International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(20s) 2026 pp. 94-102
https://doi.org/10.70917/ijcisim-2026-5154
Article

Exploring Optimization and Machine Learning for Effective Intrusion Detection Systems

Dr. Bharati Vishwas Yelikar[1], Dr. Jawed Sharfuzama Khan[2], Dr. Anuradha Shantanu Kanade[3], Mrs. Rupali Amol Sonar[4], Dr. Deepa Nitin Jamnik[5], Dr. Rajesh Kumar Kashyap[6]

[1] Associate Professor, Department of Computer Application, Bharati Vidyapeeth Institute of Management and Entrepreneurship Development, affiliated to Bharati Vidyapeeth (Deemed to be University), Pune, Maharashtra – 411038, India.
Email: bharati.yelikar@bharatividyapeeth.edu ORCID: 0000-0001-5875-7944
[2] Associate Professor and Head, Department of MCA, MCES Allana Institute of Computer Application and Information Technology, Dr. P. A. Inamdar University, Pune, Maharashtra, India. Email: jawedkhan2@gmail.com
ORCID: 0009-0009-8934-6318
[3] Associate Professor, Department of Computer Science and Applications, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India. Email: anuradha.kanade@mitwpu.edu.in ORCID: 0000-0003-4426-1053
[4] Assistant Professor, Department of MCA/BCA, Savitribai Phule Pune University, Pune, Maharashtra, India.
Email: rsr3346@gmail.com ORCID: 0009-0004-7150-7252
[5] Assistant Professor and Head, Department of IMBA, Dean Events & HOD IMBA, Savitribai Phule Pune University, Pune, Maharashtra, India. Email: deepa.jamnik@indiraicem.ac.in ORCID: 0009-0002-1289-4614
[6] Professor, Department of MBA, G. H. Raisoni College of Engineering and Management, Savitribai Phule Pune University, Wagholi, Pune, Maharashtra, India. Email: rajdlw@gmail.com ORCID: 0009-0003-3840-6970

Abstract: Intrusion Detection Systems (IDS) are essential elements of contemporary cyber security frameworks, intended to identify unauthorised access and nefarious activity within networks and computer systems. This survey examines the classification of IDS technologies, methodologies, and approaches, emphasising the benefits and obstacles related to their use. Numerous optimisation methods in Intrusion Detection Systems (IDS) are examined, emphasising feature selection and machine learning (ML) algorithms that improve detection precision and efficacy. A comprehensive analysis of recent research investigates progress in IDS, highlighting the utilisation of Genetic Algorithms (GA), Particle Swarm Optimisation (PSO), and ML models like Decision Trees (DT), Support Vector Machines (SVR) and Neural Networks (NN). Furthermore, the emerging security concerns in Internet of Things (IoT) contexts and the significance of feature optimisation for enhancing IDS performance are discussed. This review synthesises ideas from current research to offer a thorough overview of IDS technologies and their function in enhancing network security against growing cyber threats.
Keywords: Intrusion Detection Systems, Feature Selection, Machine Learning, Optimisation Techniques

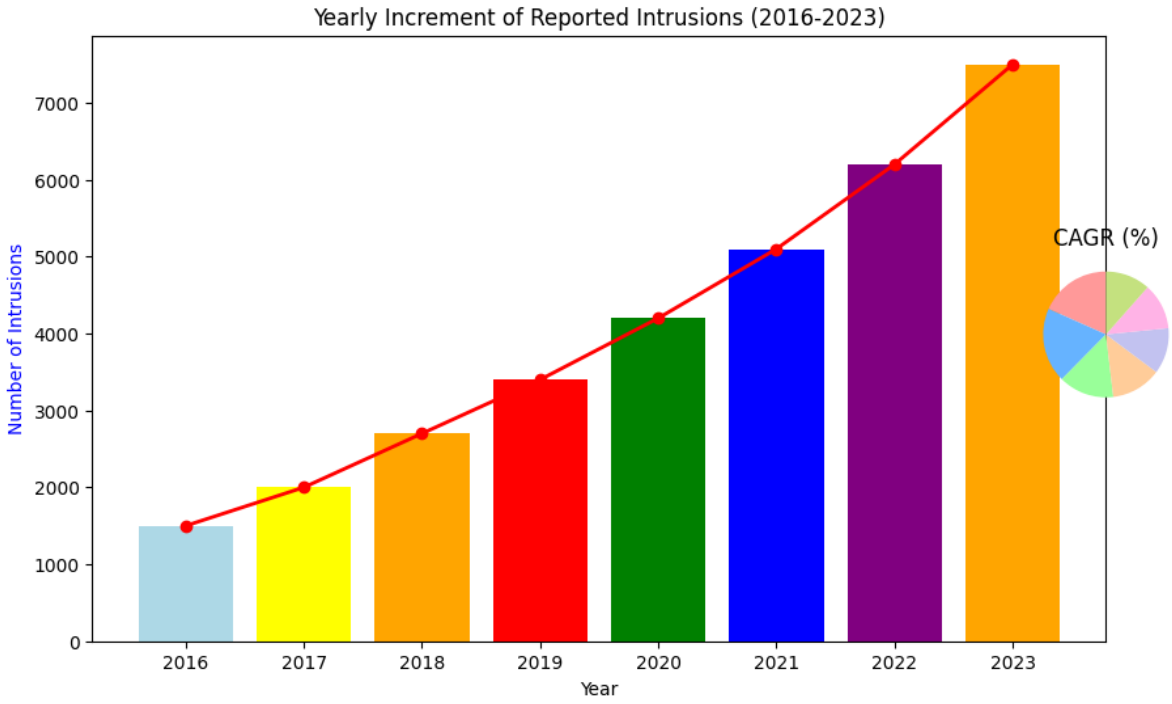
1. Introduction

Nonetheless, as cyber threats evolve in complexity, the shortcomings of conventional Intrusion Detection Systems are becoming evident. Static, signature based detection techniques frequently fail to recognise novel, emerging threats or sophisticated attacks engineered to circumvent detection. This disparity highlights the pressing necessity for sophisticated IDS systems capable of adapting to the changing threat environment. Contemporary IDS must integrate advanced methodologies, like ML, DL and feature optimisation, to improve their detection efficacy. These sophisticated systems can detect known dangers and identify previously unrecognised patterns and abnormalities that may suggest an attack.

The implementation of advanced IDS systems has shown notable progress in recent years, especially with the incorporation of ML techniques and feature optimisation methodologies. These developments have improved the

capability of IDS to identify nuanced and complex intrusions with increased precision, while reducing false positives and negatives. The advancement of IDS has been shaped by the increasing prevalence of IoT devices, which provide novel security concerns owing to their resource limitations and varied operational contexts. Consequently, contemporary IDS must be versatile, effective, and proficient in managing the extensive data produced by these interconnected devices [1].

To elucidate the significance of IDS in the contemporary cyber security ambiance, one must acknowledge the escalating frequency of network intrusions in recent years. The subsequent graph depicts the annual increase in reported intrusion occurrences, highlighting the growing menace of cyber-attacks. This growth highlights the necessity for sophisticated and resilient IDS solutions capable of efficiently identifying and alleviating these threats, thus guaranteeing the protection of network infrastructures.



The above image presents a line graph showing the yearly rise of reported intrusions from 2016 to 2023 [2]. The x-axis represents the years, and the y-axis signifies the number of intrusions. The line graph links the data points to illustrate the increasing trend in incursions during this timeframe. A pie chart is included to depict the compound annual growth rate (CAGR) for the full period, emphasising the total percentage increase in reported incursions.

2. Literature Review

The table below provides a comparative study of IDS techniques emphasising significant findings from recent studies. The table delineates the methodology utilised datasets employed and the notable results attained by various scholars. This research highlights breakthroughs in IDS particularly through the use of ML and optimisation approaches, resulting in enhanced detection accuracy, diminished false positives, and improved management of high dimensional datasets. The results highlight the efficacy of these methods in improving IDS performance across various network security contexts.

Table.1. Comparative Analysis of IDS Methodologies and Results

Author(s)	Methodology	Datasets Used	Key Findings

Kunhare et al. (2020) [3]	Feature selection utilising Random Forest (RF) and PSO	NSL-KDD	Achieved an accuracy of 99.32% with a reduction to 10 out of 41 features, hence minimising false alarms.
Liu et al. (2021) [4]	PSO-LightGBM model with OCSVM	UNSW-NB15	Demonstrates high efficacy in identifying both normal and malicious data, and is effective against small-sample attacks.
Halim et al. (2021) [5]	Genetic Algorithm-based Feature Selection (GbFS)	CIRA-CIC-DOHBrw-2020, UNSW-NB15, Bot-IoT	Attained a classification accuracy of up to 99.80%, demonstrating efficacy in high-dimensional datasets.
Pandeeswari & Jeyanthi (2022) [6]	RF for anomaly detection	ADFA	Demonstrating improved accuracy in attack detection through ML methodologies.
Sivagaminathan et al. (2023) [7]	PSO integrated with KNN and ANN	Various datasets	Demonstrating that the PSO-ANN combination surpassed other models in anomaly detection and feature optimisation.
S. Miryahyaie et al. (2021) [8]	Fuzzy rules with Ant Colony Optimization (ACO)	NSL-KDD	Achieving superior accuracy and reduced error risk relative to conventional approaches.
Disha & Waheed (2022) [9]	GIWRF for feature selection	UNSW-NB15, Network TON_IoT	GIWRF-optimized Decision Tree surpassed alternative models, attaining a greater F1 score.
Musa et al. (2020) [10]	Review of ML classifiers (single, hybrid, ensemble)	Seven datasets	A thorough guide for enhancing intrusion detection systems using ML approaches.
Qureshi et al. (2023) [11]	ML-based IDS	Virtual environment	Demonstrated superior performance compared to classical IDS in identifying unauthorised activity.
Musa et al. (2021) [12]	Review of ML, Bayesian algorithms, meta-heuristics	Various datasets	Comparative review of methodologies to improve IDS capabilities.

Ogundokun et al. (2022) [13]	ICA utilising RF	DARPA KDD 99	Attained an accuracy of 99.6% with a minimal false alarm rate.
Sadia et al. (2024) [14]	CNN for optimized feature selection	WSN datasets	CNN model attained 97% accuracy and a minimal loss metric, surpassing other ML methodologies.
E and S (2024) [15]	MLP with CatBoost and Pelican Optimization Algorithm (POA)	CSE-CIC-IDS2018, AWID, UNSW-NB15	High accuracy and minimum false positives, enhancing threat detection.
Kiran et al. (2023) [16]	Review of ML integration in IDS	Various datasets	ML substantially enhances the performance of intrusion detection systems in detecting network intrusions.
Subbiah et al. (2022) [17]	Boruta feature selection with Grid Search Random Forest (BFS-GSRF)	NSL-KDD	Obtained 99% accuracy, surpassing SVM and KNN.

3. Intrusion Detection System (IDS)

The IDS is an essential security technology employed to oversee network traffic and computer systems for indications of hostile activity or unauthorised access. It improves comprehensive network security by monitoring traffic and system behaviour to identify potential attacks. Upon detecting suspicious behaviour an IDS notifies network administrators, enabling them to implement requisite measures to safeguard data and systems. In contrast to firewalls that obstruct external threats of an IDS concentrates on identifying breaches misuse originating from both external attackers and internal sources and issues notifications upon unauthorised access attempts or system penetration [18].

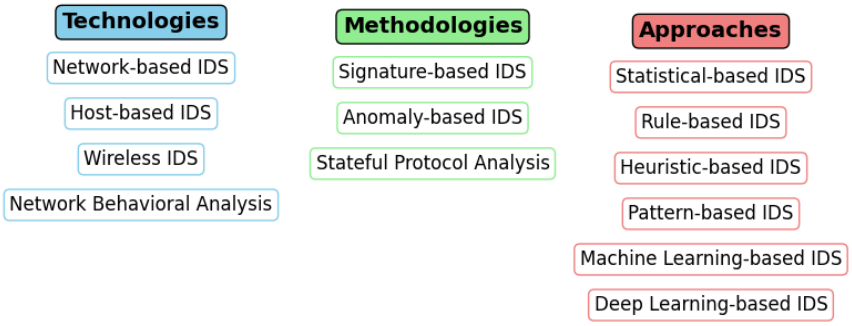


Fig.1.IDS Taxonomy

The above image delineates the taxonomy of IDS by classifying them into three primary categories: Technologies, Methodologies, and Approaches. It delineates many deployment types under Technologies, including network based, host based, and wireless IDS. Methodologies encompass several detection strategies such as signature based; anomaly based, and state full protocol analysis. Approaches encompass several methodologies employed by IDS, including statistical, rule-based, heuristic, pattern based, ML and DL techniques. This taxonomy provides a systematic overview of the varied IDS landscape.

Table 1: Merits and drawbacks of IDS

Advantages	**Disadvantages**
Early Detection of Threats	**Incorrect Positives/Negatives**
IDS provide early detection of possible threats, facilitating timely intervention.	IDS may provide false warnings or overlook genuine threats, so diminishing their efficacy.
Augmented Security Surveillance	**High Resource Demand**
On-going surveillance provides an extensive perspective on network security.	Demands substantial processing and storage resources, potentially impacting system performance.
Regulatory Compliance	**Complexity**
Facilitates compliance with industry requirements and standards for data protection.	Implementing and operating Intrusion Detection Systems can be intricate and necessitate specialised expertise.
Incident Response Assistance	**Potential for Evasion**
Offers essential logs and data for the examination of security incidents.	Advanced adversaries may employ evasive strategies to circumvent detection.
Adaptability	**Limited by Signature Databases**
Can interface with additional security tools for a multi-layered defence.	Signature based IDS may fail to identify novel or emerging threats that are not encompassed by current signatures.
	Operational Overhead
	Frequent updates, adjustments, and alert evaluations may generate operational burdens.

Above Table delineates the merits and demerits of IDS. It emphasises its advantages, including early threat identification and improved security monitoring, while also confronting problems like as false positives, resource demands, and the possibility of evasion.

4. Optimisation of features in Intrusion Detection Systems

Feature optimisation in IDS improves detection accuracy by identifying and refining the most pertinent features. Techniques such as GA, PSO, and Ant Colony Optimisation (ACO) optimise feature subsets to enhance system efficiency and minimise false positives, whereas methods like Recursive Feature Elimination and Principal Component Analysis concentrate on eliminating less significant features or decreasing dimensionality [8].

- **Genetic Algorithms (GA):** Iteratively develops feature subsets across generations to optimise detection efficacy.
- **Particle Swarm Optimisation (PSO):** Simulates social behaviour to identify optimal feature pairings.
- **Ant Colony Optimisation (ACO):** Employs pheromone trails to direct the search for significant characteristics.
- **Simulated Annealing (SA):** Investigates feature subsets and progressively diminishes exploration to identify near optimal solutions.
- **Recursive Feature Elimination (RFE):** Eliminates less significant features according to classifier efficacy.
- **Boruta Algorithm:** Employs RF to discern and preserve important traits.

- **Feature Significance from Tree based Techniques:** Determines significant features utilising approaches such as RF and Gradient Boosting Trees.
- **Principal Component Analysis (PCA):** Minimises dimensionality by emphasising principal components.

Kunhare et al. (2020) tackle the issue of erratic network traffic in IDS by proposing feature selection utilising the RF algorithm to improve efficiency. Their comparative examination of classifiers, integrated with PSO on the NSL-KDD dataset, attained a high accuracy of 99.32% and minimised false alarms by utilising 10 pertinent features from a total of 41. Liu et al. (2021) present a PSO based Gradient Boosting Machine (PSO-LightGBM) model combined with a one class Support Vector Machine (OCSVM) to address IoT security concerns. The model, evaluated using the UNSW-NB15 dataset, exhibited significant effectiveness in identifying both benign and malicious data, particularly for low sample attacks. Halim et al. (2021) present a GA based Feature Selection (GbFS) technique to enhance IDS accuracy. Evaluated on benchmark datasets (CIRA-CIC-DOHBrw-2020, UNSW-NB15, Bot-IoT), GbFS attained a classification accuracy of up to 99.80%, underscoring its efficacy in high dimensional network security contexts. Pandeeswari and Jeyanthi (2022) concentrate on enhancing IoT security with the application of the RF technique for anomaly identification within the ADFA dataset. Their findings underscore the efficacy of machine learning methodologies in improving attack detection precision.

Sivagaminathan et al. (2023) investigate PSO for feature selection in Network IDS, including it with ML models such as K-Nearest Neighbours (KNN) and Artificial Neural Networks (ANN). PSO integrated with ANN surpassed alternative models, demonstrating enhanced anomaly identification and feature optimisation. S. Miryahyaie et al. (2021) introduce an innovative approach that integrates fuzzy association rules with Ant Colony Optimisation (ACO), evaluated using the NSL-KDD dataset. Their methodology exhibited superior accuracy and reduced error risk relative to conventional techniques. Disha and Waheed (2022) present a ML based IDS utilising Gini Impurity based Weighted RF (GIWRF) for feature selection. Their assessment of multiple models on the UNSW-NB 15 and Network TON_IoT datasets indicated that the GIWRF optimized DT surpassed other models for the F1 score. Feature optimisation in IDS include the selection and refinement of pertinent features to improve detection accuracy and efficiency, including methodologies such as Genetic Algorithms, Particle Swarm Optimisation, and Boruta to mitigate computing complexity and minimise false positives.

5. ML Algorithms in IDS

Classification in IDS employs ML methods to categorise network traffic or system behaviour as either legitimate or malicious. Techniques such as Decision Trees (DT), Support Vector Machines (SVM), and K-Nearest Neighbours (KNN) facilitate the identification of patterns and anomalies, whilst sophisticated methods like Neural Networks and Gradient Boosting Machines (GBM) improve accuracy by discerning intricate data linkages.

- **DT:** Employ tree like structures to facilitate decision making based on feature values, offering a transparent and interpretable paradigm for classifying network data.
- **SVM:** Construct hyper planes to distinguish various data classes, proficient in high dimensional spaces and adept at identifying intricate patterns.
- **KNN:** Classifies data points according to the predominant class of their nearest neighbours, effective for detecting anomalies using distance measurements.
- **NN:** utilise layers of interconnected nodes to discern intricate patterns in data, with variations such as CNNs and RNNs employed for distinct data types.
- **RF:** Employ an ensemble of decision trees to enhance classification accuracy and resilience, mitigating overfitting and augmenting performance.
- **Gradient Boosting Machines (GBM)** construct models in a sequential manner, with each model rectifying the flaws of its predecessor, hence improving predicting accuracy and managing intricate interactions.

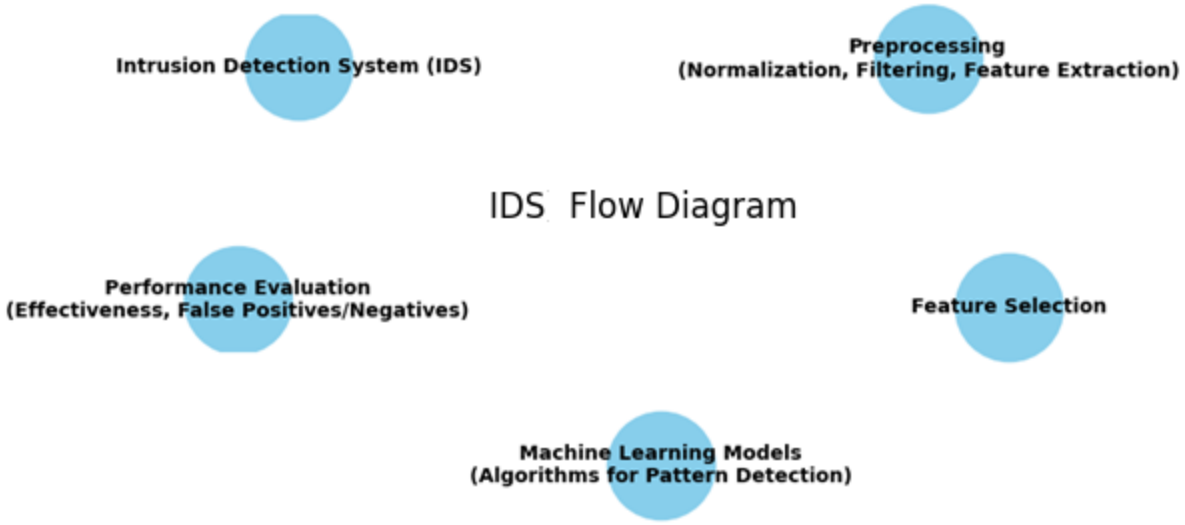
- **Naive Bayes (NB)**: Utilise probabilistic models derived from Bayes' theorem, supposing feature independence, to categorise data and identify anomalies.
- **Clustering algorithms (e.g., K-Means, DBSCAN)** aggregate analogous data points to discern patterns and discover outliers or anomalies in network traffic.

Musa et al. (2020) examine difficulties in network security, emphasising IDS methodologies: signature-based detection and anomaly detection. The research assesses multiple machine learning classifiers single, hybrid, and ensemble across seven datasets, providing an extensive framework for future enhancements in IDS utilising ML methodologies. Qureshi et al. (2023) examine security vulnerabilities in Power Line Communication (PLC) networks, offering a ML based IDS to improve detection efficacy and performance. The suggested approach surpasses conventional IDS in identifying unauthorised actions, as evidenced in a simulated environment. Musa et al. (2021) examine the increase in cyber-attacks, highlighting the significance of IDS in network security. They evaluate multiple performance enhancing methodologies, including ML, Bayesian algorithms, meta-heuristics, swarm intelligence, and Markov neural networks, contrasting them across diverse datasets. Ogundokun et al. (2022) offer a machine learning-based intrusion detection system that integrates Independent Component Analysis (ICA) with RF for feature extraction. Evaluated using the DARPA KDD 99 dataset, the ICA+RF classifier attained an accuracy of 99.6% and exhibited a minimal false alarm rate, indicating enhanced efficacy compared to traditional approaches.

Sadia et al. (2024) concentrate on Wireless Sensor Networks (WSNs), creating an enhanced Network Intrusion Detection System with refined feature selection. The model attained 97% accuracy and a minimal loss metric via a Convolutional Neural Network (CNN), surpassing previous ML methodologies in WSN scenarios. E and S offer a hybrid IDS for WSNs that integrates a multilayer perceptron (MLP) with CatBoost and the Pelican Optimisation Algorithm (POA) for hyper parameter tuning and feature selection. The model demonstrated excellent accuracy and few false positives across diverse datasets, enhancing threat detection. Kiran et al. (2023) underscore the growing necessity for efficient IDS, accentuating the incorporation of ML to improve detection precision. Their research illustrates that ML substantially enhances intrusion detection system efficacy in recognising network intrusions. Subbiah et al. (2022) present an innovative IDS architecture for WSN employing Boruta feature selection in conjunction with Grid Search Random Forest (BFS-GSRF). The BFS-GSRF model attained 99% accuracy on the NSL-KDD dataset, surpassing conventional methods such as SVM and KNN. ML automates the generation of models using data driven learning. Techniques encompass Supervised Learning (utilising labelled data), Unsupervised Learning (examining unlabelled data), Semi-Supervised Learning (integrating labelled and unlabelled data), and Reinforcement Learning (acquiring knowledge via trial and error). These strategies emphasise pattern recognition and decision making with less human involvement.

6. IDS Workflow and Enhancement Cycle

The IDS workflow encompasses the surveillance of network traffic, data processing, feature selection, and model training to identify intrusions. It encompasses continuous performance assessment and optimisation to improve detection precision and system efficacy.



The diagram depicts an IDS workflow, wherein the central IDS supervise the monitoring of network traffic. The data undergoes initial pre-processing involving normalisation, filtering, and feature extraction. Subsequently, pertinent features are chosen to train ML models that detect patterns and abnormalities suggestive of intrusions. The system's performance is assessed through accuracy and false positives/negatives, with data utilised to enhance the models, establishing a continuous cycle of development.

7. Conclusion

This report highlights the essential function of IDS in protecting contemporary networks from progressively complex cyber threats. Conventional IDS methodologies are increasingly ineffective, requiring the implementation of new approaches such as ML, DL, and feature optimisation to improve detection efficacy. The amalgamation of these technologies has markedly enhanced the precision and efficacy of IDS, facilitating the identification of both established and novel threats. As the IoT proliferates, IDS must adapt to confront emerging security issues, guaranteeing strong protection in varied and data-intensive settings. Continuous study and optimisation of IDS are crucial for sustaining robust defences in the always changing realm of cyber security.

REFERENCES

1. Liu, H.; Lang, B. Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. Appl. Sci. 2019, 9, 4396.
2. Talukder, M.A., Islam, M.M., Uddin, M.A. et al. Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction. J Big Data 11, 33 (2024).
3. A.Kiran, S. W. Prakash, B. A. Kumar, Likhitha, T. Sameeratmaja and U. S. S. R. Charan, "Intrusion Detection System Using Machine Learning," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-4.
4. U. S. Musa, M. Chhabra, A. Ali and M. Kaur, "Intrusion Detection System using Machine Learning Techniques: A Review," 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2020, pp. 149-155.
5. K. N. Qureshi, N. Arshad and T. Newe, "Intrusion Detection Systems for Cyber Attacks Detection in Power Line Communications Networks," 2023 31st Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), Naples, Italy, 2023, pp. 193-199.
6. U. S. Musa, S. Chakraborty, M. M. Abdullahi and T. Maini, "A Review on Intrusion Detection System using Machine Learning Techniques," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2021, pp. 541-549.
7. R. O. Ogundokun, S. Misra, A. N. Babatunde and S. Chockalingam, "Cyber Intrusion Detection System based on Machine Learning Classification Approaches," 2022 International Conference on Applied Artificial Intelligence (ICAPAI), Halden, Norway, 2022, pp. 1-6.
8. Kunhare, N.Tiwari, R. & Dhar, J. Particle swarm optimization and feature selection for intrusion detection system. Sādhanā 45, 109 (2020).

9. J. Liu, D. Yang, M. Lian and M. Li, "Research on Intrusion Detection Based on Particle Swarm Optimization in IoT," in IEEE Access, vol. 9, pp. 38254-38268, 2021.
10. Zahid Halim, Muhammad Nadeem Yousaf, Muhammad Waqas, Muhammad Sulaiman, Ghulam Abbas, Masroor Hussain, Iftekhar Ahmad, Muhammad Hanif,An effective genetic algorithm-based feature selection method for intrusion detection systems,Computers & Security,Volume 110,2021,102448,ISSN 0167-4048.
11. G. J. Pandeeswari and S. Jeyanthi, "Analysis of Intrusion Detection Using Machine Learning Techniques," 2022 Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE), Bangalore, India, 2022, pp. 1-5.
12. H. Sadia et al., "Intrusion Detection System for Wireless Sensor Networks: A Machine Learning Based Approach," in IEEE Access, vol. 12, pp. 52565-52582, 2024.
13. E, G., S, S. An optimized intrusion detection model for wireless sensor networks based on MLP-CatBoost algorithm. Multimed Tools Appl 83, 66725–66755 (2024).
14. Sivagaminathan, V., Sharma, M. & Henge, S.K. Intrusion detection systems for wireless sensor networks using computational intelligence techniques. Cybersecurity 6, 27 (2023).
15. S. Miryahyaie, H. Ebrahimpour-komleh and A. M. Nickfarjam, "ACO-based Intrusion Detection Method in Computer Networks using Fuzzy Association Rules," 2021 5th International Conference on Pattern Recognition and Image Analysis (IPRIA), Kashan, Iran, 2021, pp. 1-5, doi: 10.1109/IPRIA53572.2021.9483486.
16. Disha, R.A., Waheed, S. Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. Cybersecurity 5, (2022).
17. S. Subbiah, K. S. M. Anbananthen, S. Thangaraj, S. Kannan and D. Chelliah, "Intrusion detection technique in wireless sensor network using grid search random forest with Boruta feature selection algorithm," in Journal of Communications and Networks, vol. 24, no. 2, pp. 264-273, April 2022, doi: 10.23919/JCN.2022.000002.
18. Venčkauskas, A.; Toldinas, J.; Morkevičius, N.; Sanfilippo, F. An Email Cyber Threat Intelligence Method Using Domain Ontology and Machine Learning. Electronics 2024, 13, 2716.