for classification. In order to accurately detect attacks and abnormalities on IoT systems, the performances of multiple machine-learning models have been compared in this study. Test accuracy for Decision Tree, Random Forest, and ANN was 99.9% for the system. [10] improve detection accuracy by combining several learners with the most realistic intrusion detection dataset available, NSL-KDD. For the binary classification and attack classification, the overall accuracy is 85.81% and 84.25%, respectively. [11] presents a brand-new anomaly-detection system that is built around the natural fusion of many deep-learning methodologies. First, we extracted features from network traffic using the damped incremental statistics technique. Next, we trained the auto-encoder using a small set of label data. In [12] limiting DoS attacks in a single MANET, the suggested solution has an acceptable latency, a low communication overhead, and a singularity in the IPv4 address provision. In order to do this, a hypothesis is put forth, assuming that machine-learning techniques can be used for the static analysis of IoT systems. We provide a research plan that reflects the ontology of the study and aims to validate the hypothesis. In [17], alternative techniques, including recurrent neural networks (RNN), neural networks (NN), and extreme gradient boosting (XGBoost), also offer superior performance. This investigation also sheds light on the AI roadmap for threat detection according to attack categories. in [18] Researchers in the fields of computational science, management, science, and engineering may be able to address current challenges and provide solutions to the most pressing real-world issues thanks to these methodological advancements. In [19], the most common cybersecurity threats are mentioned in this paper, along with a description of the machine-learning algorithms that are commonly employed for attack detection and prediction. In order to detect botnet DDoS attacks, this research conducted an experimental analysis of machine-learning techniques. The assessment is carried out on the well-known publicised datasets for botnet DDoS attack detection, UNBS-NB 15, and KDD99. in [21] With a focus on the Internet of Things and machine learning, the study seeks to conduct current, in-depth research on pertinent works that deal with various intelligent techniques and their applicable intrusion detection structures in computer networks. in [22] Along with a detailed comparison of datasets, the current chapter also describes some of the most well-known cybersecurity datasets and machine learning methods. Therefore, it is important to acknowledge the significance of human interaction in crucial security scenarios where domain specialists play a crucial role, while also recognizing the powerful automated capability of ML. [23] presents a scalable architecture for a system that can identify falls, monitor thousands of senior citizens, and alert carers. Additionally, scalability tests were carried out to reveal the prerequisites for large-scale system operations. In [24], the impact of applying readily classified “compact” AI methods and conditionally classified “distributed” AI methods on various cyberthreats has been examined. Furthermore, the study explores the potential applications and challenges of these methods in the field of cybersecurity. In [25], we propose an Anomaly Detection IoT (AD-IoT) system, which is an intelligent anomaly detection based on Random Forest Machine-Learning Algorithm, to handle the IoT-Cybersecurity-Concerns in a smart city. Our results demonstrate that the AD-IoT can efficiently get the lowest false positive rate and maximum classification accuracy of 99.34%. in [26] Numerous machine-learning applications in cybersecurity, including spam classification, malware detection, intrusion detection, and more, have been discussed in this article. in [27] Utilising visualisation representations is a cutting-edge technique for detecting malware and network threats. Greyscale, RGB, or Markov picture representations are available for the network-pcap files. These images are used to train two-dimensional CNNs that can identify network assaults with a high degree of covariance. in [29] deep learning and machine learning methods to accurately extract key features from a real-world BoT-IoT dataset with realistic network traffic. XGBoost and CatBoost classifiers, in particular, had impressive accuracy rates of 98.50% and 98.19%, respectively. in [30] machine learning (ML) in the context of cyber security, which is crucial for the field's further advancement. Two actual case studies detailing industrial applications of machine learning as defined against cyberthreats.

III. Proposed Methodology

This section describes the proposed algorithm for cyberattack detection in cloud computing. The proposed algorithm is an encapsulation of the clustering and classification algorithms. the process of clustering an algorithm group of data on pre-processing network traffic. The classification algorithm labels classes as normal and abnormal. The proposed model is shown in Figure 1. The processing of the algorithm is described here.

The handling of traffic data clustering is very difficult to normal clustering algorithm and density-based clustering algorithm. The density-based clustering enhances with threshold function and applied on traffic data for the process of clustering. The threshold function describes as

$$Th_{pt} = \log_{\lambda} \frac{\alpha}{\alpha} - N(1 - 2^{-\lambda})$$

Here Th is the value of threshold for the selection of new feature points of traffic data. the MCM is main algorithm of traffic data clustering [12]. The MCM algorithm incorporate with density-based clustering and formed cluster for traffic data.

Input: traffic data, Mints, .λ and α

Output: normal or abnormal

1. Define the threshold as derivation

2. T_p=0

3. Process traffic data do

4. Read data point x form data traffic

5. Estimate nearest mini-cluster to x

6. If dist(x, centers) <rmcm then

7. Merge x to MCM

8. Else

9. Map the new data points x to the MCM

11 Update MCM

12. End if

13. Return normal or abnormal.

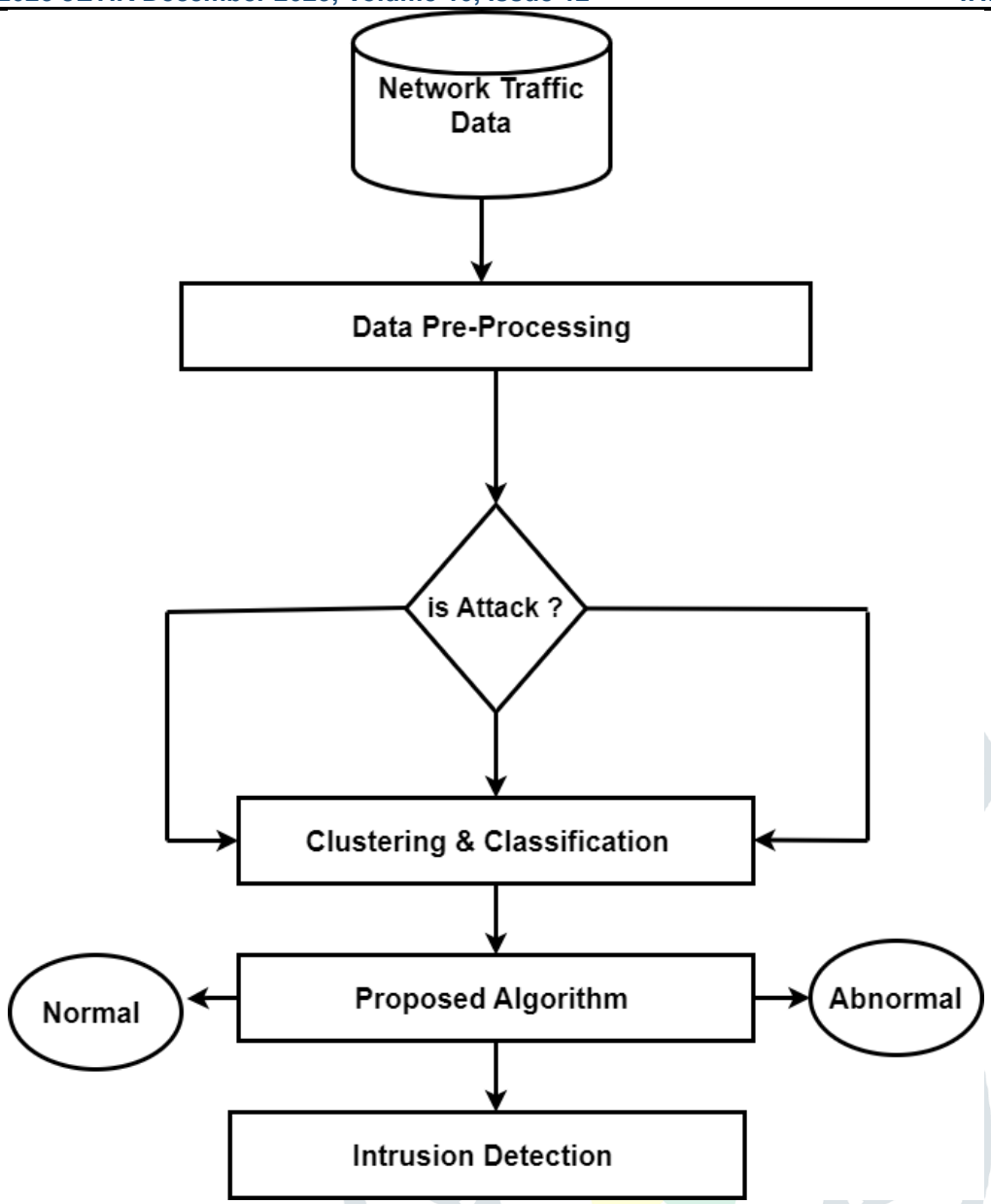

Figure 1 Proposed Model of Intrusion Detection Based On Machine Learning Algorithm

IV. Experimental Analysis

To evaluate the performance of proposed ensemble classifier for the detection of cyber-attacks in cloud computing use MATLAB software tool. The MATLAB tool provides rich library of machine learning algorithm and optimization function for the processing of classification. the system configuration of experimental machine is I7 processor, 16GB RAM and windows operating system. For the validation of algorithm applies two different datasets, named of dataset are CICIDS 2017. The description of dataset mention below. The performance of classification algorithm estimates using confusion matrix of classifiers. The parameters true negative (TN) meaning true prediction of normal behaviour, true positives (TP) implying true prediction of attack behaviour, false positives (FP) showing false prediction of normal behaviour as an assault and false negatives (FN) indicating false prediction of attack as normal [15,16,17,18,19,20]. The performance metrics generated using the confusion matrix which will be used for the evaluation of the proposed IDS are accuracy rate, F-score and detection rate.

$$Accuracy = \frac{TP + FN}{TP + TN + FN + FN} \dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

$$Detection\ rate = \frac{TP}{TP + FN} \dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

$$f - score = \frac{2X\ detection\ rate\ Xprecision}{Detection\ Rate + precision} \dots\dots\dots\dots\dots\dots\dots\dots\dots (3)$$

CICIDS 2017

CICIDS 2017 is a dataset generated by the Canadian Institute for Cyber security that contains the actions of 25 user-based protocols while capturing data [21,22,23]. It is span-ned over eight different CSV files. In those eight CSV files, there are 2,830,743 rows containing 80 features which are labelled as normal and attack. This dataset gives 14 different categories of attacks. The data was captured continuously for 5 days that is from Monday to Friday and categories of attacks contained are distributed denial of service

attack, port scan attack, denial of service attack, web-based attacks, infiltration attacks, and brute force attack. The main characteristics CICIDS 2017 are its huge volume, diversity, public availability, large variety of attacks, reliability etc.

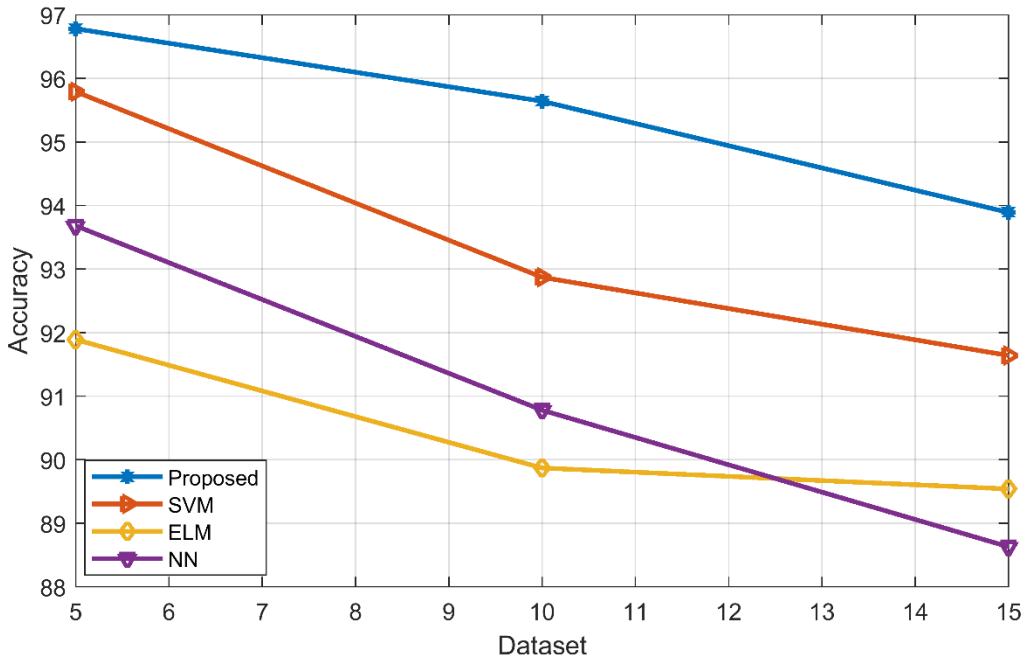

Fig 2: Comparative analysis of accuracy using NN, ELM, SVM, and Proposed techniques with CICIDS-2017 dataset. We observe that the accuracy of that proposed is better than other three techniques NN, ELM, SVM.

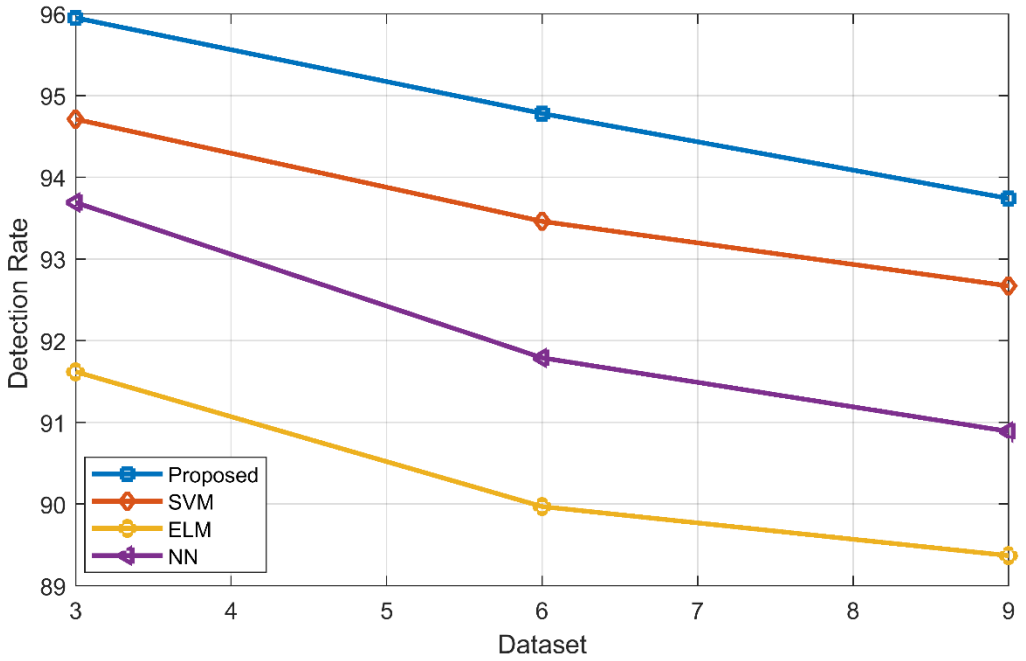

Fig 3: Comparative analysis of Detection rate using NN, ELM, SVM, and Proposed techniques with CICIDS-2017 dataset. We observe that the Detection rate of that proposed is better than other three techniques NN, ELM, SVM.

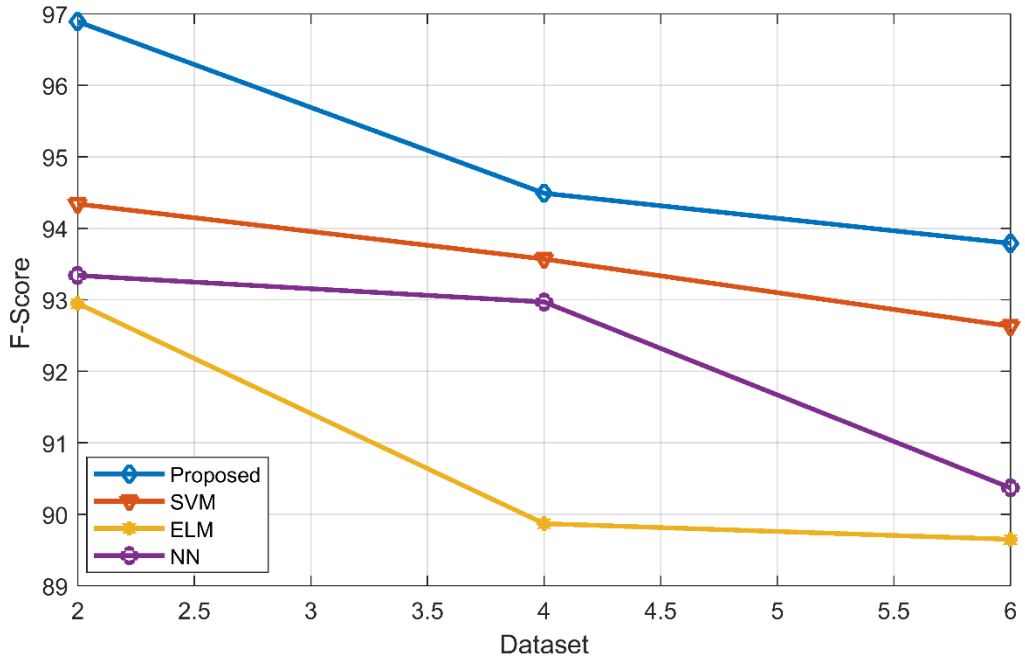

Fig 4: Comparative analysis of F-Score using NN, ELM, SVM, and Proposed techniques with CICIDS-2017 dataset. We observe that the F-Score of that proposed is better than other three techniques NN SVM, ELM.

V. Conclusion & Future Work

This paper proposes novel intrusion detection in cloud computing. The proposed algorithm is a prototype machine-learning algorithm. The prototype algorithm is based on a density-based clustering algorithm and a miner class. The miner class is a classification algorithm. The proposed algorithm effectively detects intrusion in all categories of attacks in cloud networks. During the validation phase, we used a sensitivity analysis to identify discrepancies between the expected output and the target values by examining the metrics TP, TN, and FP. The NN and SVM algorithms produce fewer prediction errors in binary and multiclass classification. In the validation stage, the ELM approaches fared well, with the NN and SVM algorithms significantly outperforming the competition. We verified the validity and effectiveness of our findings by comparing them with those of other recent studies. While the NN and decision accuracy were 99%, both of the proposed classifiers still achieved high accuracy, proving their superiority over other cutting-edge classifier models. Our goal is to integrate our system with an actual traffic system in the future to secure cloud networks.

References

[1]. Uyyala, Prabhakara. "Detection of Cyber Attack in Network Using Machine Learning Techniques." Journal of interdisciplinary cycle research 14, no. 3 (2022): 1903-1913.
[2]. Aldhyani, Theyazn HH, and Hasan Alkahtani. "Attacks to automatous vehicles: A deep learning algorithm for cybersecurity." Sensors 22, no. 1 (2022): 360.
[3]. Alkahtani, Hasan, and Theyazn HH Aldhyani. "Developing cybersecurity systems based on machine learning and deep learning algorithms for protecting food security systems: industrial control systems." Electronics 11, no. 11 (2022): 1717.
[4]. Hyder, Muhammad Faraz, Muhammad Umer Farooq, Usama Ahmed, and Wajahat Raza. "Towards Enhancing the Endpoint Security using Moving Target Defense (Shuffle-based Approach) in Software Defined Networking." Engineering, Technology & Applied Science Research 11, no. 4 (2021): 7483-7488.
[5]. Bakr, Ahmed, A. A. El-Aziz, and Hesham A. Hefny. "A survey on mitigation techniques against ddos attacks on cloud computing architecture." International Journal of Advanced Science and Technology 28, no. 12 (2019): 187-200.
[6]. Chiba, Zouhair, Noreddine Abghour, Khalid Moussaid, and Mohamed Rida. "Intelligent approach to build a Deep Neural Network based IDS for cloud environment using combination of machine learning algorithms." computers & security 86 (2019): 291-317.
[7]. Ahmad, Zeeshan, Adnan Shahid Khan, Cheah Wai Shiang, Johari Abdullah, and Farhan Ahmad. "Network intrusion detection system: A systematic study of machine learning and deep learning approaches." Transactions on Emerging Telecommunications Technologies 32, no. 1 (2021): e4150.
[8]. Kilincer, Ilhan Firat, Fatih Ertam, and Abdulkadir Sengur. "Machine learning methods for cyber security intrusion detection: Datasets and comparative study." Computer Networks 188 (2021): 107840.
[9]. Hasan, Mahmudul, Md Milon Islam, Md Ishrak Islam Zarif, and M. M. A. Hashem. "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches." Internet of Things 7 (2019): 100059.
[10]. Illy, Poulmanogo, Georges Kaddoum, Christian Miranda Moreira, Kuljeet Kaur, and Sahil Garg. "Securing fog-to-things environment using intrusion detection system based on ensemble learning." In 2019 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1-7. IEEE, 2019.
[11]. Zhong, Ying, Wenqi Chen, Zhiliang Wang, Yifan Chen, Kai Wang, Yahui Li, Xia Yin, Xingang Shi, Jiahai Yang, and Keqin Li. "HELAD: A novel network anomaly detection model based on heterogeneous ensemble learning." Computer Networks 169 (2020): 107049.

[12]. Kumar, Anand, Dharmesh Dhabliya, Pankaj Agarwal, Nagender Aneja, Pankaj Dadheech, Sajjad Shaukat Jamal, and Owusu Agyeman Antwi. "Cyber-internet security framework to conquer energy-related attacks on the internet of things with machine learning techniques." Computational intelligence and neuroscience 2022 (2022).

[13]. Kotenko, Igor, Konstantin Izrailov, and Mikhail Buinevich. "Static analysis of information systems for IoT cyber security: a survey of machine learning approaches." Sensors 22, no. 4 (2022): 1335.

[14]. Ghillani, Diptiban. "Deep learning and artificial intelligence framework to improve the cyber security." Authorea Preprints (2022).

[15]. Stergiou, Christos L., Andreas P. Plageras, Konstantinos E. Psannis, and Brij B. Gupta. "Secure machine learning scenario from big data in cloud computing via internet of things network." In Handbook of computer networks and cyber security, pp. 525-554. Springer, Cham, 2020.

[16]. Haq, Amin Ul, Jian Ping Li, Jalaluddin Khan, Muhammad Hammad Memon, Shah Nazir, Sultan Ahmad, Ghufran Ahmad Khan, and Amjad Ali. "Intelligent machine learning approach for effective recognition of diabetes in E-healthcare using clinical data." Sensors 20, no. 9 (2020): 2649.

[17]. Abdullahi, Mujaheed, Yahia Baashar, Hitham Alhussian, Ayed Alwadain, Norshakirah Aziz, Luiz Fernando Capretz, and Said Jadid Abdulkadir. "Detecting cybersecurity attacks in internet of things using artificial intelligence methods: A systematic literature review." Electronics 11, no. 2 (2022): 198.

[18]. Annamalai, Chinnaraji. "Combinatorial and Multinomial Coefficients and its Computing Techniques for Machine Learning and Cybersecurity." The Journal of Engineering and Exact Sciences 8, no. 8 (2022): 14713-01i.

[19]. Bahassi, Hanan, Nahid Edddermoug, Abdeljebar Mansour, and Azmi Mohamed. "Toward an exhaustive review on Machine Learning for Cybersecurity." Procedia Computer Science 203 (2022): 583-587.

[20]. Tuan, Tong Anh, Hoang Viet Long, Le Hoang Son, Raghvendra Kumar, Ishaani Priyadarshini, and Nguyen Thi Kim Son. "Performance evaluation of Botnet DDoS attack detection using machine learning." Evolutionary Intelligence 13, no. 2 (2020): 283-294.

[21]. da Costa, Kelton AP, João P. Papa, Celso O. Lisboa, Roberto Munoz, and Victor Hugo C. de Albuquerque. "Internet of Things: A survey on machine learning-based intrusion detection approaches." Computer Networks 151 (2019): 147-157.

[22]. Kumar, Koushal, and Bhagwati Prasad Pande. "Applications of machine learning techniques in the realm of cybersecurity." Cyber Security and Digital Forensics (2022): 295-315.

[23]. Mrozek, Dariusz, Anna Koczur, and Bożena Małysiak-Mrozek. "Fall detection in older adults with mobile IoT devices and machine learning in the cloud and on the edge." Information Sciences 537 (2020): 132-147.

[24]. Naik, Binny, Ashir Mehta, Hiteshri Yagnik, and Manan Shah. "The impacts of artificial intelligence techniques in augmentation of cybersecurity: a comprehensive review." Complex & Intelligent Systems 8, no. 2 (2022): 1763-1780.

[25]. Alrashdi, Ibrahim, Ali Alqazzaz, Esam Aloufi, Raed Alharthi, Mohamed Zohdy, and Hua Ming. "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, 2019.

[26]. Bharadiya, Jasmin. "Machine Learning in Cybersecurity: Techniques and Challenges." European Journal of Technology 7, no. 2 (2023): 1-14.

[27]. Dhanya, K. A., Sulakshan Vajipayajula, Kartik Srinivasan, Anjali Tibrewal, T. Senthil Kumar, and T. Gireesh Kumar. "Detection of Network Attacks using Machine Learning and Deep Learning Models." Procedia Computer Science 218 (2023): 57-66.

[28]. Karuniawan, Rickho Rizky, Sugeng Santoso, Muhamad Al Fikri, M. Argadilah, and Wisnu Ardi Pamungkas. "Learning Cyber Security and Machine Engineering at the University." Blockchain Frontier Technology 3, no. 1 (2023): 7-12.

[29]. Alkhudaydi, Omar Azib, Moez Krichen, and Ans D. Alghamdi. "A Deep Learning Methodology for Predicting Cybersecurity Attacks on the Internet of Things." Information 14, no. 10 (2023): 550.

[30]. Apruzzese, Giovanni, Pavel Laskov, Edgardo Montes de Oca, Wissam Mallouli, Luis Brdalo Rapa, Athanasios Vasileios Grammatopoulos, and Fabio Di Franco. "The role of machine learning in cybersecurity." Digital Threats: Research and Practice 4, no. 1 (2023): 1-38.