

Innovative Computing Review (ICR)

Volume 2 Issue 2, Fall 2022


ISSN(P): 2791-0024 ISSN(E): 2791-0032

Homepage: <https://journals.umt.edu.pk/index.php/ICR>



Article QR



- Title:** **Machine Learning for Intrusion Detection in Cyber Security: Applications, Challenges, and Recommendations**
- Author (s):** Aqib Ali¹, Samreen Naeem¹, Sania Anam², Muhammad Munawar Ahmed³
- Affiliation (s):** ¹College of Automation, Southeast University, Nanjing, China.
²Govt Associate College for Women Ahmadpur East, Bahawalpur, Pakistan.
³Islamia University Bahawalpur, Bahawalpur, Pakistan.
- DOI:** <https://doi.org/10.32350.icr.22.03>
- History:** Received: October 10, 2022, Revised: November 11, 2022, Accepted: December 2, 2022
- Citation:** A. Ali, S. Naeem, S. Anam, and M. M. Ahmed, "Machine learning for intrusion detection in cyber security: Applications, challenges, and recommendations," *UMT Artif. Intell. Rev.*, vol. 2, no. 2, pp. 41-64, 2022, doi: <https://doi.org/10.32350.icr.22.03>
- Copyright:** © The Authors
- Licensing:**  This article is open access and is distributed under the terms of [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)
- Conflict of Interest:** Author(s) declared no conflict of interest



A publication of
School of Systems and Technology
University of Management and Technology, Lahore, Pakistan

Machine Learning for Intrusion Detection in Cyber Security: Applications, Challenges, and Recommendations

Aqib Ali^{*}, Samreen Naeem¹, Sania Anam², and Muhammad Munawar Ahmed³

¹College of Automation, Southeast University, Nanjing, China.

²Department of Computer Science, Govt. Associate College for Women Ahmadpur East, Bahawalpur, Pakistan.

³Department of Information Technology, Islamia University Bahawalpur, Pakistan.

Abstract-Modern life revolves around networks and cybersecurity has emerged as a critical study field. The health of the software and hardware running on a network is monitored by an Intrusion Detection System (IDS) which is a fundamental cybersecurity approach. After decades of research, the existing IDSs have developed the capability to confront hurdles in order to improve detection accuracy, reduce false alarm rates, and detect unexpected attacks. Many academics have concentrated on designing such IDSs that employ machine learning approaches to overcome the aforementioned difficulties. Machine learning approaches are capable to discover important distinctions that exist between normal and aberrant data with great accuracy. Moreover, these approaches are also very generalizable which allows them to detect unknown attacks. The survey conducted in the current study offers ataxonomy of IDS based on machine learning that uses data objects as the

critical dimension to classify and summarize the IDS literature. This form of classification structure is appropriate for cyber security researchers.

Index Terms-classification, feature optimization, Intrusion Detection System, machine Learning Classification

I. Introduction

The Internet has become a vital aspect of modern lives as the digital world has grown considerably [1]. With the emergence of smart cities, self-driving cars, health monitoring via wearables, and mobile banking, among many other things, internet addiction is on the rise. While these technologies assist individuals and societies at a large scale, they also pose several concerns [2]. For instance, hackers could take advantage of weaknesses, resulting in theft and sabotage that harm people worldwide. Cyberattacks

*Corresponding Author: aqibcsit@gmail.com

may be costly to organizations regarding both cash losses and reputational damage. As a result, network security has become a significant concern. Organizations' use of traditional measures, such as firewalls, encryption, and antivirus software packages play a considerable part to safeguard their network infrastructure. These approaches; however, only provide the first line of protection and cannot fully defend networks and systems against malware and advancing attacks. Consequently, some intruders are nevertheless, able to get access which may result in a breach [3].

Intrusion Detection refers to the security of computer systems against illegal usages, such as hackers and any form of misuse from lawful access, such as insider threats (ID). A breach in the computer system may result in data loss, restricted access to internet resources, the loss of sensitive data, and the exploitation of private resources [4]. Denning (1987) was the first to construct the Intrusion Detection System (IDS). Therefore, it has become a hot study area as a vital tool for computer network security since then. Given the

current level of cybercrime, there is little doubt that the IDS plays a critical role. The classification of IDS taxonomy is shown in Fig. 1. The IDS could be regarded as a hardware or software system that monitors, detects, and warns the computer or network of attacks or intrusions [5]. This warning report assists the administrator or user to locate and resolve the system or network vulnerability. An attempt to access the data, change it, or render the system unworkable after an intrusion might be purposeful or a criminal act. The area of cybersecurity aids to prevent and detect the illicit computer activity. Data in both hardware and software is safeguarded against destruction and disruption [6]. Computer security prevents the intruders to use the computers for their personal benefits. Firewalls, suites, antiviruses, and other cybersecurity tools are the various examples to protect the system. Data availability at the right moment, asset authentication, document confidentiality, integrity, and all specified data are the four primary categories that may be classed under this specific domain, notably in cybersecurity [7].

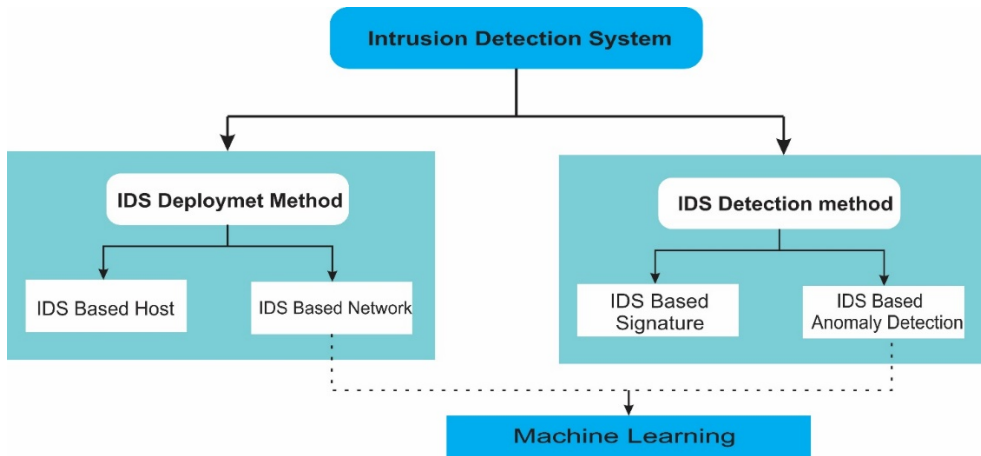


Fig. 1. Classification of IDS taxonomy

According to the World Internet Statistics report, the Internet's growth rate from 2000 to 2019 was 1.114 percent, with more than two quintillion bytes of data created per day [8]. This demonstrates that the data accumulation from diverse sources was relatively rapid, while the development of hacking tools and procedures also increased rapidly. To secure data from intrusion, information security and data analysis are necessary measures. The typical detection system cannot detect intruders due to the enormous volume and high data velocity. Significant data approaches are employed to efficiently handle the intrusion. The 7v defines big data as Volume: data size, Speed: data generation pace, Variety: diverse sorts of data, Value: the data's worth, Truthfulness: the data's

dependability, Variability: the data's meaning is changing over time, and Visualization: the data's simple access or reading [9].

Because of the exponential rate of data expansion, traditional data management systems are incredibly complicated and are time and resource-intensive. The accumulation of vast data is inherently complex, necessitating solid technology and knowledgeable algorithms to handle it. To identify the attacks, IDS is crucial. An IDS monitors the network traffic to detect unusual behaviors and known threats. The administrators could then be informed on the discovery of such conducts to avoid any trouble. ML algorithms may be used to efficiently handle and categorize the attacks [10]. Intrusion detection is

divided into two kinds based on how it works and they are as follows:

A. Active IDS

Active IDS are similar to passive IDS in that they prevent attacks by blocking suspicious traffic.

B. Passive IDS

These IDS merely monitor and analyze traffic by notifying the administrator of attacks and vulnerabilities [11].

II. Applications of Intrusion Detection Systems

Intrusion Detection Systems are vital to prevent cyber-attacks. All transactions and data processing occurs through the Internet, which is very susceptible to fraudulent activities. It is essential that the Information security must be emphasized. Fig. 2 summarizes the IDS based applications.

A. IDS for Internet of Things

The Internet of Things (IoT) is a network of things or devices that can detect, collect, and transmit data without human or computer interaction. Low-power IoT devices use lightweight protocols. The Reference [12] discussed smart grid IoT devices. Attackers may manipulate the sensor data. Physical, side channel, environmental, cryptanalysis, black

hole, and Sybil attacks are common types of IoT attacks. The Reference [13] proposed supervised light intrusion detection. SVM was created to identify attacks (DDoS target).

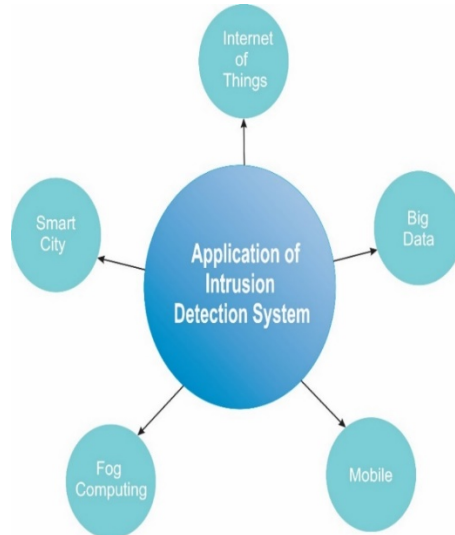


Fig. 2. Application of IDS

B. IDS of Smart City

The Reference [14] described the intelligent city intrusion detection. The author utilized an intelligent water distribution system dataset. Smart city DDoS attacks must be detected. The approach suggested in the current paper consists of two parts, that is, RBM and classifier. This RBM model helps unsupervised high-level learning. Classification is used to differentiate DDOS attacks. The FFNN, AFNN, RF, and SVM classifiers were employed. RBM

model processes the K-Means method and contains up to 5 layers that give five subversions of each clustering algorithm with a distinct k value. Four classifiers as they are used for every five cluster-generated datasets and 20 tests are run.

C. IDS for Big Data

Big data refers to as being heterogeneous, organized, unstructured, and semi-structured. Traditional intrusion management cannot handle excessive data therefore, ML is needed for Big Data IDS. The Reference [15] utilized Apache Spark Big Data to identify intrusion detection. The preprocessed model uses Mlib spark unit variance. NumTopFeatures is used to pick features using Chisqselector and SVM. SVM soft margin reduces misclassification. The slack variable swaps margin and classification error. Their results reveal faster and more efficient big data intrusion detection.

D. IDS for Fog Computing

Fog computing is a type of novel processing paradigm that moves analytics to the edge to boost performance. Fog computing has a cloud, fog, and user levels. The fog service layer has a globally dispersed fog node comprising a

router, gateway, and edge server. Fog nodes allow heterogeneous processing, making them vulnerable to attacks like DDoS, Remote-to-Local (R2L), User-to-Root (U2R), and PROBE. The Reference [16] added to the DDoS attack process in fog computing and studied the fog node and hypergraph-based DDoS. Load factor helps to determine the fog node status. The fog node's threshold charge level determines its condition. This approach is used to assess a DDoS attack's association with cloud nodes.

E. IDS for Mobile

People in the modern era increasingly use mobile phones to communicate and store their sensitive information. Mobile vulnerabilities include apps, devices, networks, online sources, and content vulnerabilities. Therefore, IDS is needed to handle these vulnerabilities and threats. The Reference [17] presented a 5G-oriented cyber protection architecture to recognize 5G mobile network's cyber threats. The incursions were defined by dividing anomaly detection into two levels, that is, ASD and NAD. The NAD uses a supervised variant of LSTM (Recurrent Short-Term Memory Networks), while the ASD module uses a supervised or semi-

supervised two-level form of DBN and SAE.

III. The Role of Machine Learning in IDS

One of the machine learning activities is classification which is also a paradigm of supervised learning. It is employed in intrusion detection systems that are binary-based or multiclass. The data is always labeled in supervised learning, with each record in a data set being assigned to a specific class. All network traffic is categorized into normal or abnormal classification techniques by an IDS based on a classification model. The enormous volume of data impedes to create the model. Data preparation is required by classification techniques, which may handle various challenges in model construction, especially with high-dimensional data. The confusion matrix and accuracy performance assessment criteria determines the optimal ranking

method [18]. The two rounds of training and testing are involved to categorize the data in the dataset. A target classifier is learned during the training and learning phase. The generated model predicts class labels for provided data during the second phase, that is, the test phase. It is critical to determine that how much time each classifier takes during the training and testing phases. Data preparation helps the classification model to minimize time and complexity by eliminating unnecessary data and improving the performance of classification methods before applying classifiers. For the IDS dataset categorization, the cross-validation procedure is divided evenly into two groups, that is, one group would be utilized for testing and the remainder would be used for training. Only a few algorithms may reliably discriminate between unusual attacks and typical attacks [19] as shown in Fig. 3

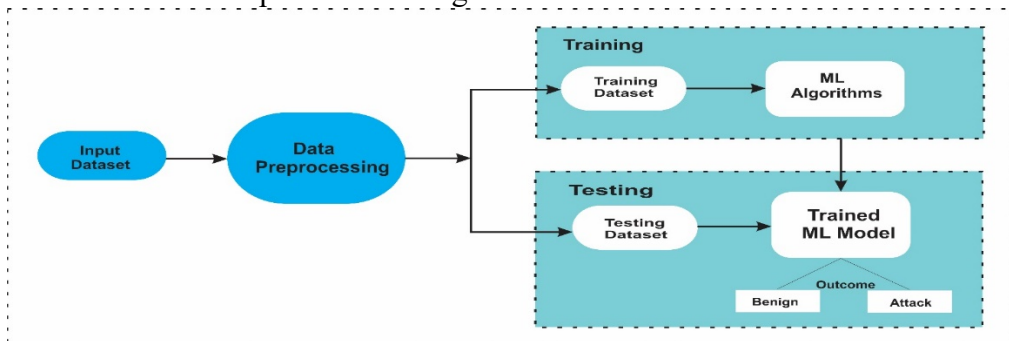


Fig. 3. Generalized machine learning based IDS methodology

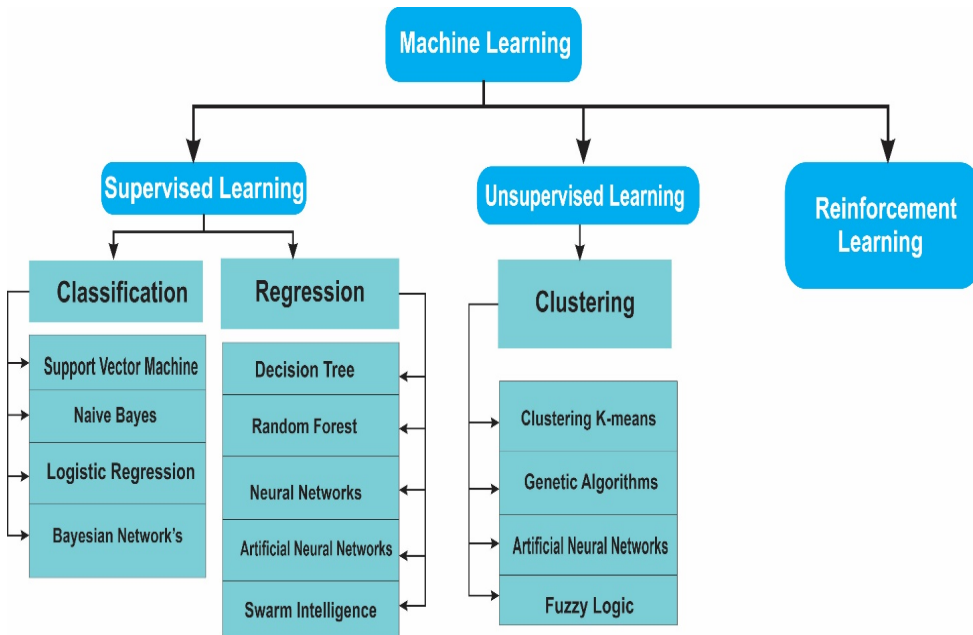


Fig. 4. Machine Learning Classifiers Categorization

The most commonly used machine learning classifiers in IDSs discussed below in Fig. 4.

A. Neural Networks

The brain and other biological nervous systems handle information similarly as neural networks do. Artificial neural networks [20] have the ability to recognize and categorize the network activity depending on various characteristics.

B. Bayesian Network's

Bayesian Networks is a probabilistic learning model that depicts an acyclic graph with conditional and unconditional

relationships. The Reference [21] proposed an automated intrusion detection system.

C. Genetic Algorithms

A genetic algorithm based on a natural selection process is used to address both bound and unbound problems. To identify intrusions, genetic algorithms may be employed effectively [22].

D. Artificial Neural Networks (ANN)

ANNs are based on biological neural networks. They execute tasks using examples from noisy and partial data. ANN was designed to address brain-like issues. Data-

intensive applications employ such systems. This section discusses ANN kinds, contributions, and intrusion detection performance [23].

E. Support Vector Machine (SVM)

This supervised model classifies, regresses, and detects the outliers. Hyperplane-based data linearization. SVM maps the data into feature space and separates it into classes using a hyperplane with the most significant class margin and transforms it into multiclass ensemble. SVM excels the nonlinear data. Using SVM, researchers have detected intrusions. The Reference [24]–[25] used SVM to detect network breaches. Authors say that the high-quality training data improves detection efficiency and they also presented an SVM-based IDS. To improve SVM detection, they log-transformed marginal density ratio (LMDRT). The results indicated excellent DR and decent efficiency.

F. Naïve Bayes (NB)

Naïve Bayes is a Theorem-based categorization method. This classifier believes that each characteristic's class probability is independent of others [26]. This method is used to calculate the instance probabilities of each class

and choose the highest likelihood. NB is also used to detect intrusions.

G. Regression Logistic (LR)

The LR estimates Zero or One from independent values. The fit data predicts the logistic function's event. The Reference [27] presented the network anomaly which defined that the detection method is based on Internet traffic's nonlinear invariant features. The findings demonstrated that this approach separates a wide variety of volumetric DoS attacks with great accuracy and precision.

H. Decision Tree (DT)

A chart or tree model is used to make decisions and examine the potential ramifications of those actions, including the outcomes of random occurrences. A decision tree has symbolic labels, while a regression tree has continuous values. This method attempts to sort the sample through a tree of options, with each decision affecting the next. These decisions are tree-structured. CART creates decision trees, whereas DT is used to detect intrusions. The Reference [28] suggested a misuse and anomaly-based hybrid intrusion detection approach. The experiment used NSL-KDD. The proposed strategy improved DR, FPR, and complexity. The recommended

method's time-saving techniques were not great. However, the future study would prove helpful to improve C4.5's decision tree algorithm.

I. Random Forest (RF)

RF builds a decision tree, as the name indicates. It is made by integrating many decision trees and averaging their forecasts. Single indication is typically less accurate. A forest looks healthier with more trees. The Reference [29] suggested an RF intrusion detector model. RF surpassed other conventional classifiers in ranking successful attacks.

J. Clustering K-means

Clustering with K-means is an unsupervised ML algorithm. Unsupervised algorithms don't label the data. Data search groups drive this algorithm. Groups items are based on similarities and contrasts. K-means is used to pattern-match time series data. K-Means is incapable to handle non-spherical findings. Using K-mean, researchers have detected intrusions [30].

K. Fuzzy Logic (FL)

Fuzzy Logic is utilized to examine the safety of a place and to begin scientific research. For quantitative and security reasons, fuzzy logic was employed for

intrusion detection. Fuzzy logic allows an item to belong to many classes at once, it also proves useful when class differences are unclear. Fuzzy theory may identify intruders when normal and abnormal classifications aren't correctly defined [31].

L. Swarm Intelligence (SI)

It solves complicated issues through agent-environment interactions. SI requires self-organization and work division. Self-organization is a system's capacity to restore its agents without outside aid. Parallel task execution allows him to solve complicated challenges. The ACO and PSO are swarm-inspired algorithms. ACO replicates ant behavior and solves discrete optimization issues, whereas PSO solves nonlinear optimization problems [32]–[33].

These approaches allow algorithms to move beyond merely static program instructions, producing data-driven predictions or judgments by constructing a model from sample inputs. It may be utilized in various computing tasks when explicit methods cannot be designed or programmed, such as network infiltration and security breach.

IV. Literature Review

Several studies have been conducted to enhance IDS to detect and prevent cyberattacks in the previous decade. This section examines the data preparation, feature selection, number of features picked, classification methods, and assessment algorithms used in intrusion detection classification.

The Reference [34] developed a machine learning-based wireless network IDS. Before training, the preparation phase converts the dataset values to integers, scales huge data ranges, and normalizes them into smaller fields. The current study employed multiple ML classifiers and focused to improve classification algorithms' feature optimization efficiency, which improves accuracy and detection time. The numbers 32, 10, 7, and 5 were chosen as valuable functions for the training model. The random forest classifier with 32 specified attributes showed the best performance in the experiments. The classification methods represented 99.64% accuracy, 0.995 precision, and 0.966 recalled—the suggested system used AWID wireless data. The comparison of the proposed approach with various categorization methods helps to validate the results.

The Reference [35] observed the performance of 4 ML classifiers. Apache Spark tools were used to categorize the network traffic intrusion detection. The model uses 42 characteristics from the UNSW-NB15 public network intrusion dataset. Among various classifiers, a random forest classifier has the most remarkable accuracy of 97.49%, specificity of 97.75% and sensitivity of 93.53%.

The Reference [36] suggested a Hybrid Filter-based Selection Algorithm (HFSA). HFSA optimized a subset of the most relevant and highest-ranking classifier functions. This model uses real-time Jpcap packets. Nave Bayes classifies regular attacks as harmful ones. Preprocessing involves two stages. Firstly, the difficulty to transform input into a quantitative value. Secondly, during data normalization, each record's attributes are scaled from (0,1) to (0,1). Naive Bayes feature selection and Naive Bayes regression methods were used to detect six standard classes. HFSA improves the categorization system. The model's total accuracy was 92%, with 95% accuracy and 90% recall.

The Reference [37] introduced an IDS based on SVM and Nave Bayes algorithms. The function selection correlation subset type

selected 24 of 42 NSL-KDD functions. The data preparation methods convert the characteristics to binary numbers and normalizes the data. SVM showed 93.95% overall promising accuracy.

The Reference [38] proposed a supervised approach to detect malicious network traffic. The current study employed ANN and SVM algorithms to classify the data. Both filter and wrapper feature selections were used, that is, Chi-Square and correlation. The 25,191-record NSL-KDD training dataset. The wrapper technique is based on 17 of 41 essential features. A chi-squared filter selects 35 more interesting and relevant training model attributes. The wrapper strategy, which picks 17 features, has the maximum ANN accuracy of 94.02%.

The Reference [39] introduced a new feature of categorization and selection technique using ART and Random Forest. HAIDS is the system (Hybrid Anomaly-Based Intrusion Detection System). The hybrid technique showed very promising accuracy of 87.74%.

The Reference [40] introduced IDS based hybrid system which combined KNN, ELM, and HELM. The suggested system's KDD Cup 99 results revealed 84.29%

accuracy. The approach showed a 77.18% attack detection rate.

The Reference [41] developed an embedded approach for SVM-based intrusion detection that uses Naive Bayes. The embedding model was used in numerous datasets to identify different sorts of attacks, including NSL-KDD and Kyoto 2006+. Based on the embedded system against a single SVM algorithm, the suggested technique found that the combination of Naive Bays with SVM improves detection accuracy. NSL-KDD represented the highest accuracy of 99.36%.

The Reference [42] revealed that IDS uses a hybrid classification algorithm with profile augmentation. Hybrid classification methods use Nave Bayes and SVM. It also preprocesses the data. Normalizing data, scaling attributes to 0,1, and picking the suitable real-time dataset characteristics improve model accuracy. This hybrid technique achieved an overall accuracy of 93.10%.

The Reference [43] proposed hybrid IDS categorization in 2020. Hybrid of Decision Tree J48 was performed with SVM. The SVM overcomes high-dimensionality. Particle Swarm Optimization (PSO) was utilized to extract features, selecting nine out of 42 that were meaningful. Training and testing

was carried out using KDD99. The data collection was proportioned. The results revealed that 70% testing and 30% training proved best for accuracy and false alarm rate. The hybrid model achieved 99.1% of accuracy.

The Reference [44] updated the electricity smart grid to identify regular harmful attacks. A Hybrid Decision Trees (HDTs) approach was devised to identify the attacks. The proposed hybrid method's presentation was also compared with SVM. The trials demonstrated that the proposed strategy (HDT) was more efficient with a measuring accuracy of 97.2193% using NSLKDD.

The Reference [45] suggested a DDoS detection approach to increase network security in 2020. The classification was carried out using K-Nearest Neighbor and Nave Bayes, while feature extraction employed correlation. The proposed model was compared against NSL-KDD and KDD Cup 99 learning models. The eight-character KNN technique surpassed Naive Bayes. Performance was calculated to be 98.51 percent and accuracy 98.9%.

The Reference [46] explained the usage of feature reduction in the classification model. Intelligent IDS were presented employing various ML classifiers. The current

study compared the results of classifiers using all 41 features vs. 11, 12, 13, and 15 feature sets. The reduction of characteristics enhanced precision in the experiment. Random Forest Classification Algorithm performed better with the DoS class at 99.63% accuracy.

The Reference [47] developed an intrusion detection system employing a random forest classifier with PCA scaling. Decision trees, naive Bayes, and SVM were compared to the suggested technique. The proposed approach obtained the maximum accuracy of 96.78 percent, an error rate of 0.21 percent, and built the 3.42 model which proved to be the fastest.

The Reference [48] provided a technique for anomalous IDS based on ML classifier. The CSE-CIC-IDS2018 dataset model showed 80 features. This ensemble feature optimization approach used Chi-square to calculate high feature rank correlation. The hybrid technique picked 23 of 80 features. The suggested model outperformed the three overall classifiers' accuracy of 98.8%.

V. Comparative Analysis of Various ML Algorithms Used for IDS

The survey of intrusion detection using ML algorithm was

provided and addressed in the current study. Various IDS apps were thrown out, as well as a performance evaluation. The survey's results are summarized in Table I.

Table I
Summary of Literature Review

Ref	Dataset	Feature Optimization Approach	Classifier	Accuracy
[34]	AWID	ZeroR	Random Forest	99.64%
[35]	UNSW-NB15	ZeroR	Random Forest	97.49%
[36]	KDD Cup 99	HFSA	Naïve Bayes	92%
[37]	NSL– KDD	CFS Subset Eval	SVM	93.95%
[38]	NSL– KDD	Correlation Chi-Square	ANN	94.02%
[39]	UNSW-NB15	Random Forest	Regression Tree	87.74%
[40]	NSL– KDD	Software SDN	KNN	84.29%
[41]	NSL– KDD	Naïve Bayes	Hybrid SVM	99.36%
[42]	Real World Log	Naïve Bayes	Naïve Bayes	95.3%
[43]	KDD'99	PSO	J48	99.1%
[44]	NSLKDD	CART tree	Decision Tree	97.21%
[45]	KDD Cup 99	Correlation	KNN	98.9%
[46]	NSL KDD	feature reduction PCA - RFE	Random Forest	99.63%
[47]	KDD	PCA	Random Forest	96.78%
[48]	SE-CIC-IDS2018	Chi-square Correlation	Decision Tree	98.8%

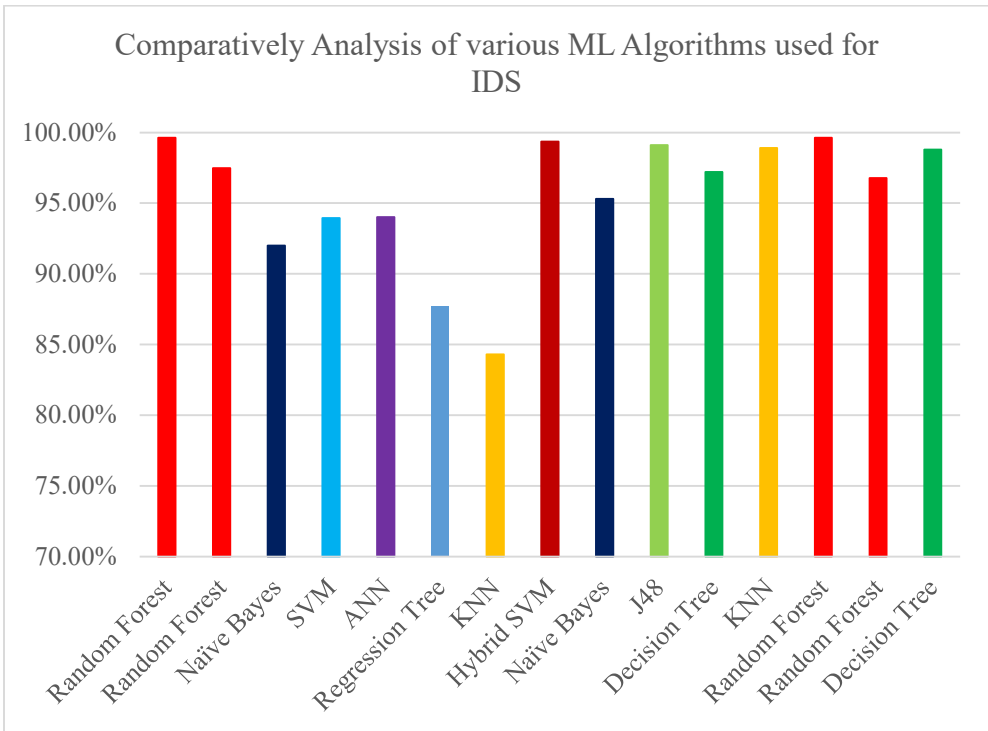


Fig. 5. Graphically comparatively analysis of various ml algorithms used for IDS

Most researchers compared the suggested that Random Forest and Decision Tree models according on the literature review. The highest accuracy was calculated to be 99.64% which was obtained by using Random Forest as shown in Fig. 5.

VI. Research Challenges

This section discusses IDS research challenges as shown in Fig. 6.



Fig. 6. IDS based research challenges

A. No Systematic Dataset

The current study emphasized the lack of an up-to-date dataset reflecting recent network threats. Most of the proposed approaches couldn't detect zero-day attacks because their models lacked adequate kinds and patterns of the attack. Earlier and newer attacks must be evaluated and confirmed for an effective IDS model. By incorporating the maximum number of attacks in a dataset, ML/DL may learn more patterns and guard against maximal incursions. Dataset creation is expensive and requires expertise. One of the IDS's research problems is building an up-to-date dataset with fine examples of practically all attack types. The dataset should be updated periodically and made accessible to benefit researchers [49].

B. Lower Detection Accuracy Owing to an Imbalanced Dataset

According to the current study, most of the proposed IDS approaches have poorer detection accuracy for particular attack types than the model overall. Unbalanced data causes this difficulty. Low-frequency attacks have insufficient detection accuracy than frequent strikes. To combat this problem two solutions have been proposed. Firstly, create a balanced, up-to-date dataset. Secondly, increase the

number of minority attack occurrences to balance the dataset. Recently, researchers applied SMOTE, Random Over Sampler, and ADASYN Algorithm to reduce the dataset imbalance ratio and improve performance [50].

C. Real-World Performance

The Real-world performance is another IDS research problem. Most suggested approaches are lab-tested using public datasets. None of the offered methods is field-tested. It's unclear how they would function in real-world situations. Still old datasets for testing are being used. The proposed procedure must be equally effective as in lab testing. The suggested solution should be evaluated in real-time to ensure its usefulness for current networks [51].

D. Complex Models Take Resources

Most IDS strategies provided by the researcher need a lot of processing time and computational resources (almost 80 percent DL-based methods or ML-based methods). This may add processing costs and degrade the IDS performance. A multi-core GPU may speed up the calculation and minimize time, however it is expensive. Similarly, an efficient feature selection method is needed

to choose the most significant features for speedy processing. Researchers are exploring different optimization techniques for feature selection, however, there is still a room for improvement. More study is needed to develop an efficient approach [52].

E. Lightweight IoT Security

An IDS can secure the IoT network and sensor nodes. In IoT, sensor nodes collect and exchange critical data online. Sensor nodes have limited CPU, storage, and battery life. IDS may be installed where internet traffic enters the IoT network or is dispersed over sensor nodes. In the first case, the NIDS must identify malicious attacks efficiently and face the same obstacles. Secondly, resource-limited sensor nodes need a lightweight IDS paradigm. Designing a lightweight IDS model, efficient in processing power, training time, and intrusion detection rate is a problem [53].

VII. Conclusion

The effectiveness of various machine learning strategies is required since it plays an important role to enhance the IDS performance. Classification algorithms play a crucial part to help the IDSs differentiate between multiple forms of attacks. The

current article aimed to evaluate the performance of various/differently ranking algorithms by using a variety of criteria and compared their results. A variety of metrics were used to evaluate the performance of the classifiers, out of which the random forest method produced satisfactory results. It proved to be one of the excellent and accurate methods to identify the various kinds of attacks. To obtain good performance from the model, most researchers chose to construct IDSs by utilizing the hybrid classification method, rather than using individual classification. In big data sets, the success of size reduction in lowering the complexity leads to selecting outstanding features which, in turn, leads to improved classification performance in terms of accuracy and speed.

Conflict of Interest

The authors declare that they have no conflict of interest regarding the publication of this paper.

Acknowledgment

The authors would like to thank the referees for their careful reading and for their comments, which significantly improved the paper. Additionally, thanks to Dr. Salman Qadri, (Associate Professor,

Chairman Department of Computer Science, MNS University of Agriculture, Multan, Pakistan) and Dr. Farrukh Jamal, (Assistant Professor, Department of Statistics, The Islamia University of Bahawalpur, Pakistan) for his motivational support.

References

- [1] I. Levin and M. Dan, "Culture and society in the digital age," *Information*, vol. 12, no. 2, Art. no. 68, Feb. 2021, doi: <https://doi.org/10.3390/info12020068>
- [2] N. A. Usmani, T. Ahmed, and M. Faisal, "An IoT-based Framework toward a Feasible Safe and Smart City Using Drone Surveillance," in *Smart Cities*, K. Kumar, G. Saini, D. Manh Nguyen, N. Kumar, and R. Shah, Eds., CRC Press, 2022, pp. 97–112.
- [3] K. F. Steinmetz, A. Pimentel, and W. R. Goe, "Performing social engineering: A qualitative study of information security deceptions," *Comput. Hum. Behav.*, vol. 124, Art. no. 106930, 2021, doi: <https://doi.org/10.1016/j.chb.2021.106930>
- [4] Z. Ahmad, A. K. Shahid, C. S. Wai, J. Abdullah, and F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 1, Art. no. 4150, 2021, doi: <https://doi.org/10.1002/ett.4150>
- [5] M. Sarhan, S. Layeghy, and M. Portmann, "Towards a standard feature set for network intrusion detection system datasets," *Mobile Netw. Appl.*, vol. 1, pp. 1-14, 2021, doi: <https://doi.org/10.1007/s11036-021-01843-0>
- [6] A. Thakkar and R. Lohiya, "A survey on intrusion detection system: feature selection, model, performance measures, application perspective, challenges, and future research directions," *Artif. Intell. Rev.*, vol. 55, pp. 453–563, 2021, doi: <https://doi.org/10.1007/s10462-021-10037-9>
- [7] R. Leszczyna, "Review of cybersecurity assessment methods: Applicability perspective," *Comput. Secur.*, vol. 108, Art. no. 102376, 2021, doi: <https://doi.org/10.1016/j.cose.2021.102376>
- [8] H. Wu, N. Ba, S. Ren, et al., "The impact of internet development on the health of

- Chinese residents: Transmission mechanisms and empirical tests,” *Socio-Econom. Plann. Sci.*, vol. 81, Art. no. 101178, 2021, doi: <https://doi.org/10.1016/j.seps.2021.101178>
- [9] H. Wu, Y. Hao, S. Ren, X. Yang, and G. Xie, “Does internet development improve green total factor energy efficiency? Evidence from China,” *Energy Policy*, vol. 153, Art. no. 112247, 2021, doi: <https://doi.org/10.1016/j.enpol.2021.112247>
- [10] A. Churcher, R. Ullah, J. Ahmad, et al., “An experimental analysis of attack classification using machine learning in IoT networks,” *Sensors*, vol. 21, no. 2, Art. no. 446, 2021, doi: <https://doi.org/10.3390/s21020446>
- [11] J. Perháč, V. Novitzká, W. Steingartner, and Z. Bilanová, “Formal model of IDS based on BDI logic,” *Math.*, vol. 9, no. 18, Art. no. 2290, 2021, doi: <https://doi.org/10.3390/math9182290>
- [12] N. Abosata, S. A. Rubaye, G. Inalhan, and C. Emmanouilidis, “Internet of things for system integrity: a comprehensive survey on security, attacks and countermeasures for industrial applications,” *Sensors*, vol. 21, no. 11, Art. no. 3654, 2021, doi: <https://doi.org/10.3390/s21113654>
- [13] A. Khraisat and A. Alazab, “A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges,” *Cybersecur.*, vol. 4, no. 1, pp. 1–27, 2021, doi: <https://doi.org/10.1186/s42400-021-00077-7>
- [14] D. Chen, P. Wawrzynski, and Z. Lv, “Cyber security in smart cities: A review of deep learning-based applications and case studies,” *Sustain. Cities Soci.*, vol. 66, Art. no. 102655, 2021, doi: <https://doi.org/10.1016/j.scs.2020.102655>
- [15] M. Mahdavisarif, S. Jamali, and R. Fotohi, “Big data-aware intrusion detection system in communication networks: a deep learning approach,” *J. Grid Comput.*, vol. 19, no. 4, pp. 1–28, 2021, doi: <https://doi.org/10.1007/s10723-021-09581-z>

- [16] P. Kumar, G. P. Gupta, and R. Tripathi, "Design of anomaly-based intrusion detection system using fog computing for IoT network," *Aut. Control Comput. Sci.*, vol. 55, no. 2, pp. 137–147, 2021, doi: <https://doi.org/10.3103/S0146411621020085>
- [17] V. Ponnusamy, M. Humayun, N. Jhanjhi, A. Yichiet, and M. F. Almufareh, "Intrusion detection systems in internet of things and mobile ad-hoc networks," *Comput. Syst. Sci. Eng.*, vol. 40, no. 3, pp. 1199–1215, 2022, doi: <https://doi.org/10.32604/csse.2022.018518>
- [18] Y. Jiang and Y. Atif, "A selective ensemble model for cognitive cybersecurity analysis," *J. Netw. Comput. Appl.*, vol. 193, Art. no. 103210, 2021, doi: <https://doi.org/10.1016/j.jnca.2021.103210>
- [19] I. Castiglioni, L. Rundo, M. Codari, et al., "AI applications to medical images: From machine learning to deep learning," *Physica Med.*, vol. 83, pp. 9–24, 2021, doi: <https://doi.org/10.1016/j.ejmp.2021.02.006>
- [20] A. O. Drewek, M. Pietrołaj, and J. Rumiński, "A survey of neural networks usage for intrusion detection systems," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 1, pp. 497–514, 2021, doi: <https://doi.org/10.1007/s12652-020-02014-x>
- [21] P. G. George and V. R. Renjith, "Evolution of safety and security risk assessment methodologies towards the use of bayesian networks in process industries," *Process Saf. Environ. Prot.*, vol. 149, pp. 758–775, 2021, doi: <https://doi.org/10.1016/j.psep.2021.03.031>
- [22] A. J. Obaid, K. A. Alghurabi, S. A. Albermany, and S. Sharma, "Improving extreme learning machine accuracy utilizing genetic algorithm for intrusion detection purposes," in *Research in Intelligent and Computing in Engineering*, R. N. Kumar, N. H. Quang, V. Kumar Solanki, M. Cardona, P. K. Pattnaik, Eds., Singapore: Springer, 2021, pp. 171–177, doi: https://doi.org/10.1007/978-981-15-7527-3_17
- [23] M. Choraś and M. Pawlicki, "Intrusion detection approach based on optimised artificial

- neural network,” *Neurocomput.*, vol. 452, pp. 705–715, 2021, doi: <https://doi.org/10.1016/j.neucom.2020.07.138>
- [24] M. Ajdani and H. Ghaffary, “Design network intrusion detection system using support vector machine,” *Int. J. Commun. Syst.*, vol. 34, no. 3, Art. no. 4689, 2021, doi: <https://doi.org/10.1002/dac.4689>
- [25] M. Mohammadi, T. A. Rashid, S. H. T. Karim, et al, “A comprehensive survey and taxonomy of the SVM-based intrusion detection systems,” *J. Netw. Comput. Appl.*, vol. 178, Art. no. 102983, 2021, doi: <https://doi.org/10.1016/j.jnca.2021.102983>
- [26] M. Zubair, A. Ali, S. Naeem, F. Jamal and C. Chesneau, “Emotion recognition from facial expression using machine vision approach,” *J. Appl. Emerg. Sci.*, vol. 10, no. 1, pp. 12–21, 2020.
- [27] X. Duan, S. Ying, W. Yuan, H. Cheng, and X. Yin, “QLLog: A log anomaly detection method based on Q-learning algorithm,” *Info. Process. Manag.*, vol. 58, no. 3, Art. no. 102540, 2021, doi: <https://doi.org/10.1016/j.ipm.2021.102540>
- [28] R. Kajal, D. Syamala, and G. Ajay, “Decision tree-based Algorithm for Intrusion Detection,” *Int. J. Adv. Netw. Appl.*, vol. 7, no. 4, pp. 2828–2834, 2021.
- [29] N. Kaur, M. Bansal, and S. S. Sran, “Scrutinizing attacks and evaluating performance appraisal parameters via feature selection in intrusion detection system,” *Res. Squ.*, vol. 10, pp. 1–14, 2021, doi: : <https://doi.org/10.21203/rs.3.rs-748765/v1>
- [30] Q. V. Dang, “Studying the fuzzy clustering algorithm for intrusion detection on the attacks to the domain name system,” in *2021 5th World Conf. Smart Trends Syst. Secur. Sustainab. (WorldS4)*, London, United Kingdom, 29–30 July, 2021, IEEE, pp. 271–274, doi: <https://doi.org/10.1109/WorldS451998.2021.9514038>
- [31] M. Almseidin, J. Al-Sawwa, and M. Alkasassbeh, “Anomaly-based Intrusion Detection System Using Fuzzy Logic,” in *2021 Int. Conf. Inform. Technol.*, IEEE, Amman, Jordan, July 14–15, 2021, pp. 290-295, doi:

- <https://doi.org/10.1109/ICIT52682.2021.9491742>
- [32] A. Alsaleh and W. Binsaeedan, "The influence of salp swarm algorithm-based feature selection on network anomaly intrusion detection," *IEEE Access*, vol. 9, pp. 112466-112477, Aug. 2021, doi: <https://doi.org/10.1109/ACCESS.2021.3102095>
- [33] J. E. Fontecha, P. Agarwal, M. N. Torres, S. Mukherjee, L. J. Walteros, and J. P. Rodríguez, "A two-stage data-driven spatiotemporal analysis to predict failure risk of urban sewer systems leveraging machine learning algorithms," *Risk Anal.*, vol. 41, no. 12, pp. 122-151, Dec. 2021, doi: <https://doi.org/10.1111/risa.13742>
- [34] R. Abdulhammed, M. Faezipour, A. Abuzneid, and A. Alessa, "Effective features selection and machine learning classifiers for improved wireless intrusion detection," in *2018 Int. Symp. Netw. Comput. Commun.*, Rome, Italy, June 19–21, 2018, pp. 1–6, doi: <https://doi.org/10.1109/ISNCC.2018.8530969>
- [35] A. Ali and S. Naeem, "The controller parameter optimization for nonlinear systems using particle swarm optimization and genetic algorithm," *J. Appl. Emerg. Sci.*, vol. 12, no. 1, 2022.
- [36] K. S. Bhosale, M. Nenova, and G. Iliev, "Data mining based advanced algorithm for intrusion detections in communication networks," in *2018 Int. Conf. Comput. Tech. Electron. Mechanic. Syst.*, Belgaum, India, Dec. 21–22, 2018, pp. 297–300, doi: <https://doi.org/10.1109/CTEM.S.2018.8769173>
- [37] K. K. Gulla, P. Viswanath, S. B. Veluru, and R. R. Kumar, "Machine learning based intrusion detection techniques," in *Handbook of computer Networks and Cyber Security*, B. Gupta, G. Perez, D. Agrawal, D. Gupta. Eds., Springer, 2020. pp. 873–888.
- [38] K. A. Taher, B. M. Y. Jisan, and M. M. Rahman, "Network intrusion detection using supervised machine learning technique with feature selection," in *2019 Int. Conf. Robot. Elect. Signal Process. Tech.*, 10–12 Jan. 2019, pp. 643–64, doi:

- <https://doi.org/10.1109/ICRES.T.2019.8644161>
- [39] S. Naeem and A. Ali, “Bees algorithm based solution of non-convex dynamic power dispatch issues in thermal units,” *J. Appl. Emerg. Sci.*, vol. 12, no. 1, 2022.
- [40] M. Latah and I. Toker, “An efficient flow-based multi-level hybrid intrusion detection system for software-defined networks,” *CCF Trans. Netw.*, vol. 3, no. 3, pp. 261–271, 2020, doi: <https://doi.org/10.1007/s42045-020-00040-z>
- [41] J. Gu and S. Lu, “An effective intrusion detection approach using SVM with naïve Bayes feature embedding,” *Comput. Secur.*, vol. 103, Art. no. 102158, 2021, doi: <https://doi.org/10.1016/j.cose.2020.102158>
- [42] P. Pokharel, R. Pokhrel, and S. Sigdel, “Intrusion detection system based on hybrid classifier and user profile enhancement techniques,” in *2020 Int. Work. Big Data Inform. Secur.*, pp. 137–144, 2020.
- [43] A. Kumari and A. K. Mehta, “A hybrid intrusion detection system based on decision tree and support vector machine,” in *2020 IEEE 5th Int. Conf. Comput. Commun. Autom.*, Greater Noida, India, Oct. 30–31, 2020, pp. 396–400.
- [44] S. M. Taghavinejad, M. Taghavinejad, L. Shahmiri, M. Zavvar, and M. H. Zavvar, “Intrusion detection in iot-based smart grid using hybrid decision tree,” in *2020 6th Int. Conf. Web Res.*, Tehran, Iran, Apr. 22–23, 2020, pp. 152–156, <https://doi.org/10.1109/ICWR49608.2020.9122320>
- [45] A. V. Kachavimath, S. V. Nazare, and S. S. Akki, “Distributed denial of service attack detection using naïve bayes and k-nearest neighbor for network forensics,” in *2020 2nd Int. Conf. Innov. Mecha. Indust. Appl.*, Bangalore, India, Mar. 5–7, 2020, pp. 711–717, doi: <https://doi.org/10.1109/ICIMI448430.2020.9074929>
- [46] G. Sah and S. Banerjee, “Feature reduction and classifications techniques for intrusion detection system,” in *2020 Int. Conf. Commun. Sig. Process.*, Chennai, India, July 28–30, 2020, pp. 1543–1547, doi:

- <https://doi.org/10.1109/ICCSP.48568.2020.9182216>
- [47] S. Waskle, L. Parashar, and U. Singh, “intrusion detection system using PCA with random forest approach,” in *2020 Int. Conf. Electron. Sustain. Commun. Syst.*, Coimbatore, India, July 2–4, 2020 pp. 803–808, doi: <https://doi.org/10.1109/ICESC.48915.2020.9155656>
- [48] Q. R. S. Fitni and K. Ramli, “Implementation of ensemble learning and feature selection for performance improvements in anomaly-based intrusion detection systems,” in *2020 IEEE Int. Conf. Indust. 4.0, Artif. Intell. Commun. Technol.*, Bali, Indonesia, July 7–8, 2020, pp. 118–124, doi: <https://doi.org/10.1109/IAICT.50021.2020.9172014>
- [49] M. Ghurab, G. Gaphari, F. Alshami, R. Alshamy, and S. Othman, “A detailed analysis of benchmark datasets for network intrusion detection system,” *Asian J. Res. Comput. Sci.*, vol. 7, no. 4, pp. 14–33, 2021.
- [50] M. Ragab and M. F. S. Sabir, “Outlier detection with optimal hybrid deep learning enabled intrusion detection system for ubiquitous and smart environment,” *Sustain. Energy Technol. Assess.*, vol. 52, Art. no. 102311, Aug. 2022, doi: <https://doi.org/10.1016/j.seta.2022.102311>
- [51] Z. Wang, Y. Liu, D. He, and S. Chan, “Intrusion detection methods based on integrated deep learning model,” *Comput. Secur.*, vol. 103, Art. no. 102177, Apr. 2021, doi: <https://doi.org/10.1016/j.cose.2021.102177>
- [52] N. Jose and J. Govindarajan, “DOMAIN-Based intelligent network intrusion detection system,” in *Invent. Comput. Info. Technol.*, S. Smys, V. E. Balas, R. Palanisamy, Eds., Singapore, Springer, pp. 449–462, 2022, doi: https://doi.org/10.1007/978-981-16-6723-7_34
- [53] S. Roy, J. Li, B. J. Choi, and Y. Bai, “A lightweight supervised intrusion detection mechanism for IoT networks,” *Future Gener. Comput. Syst.*, vol. 127, pp. 276–285, Feb. 2022, doi: <https://doi.org/10.1016/j.future.2021.09.027>