Feature optimization based cyber-attack detection in public cloud network using machine learning swarm intelligence

Ashi Parashar[1], Prof.Mahendra Sahare[2], Prof.Anurag Shrivastava[3]
[1]M. tech Scholar, [2]Associate Professor, [3]HOD, Department of CSE, NIIST Bhopal

***Abstract*—In current decade growth of cloud computing-based services all area of IT-enabled service sector. The growing impact of services interconnect different application and network. the interconnection of network is easy to target of intruders and cyber threats. In order to identify complex and undetected threats, the development of learning intrusion detection systems has been the main focus. Machine learning-based models are frequently used in intrusion detection systems because of their rapid accuracy gains. This study used machine learning to detect attacks on network traffic through multiple classifications. The CICIDS2017 data set, which includes both recent and older attacks, was used to build the model. Tests were run quickly on the CICIDS2017 data set, which has about 2.8 million rows of data, using the high-performance computer. Our machine learning models performed better after we cleaned, normalised, oversampled for an uneven number of labels, and used feature selection techniques to shrink the size of the data set. Using the pre-processed data set, the random forest classifier was found to have the highest accuracy of 99.94% when compared to the decision tree, logistic regression, and other classifiers.**

***Index Terms*—Cloud Computing, Cyber-Attacks, Machine Learning, Feature optimization, SI**

I. INTRODUCTION

Cloud computing refers to a collaborative platform for data storage and computing over the internet. The appealing impacts of cloud computing survive in society and organization all the time. The scalability of services and the reliability of communication face some vulnerable protocols of security in cloud computing [1,2]. The vulnerability of security in cloud computing has been exposed to cyber threats and attacks. Furthermore, the degree of attack detection is critical to the seamless functioning of cloud computing. Stated differently, legitimate pre-processing techniques, neural network architectures, and hyper-parameter optimizers are essential to the functioning of cloud computing [3, 4]. After analysing the available intrusion detection datasets, machine learning algorithms, and optimizers to evaluate the datasets, we investigate the aforementioned issues. This provides a clear picture of the kind of network traffic, which aids in the training model's accurate labelling of the training instances. However, the current research has limited the scope of identifying DDoS attack types, intrusion detection systems, and datasets. We have attempted to address every one of these problems in our paper. For the purpose of early DDoS attack detection in cloud computing [5,6], we suggested a hybrid methodology. This makes it abundantly evident that attackers are using creative techniques to compromise the system, which is why environmental security is crucial. According to a report by the National Institute of Standards and Technology (NIST), IT-related intrusions and attacks cost American businesses as much as $65.5 billion in losses in 2017. Denial-of-service attacks continue to pose a serious threat to companies and institutions that rely on the internet, despite constant, unrelenting efforts by security researchers. These kinds of attacks have the potential to quickly degrade user resources [8]. The importance of security has also been a major source of worry over time. DDoS attacks are one such kind, and because of their impact, they are the most serious. Through the use of artificial intelligence and machine learning, a wide range of solutions for identifying such attacks are feasible. Systems for detecting and preventing intrusions can be developed using artificial intelligence and machine learning. Reported surveys suggest that machine learning and swarm intelligence are great influences in cyber-attack detection in cloud computing. Swarm intelligence-based algorithm employees for feature optimization and feature selection for machine learning algorithms. Swarm intelligence offers several algorithms for detection of intrusion, such as

support vector machines, decision trees, NBs, and artificial neural networks [7,8]. This paper proposes feature optimization-based intrusion detection for cloud computing. The rest of the paper is organised as in Section II related work, in Section III proposed methodology, in Section IV experimental analysis, and finally concludes in Section V.

II. RELATED WORK

The complex feature of intrusion degraded the detection process of intrusion in cloud computing. Feature optimization and feature selection is new approaches in cloud computing for the detection of cyberattacks. Recently, several authors proposed machine learning and a swarm intelligence-based detection approach. Here, we explore recently proposed work in intrusion detection. In order to improve the detection rate during the process of stopping DDoS attacks in clouds, a software-defined networking-based mitigation scheme and a convolutionally enhanced self-organising map are proposed in this research. In order to provide simultaneous security checks and procedural multistage security using authentication techniques, intrusion detection, and encryption, an improvised method for securing the data in cloud computing environments using parallel and multistage security mechanisms (PMSSM) is proposed in this paper. In [3] experimental findings, the suggested hybrid methodology performs better, reaching an 82.5% feature reduction ratio and 98.34% accuracy with ANOVA for XGBoost. It also aids in the early detection of DDoS attacks on Internet of Things devices. in [4] suggest a realistic and lightweight mitigation strategy to safeguard SDN architecture against DDoS-flooding attacks and maintain a safe and effective SDN-based networking environment. in [5] metaheuristic method to optimise the data characteristics and a variety of deep learning techniques to try and identify and categorise such cyber-attacks. The suggested algorithm yielded better results than the others, according to simulations, with 97.8% accuracy for binary classification, 95.6% accuracy for three-class classification, and 94.3% accuracy for multi-class classification. In [6], the presentation and analysis of machine-learning and deep-learning approaches pertain to the identification of cyberattacks in Internet-of-Things systems. In order to provide a comprehensive overview of the advancements in this field, we integrate a comprehensive list of publications published in the literature up to this point. in [7] OSI-model-based attack classification and go in greater depth about the cyberattacks that can target the various communication layers of smart grid networks. primary categories of cyberattacks and primary techniques for identifying them are being examined for this aim. The word „cyber-fault “is then introduced in the analogy to „process fault," a phrase that is well-established in process diagnostics. In order to address this issue, this work suggests a deep learning-based attack detection model for energy systems. This model may be trained using logs and data collected by phasor measurement units (PMUs). The results of the simulation demonstrated that this model outperforms the current approaches, with a detection rate of 93.6% and an accuracy rate of 93.91%. In contrast to previous research, this study analyses the frequency effects of the features in the data set to detect cyberattacks on computer networks. First, the NSL-KDD-Dataset's feature frequencies were ascertained. [11] suggests using convolutional and recurrent neural networks—two technologies that have proven effective when applied to tasks like pattern detection in images—to provide a security solution for the Internet of Things. In [12], for binary classification, we employed the Naïve Bayes algorithm as the first-layer detection method, and for multi-class classification, we used the Light GBM algorithm as the second-layer detection method. in [14] Within the unified framework of control and detection, it is shown that when both observer-based and control-signal-based residual signals are generated and employed for detection, then all kernel attacks can be structurally detected. In [15], analysis is conducted on the sustainability of IoT risk categorization, risk reduction objectives, and implementation issues. Analyses are done on the data dichotomy between sharing and privacy and the openness paradox. As a result, efforts to build security standards and IoT technology are emphasised. In [16], using the pre-processed data set, the random forest classifier was shown to have the greatest accuracy of 99.94% when compared to the decision tree, logistic regression, Naive Bayes, and other classifiers. in [17] cyber-attack detection system with machine-learning models. On an open-source website, data from previous cyberattacks was used to teach machine-learning algorithms to predict the scores of cyberattacks. in [19] Cybersecurity, convolutional neural networks, recurrent neural networks, and deep neural networks are the deep-

learning approaches examined in this paper. With binary class, the CNN model has the maximum accuracy of 98.64% and the highest precision of 98%. At 97.75%, the RNN model has the second-highest accuracy. The CNN model offers the best accuracy with multiclass class, at 98.42. In [20], the study looks at cybersecurity problems and weaknesses in 5G and 4G technologies. The study used both primary and secondary data to arrive at its conclusions. Cyberattacks can be quickly and effectively detected and mitigated by security solutions. In order to stop detrimental losses, our suggested research seeks to quickly and accurately detect network intrusions. With a high-performance accuracy of 99.9%, the thorough research results showed that the random forest approach performed better than the state-of-the-art approach. in [22]. Based on a multi-class classification technique, this study trained two intelligent network models, Dense Net and Inception Time, to identify cyberattacks. The best outcome when using the Inception Time Method on the Edge-IIoT dataset was an accuracy of 94.94%. in [23] The recent study highlights how crucial a drone network with strong security is to thwart attacks and interception. The temporal efficacy statistical measures precision (97.68%), accuracy (98.58%), recall (98.59%), F-measure (99.01%), reliability (94.69%), and stability are used to assess the model's performance. In [24], an intrusion detection system with notable accuracy in detecting DDoS attacks is implemented using an ensemble-unsupervised machine-learning approach. The objective of this study is to reduce the false positive rate and raise the accuracy of DDoS attack detection. in [25] Utilising the availability of vast amounts of network data and the development of increasingly potent computing hardware, academics have proposed machine-learning-based frameworks to secure IVNs in recent years. [26] presents a novel GAN- and LSTM-based approach for detecting such attacks: Using a collection of network traffic data from different IoT devices, we apply our model to classify incoming traffic as either benign or malicious. in [27] To help decide when to employ which of these techniques, this paper compares and examines the key differences between ML and DL-Technician. For managing larger volumes of data, DL approaches are still considered to be better. Furthermore, attacks have dynamic defences against IDS of their own. in [28] Convolutional long-short-term memory is the foundation of the approach (ConvLSTM). This model consists of a convolutional LSTM layer, two convolutional layers, and two pooling layers. Furthermore, it attains a 99% accuracy rate, while the literature-presented works reach up to 95% accuracy.

III. PROPOSED METHODOLOGY

The KNN classifier is simple algorithm of machine learning, it also knows as lazy classifier. The classification accuracy of KNN classifier varies in range of 70-80%. The major utility of KNN classifier in case of pattern recognition. The KNN classification algorithm applied on the case of continuous nature of attribute. The processing of KNN algorithm describe here

1. Estimate K training attribute which belong to unknow attribute
2. Chose the common occurring classification of K

For the estimation of similarity in class of K instance applied different distance equation. The very famous distance equation is Euclidean distance equation.

Input: A data set according to sample selection
Output: a mixed transform table data
class: E= {}, the set of the equivalence classes
QIC= {}, set of equivalence classes with
similar QI sets
CIP {}, set of attributes with similar class
DIP=number of different class values in the remaining dataset
Begin
While CIP >= attribute
Cluster T to m tables according QI
For i=1 to m
Bucketize attributes according SA values
While |DIPi|>=ℓ
Create_equivalence_ classes ()
E=E U Create_ equivalence_ classes ()
return E
Incorporate the remaining attributes to E
End
Generate equivalence class with prototype is
Input: CIP
Output: E
Begin
Randomly selection of a attributes tm
from the smallest group
E={tm}
For p=1 until attribute-1
Select a attributes tp that minimizes the gcp
E=E U tp
Remove tp from T
Remove tm from T

Return E
End
Process of cluster generation in prototype classification
Input: data set used defined
Output: QIC= {}, set of tables with attributes with similar QI sets
Begin
Insert T to the decision tree classification
QIC= {QIC1, QIC2, QICm}
return QIC
End

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 mentions 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.

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

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

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

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.

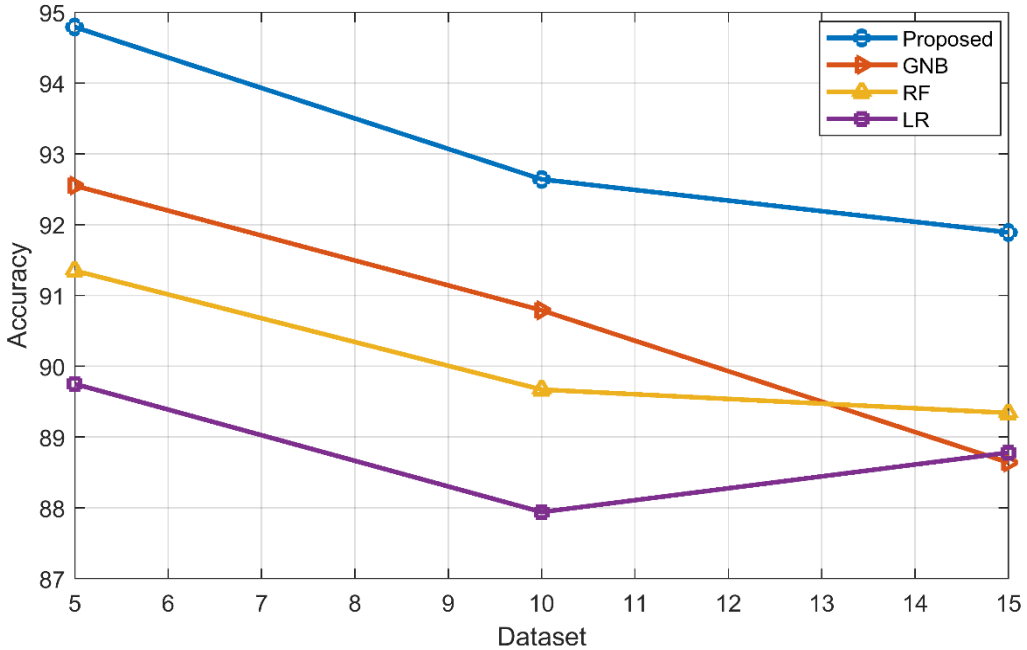

Fig 2: Comparative analysis of Accuracy using LR, RF, GNB, and Proposed techniques with CICIDS-2017 dataset. We observe that the Accuracy of that proposed is better than other three techniques LR, RF, GNB.

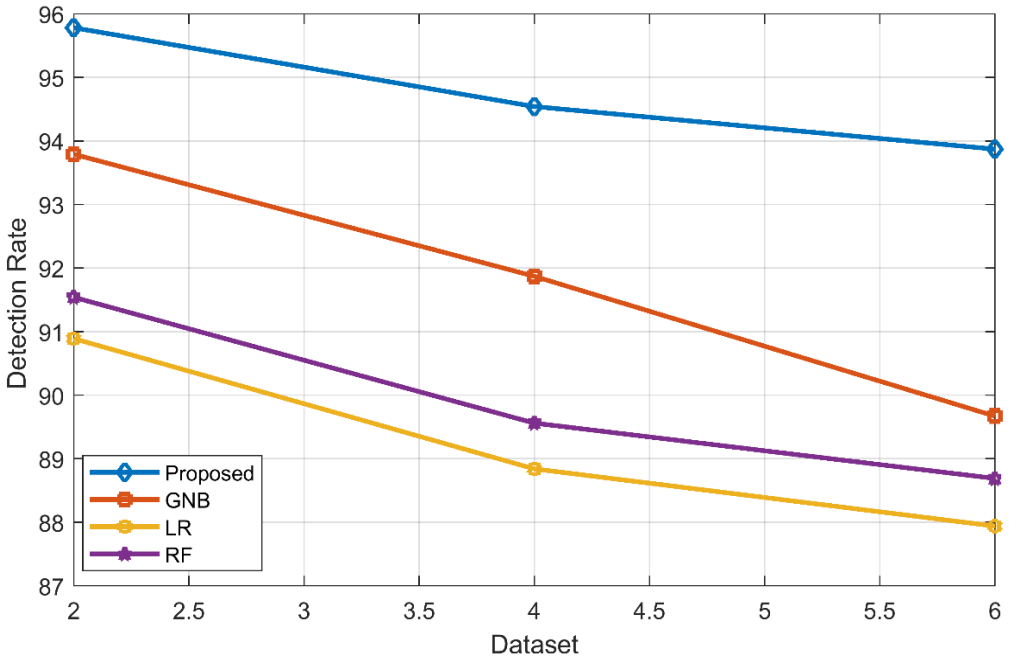

Fig 3: Comparative analysis of Detection rate using LR, RF, GNB, and Proposed techniques with CICIDS-2017 dataset. We observe that the Detection rate of that proposed is better than other three techniques LR, RF, GNB.

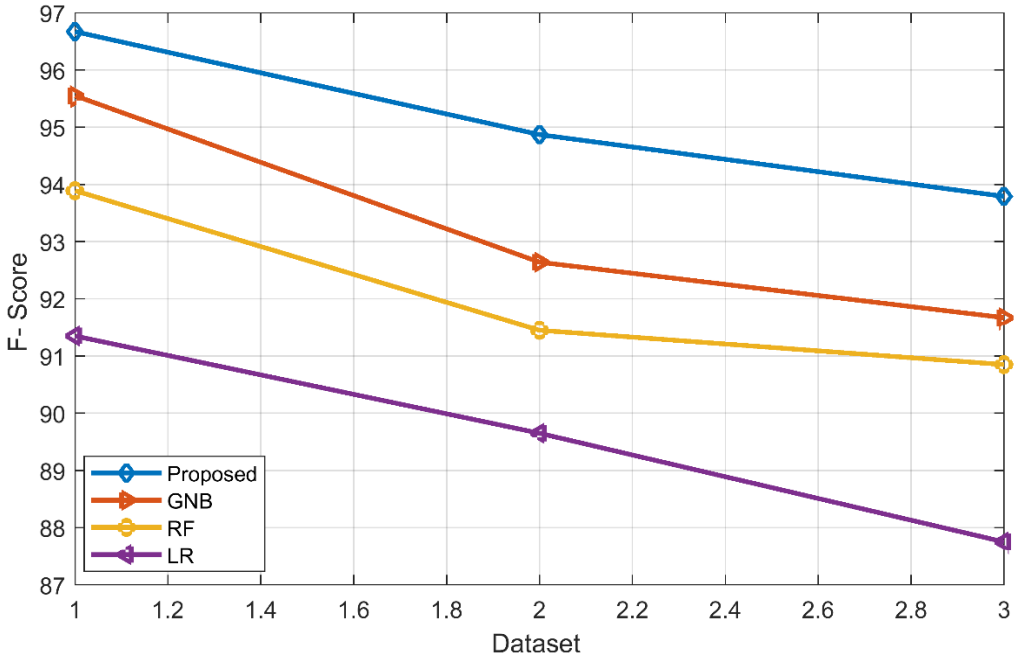

Fig 4: Comparative analysis of F-Score using LR, RF, GNB, and Proposed techniques with CICIDS-2017 dataset. We observe that the F-Score of that proposed is better than other three techniques LR, RF, GNB.

V. CONCLUSION & FUTURE WORK

This has become a serious issue recently, so it has been investigated how the machine learning model contributes to the effectiveness of IDS used for attack detection. The goal of this project is to use machine learning techniques to create a model on an unbalanced and pre-processed data set. Step-by-step research was done to determine how each operation affected the model's data set. Using the normalisation algorithms selected in the feature selection phase, the data set was normalised, and the run times and MAE values of four classifiers were

noted. Although the standard normalisation procedure was preferred for our model, an alternative normalisation technique based on the values in the table might be the best option if the model only contained one classifier. The classifications we performed on the finished data set yielded an accuracy of 99.94% with RF, 99.92% with GNB, 90.70% with logistic regression, and 87.31% with GNB. The pre-processing phases of the data were found to have a significant impact on the model's performance when it came to detecting multiple attacks on unbalanced data sets.

REFERENCES

[1] Harikrishna, Pillutla, and A. Amuthan. "SDN-based DDoS attack mitigation scheme using convolution recursively enhanced self-organizing maps." Sādhanā 45 (2020): 1-12.

[2] Goyal, Ranjan, R. Manoov, Prabu Sevugan, and P. Swarnalatha. "Securing the data in cloud environment using parallel and multistage security mechanism." In Soft Computing for Problem Solving: SocProS 2018, Volume 2, pp. 941-949. Springer Singapore, 2020.

[3] Gaur, Vimal, and Rajneesh Kumar. "Analysis of machine learning classifiers for early detection of DDoS attacks on IoT devices." Arabian Journal for Science and Engineering 47, no. 2 (2022): 1353-1374.

[4] MAHRACH, Safaa, and Abdelkrim HAQIQ. "DDoS flooding attack mitigation in software defined networks." International Journal of Advanced Computer Science and Applications 11, no. 1 (2020).

[5] Diaba, Sayawu Yakubu, Miadreza Shafie-Khah, and Mohammed Elmusrati. "Cyber Security in Power Systems Using Meta-Heuristic and Deep Learning Algorithms." IEEE Access 11 (2023): 18660-18672.

[6] Inayat, Usman, Muhammad Fahad Zia, Sajid Mahmood, Haris M. Khalid, and Mohamed Benbouzid. "Learning-based methods for cyber-attacks detection in IoT systems: A survey on methods, analysis, and future prospects." Electronics 11, no. 9 (2022): 1502.

[7] Khoei, Tala Talaei, Hadjar Ould Slimane, and Naima Kaabouch. "A Comprehensive Survey on the Cyber-Security of Smart Grids: Cyber-Attacks, Detection, Countermeasure Techniques, and Future Directions." arXiv preprint arXiv:2207.07738 (2022).

[8] Syfert, Michał, Andrzej Ordys, Jan Maciej Kościelny, Paweł Wnuk, Jakub Możaryn, and Krzysztof Kukiełka. "Integrated approach to diagnostics of failures and cyber-attacks in industrial control systems." Energies 15, no. 17 (2022): 6212.

[9] Almalaq, Abdulaziz, Saleh Albadran, and Mohamed A. Mohamed. "Deep machine learning model-based cyber-attacks detection in smart power systems." Mathematics 10, no. 15 (2022): 2574.

[10] Özalp, Ahmet Nusret, and Zafer Albayrak. "Detecting Cyber Attacks with High-Frequency Features using Machine Learning Algorithms." Acta Polytechnica Hungarica 19, no. 7 (2022): 213-233.

[11] Parra, Jhon Alexánder, Sergio Armando Gutiérrez, and John Willian Branch. "A Method Based on Deep Learning for the Detection and Characterization of Cybersecurity Incidents in Internet of Things Devices." arXiv preprint arXiv:2203.00608 (2022).

[12] Ismail, Shereen, Diana Dawoud, and Hassan Reza. "A lightweight multilayer machine learning detection system for cyber-attacks in WSN." In 2022 IEEE 12th annual computing and communication workshop and conference (CCWC), pp. 0481-0486. IEEE, 2022.

[13] Lekidis, Alexios. "Cyber-security measures for protecting EPES systems in the 5G area." In Proceedings of the 17th International Conference on Availability, Reliability and Security, pp. 1-10. 2022.

[14] Ding, Steven X., Linlin Li, Dong Zhao, Chris Louen, and Tianyu Liu. "Application of the unified control and detection framework to detecting stealthy integrity cyber-attacks on feedback control systems." Automatica 142 (2022): 110352.

[15] Salam, Abdul, and Abdul Salam. "Internet of things for sustainability: perspectives in privacy, cybersecurity, and future trends." Internet of things for sustainable community development: wireless communications, sensing, and systems (2020): 299-327.

[16] Güven, Ebu Yusuf, Sueda Gülgün, Ceyda Manav, Behice Bakır, and Zeynep Gürkaş Aydın. "Multiple Classification of Cyber

Attacks Using Machine Learning." Electrica 22, no. 2 (2022): 313-320.

[17] Akhtar, Muhammad Shoaib, and Tao Feng. "Comparison of classification model for the detection of cyber-attack using ensemble learning models." EAI Endorsed Transactions on Scalable Information Systems 9, no. 5 (2022).

[18] Ghelani, Diptiben, Tan Kian Hua, and Surendra Kumar Reddy Koduru. "Cyber Security Threats, Vulnerabilities, and Security Solutions Models in Banking." Authorea Preprints (2022).

[19] Barik, Kousik, Sanjay Misra, Karabi Konar, Luis Fernandez-Sanz, and Murat Koyuncu. "Cybersecurity deep: approaches, attacks dataset, and comparative study." Applied Artificial Intelligence 36, no. 1 (2022): 2055399.

[20] Yousef Alshunaifi, Sulaiman, Shailendra Mishra, and Mohammed Alshehri. "Cyber-Attack Detection and Mitigation Using SVM for 5G Network." Intelligent Automation & Soft Computing 31, no. 1 (2022).

[21] Raza, Ali, Kashif Munir, Mubarak S. Almutairi, and Rukhshanda Sehar. "Novel Class Probability Features for Optimizing Network Attack Detection with Machine Learning." IEEE Access (2023).

[22] Tareq, Imad, Bassant M. Elbagoury, Salsabil El-Regaily, and El-Sayed M. El-Horbaty. "Analysis of ton-iot, unw-nb15, and edge-iiot datasets using dl in cybersecurity for iot." Applied Sciences 12, no. 19 (2022): 9572.

[23] Aldaej, Abdulaziz, Tariq Ahamed Ahanger, Mohammed Atiquzzaman, Imdad Ullah, and Muhammad Yousufudin. "Smart cybersecurity framework for IoT-empowered drones: machine learning perspective." Sensors 22, no. 7 (2022): 2630.

[24] Das, Saikat, Deepak Venugopal, and Sajjan Shiva. "A holistic approach for detecting ddos attacks by using ensemble unsupervised machine learning." In Advances in Information and Communication: Proceedings of the 2020 Future of Information and Communication Conference (FICC), Volume 2, pp. 721-738. Springer International Publishing, 2020.

[25] Refat, Rafi Ud Daula, Abdulrahman Abu Elkhail, and Hafiz Malik. "Machine Learning for Automotive Cybersecurity: Challenges, Opportunities and Future Directions." AI-enabled Technologies for Autonomous and Connected Vehicles (2022): 547-567.

[26] Kaushik, Priyanka. "Unleashing the power of multi-agent deep learning: Cyber-attack detection in IoT." International Journal for Global Academic & Scientific Research 2, no. 2 (2023): 23-45.

[27] Al-Shareeda, Mahmood A., Selvakumar Manickam, and Murtaja Ali. "DDoS attacks detection using machine learning and deep learning techniques: Analysis and comparison." Bulletin of Electrical Engineering and Informatics 12, no. 2 (2023): 930-939.

[28] Sedik, Ahmed, Osama S. Faragallah, Hala S. El-sayed, Ghada M. El-Banby, Fathi E. Abd El-Samie, Ashraf AM Khalaf, and Walid El-Shafai. "An efficient cybersecurity framework for facial video forensics detection based on multimodal deep learning." Neural Computing and Applications (2022): 1-18.