AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Rferred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(2026)

Multi-Feature Wavelet-Based Network Intrusion Detection Using Machine Learning and Data Balancing Techniques

Rukhaya Khanam[1], R. Rajashekar[2], Rekha Gangula[3*], Sai Naga Poojitha N[1], Guguloth Nithin[1], Polepaka Chaithanya[1]

[1]UG Student, [2]Assistant Professor, [3]Associate Professor and Head, [1,2,3]Department of Computer Science and Engineering (AI&ML)

[1,2,3]Vaagdevi Engineering College, Bollikunta, Warangal, 506005, Telangana, India

*Correspondence: Rekha Gangula (gangularekha@gmail.com)

ABSTRACT

Modern digital infrastructures such as enterprise networks, cloud systems, and IoT environments generate large volumes of dynamic network traffic, including packet data, authentication logs, and system metrics. Real-time monitoring of this data is critical for ensuring network security and preventing unauthorized access. However, traditional intrusion detection methods rely on rule-based mechanisms and manual analysis, which are often inefficient, inflexible, and unable to adapt to evolving cyber threats. These limitations result in delayed detection, high false positive rates, and poor scalability in complex environments. To overcome these challenges, a multi-feature wavelet-based network intrusion detection framework is proposed using machine learning and data balancing techniques. The system incorporates key network features such as packet size, latency, bandwidth utilization, and authentication metrics. Additionally, Discrete Wavelet Transform (DWT)-based features are used to improve signal representation and uncover hidden patterns in network behavior. Machine learning models including K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), and Naive Bayes are applied for both binary anomaly detection and multi-class classification of authentication failures. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is employed, enhancing model performance and reducing bias. Feature scaling through standardization ensures consistency across models. The framework is implemented using a modular pipeline and deployed via a Flask-based web interface for real-time prediction. Performance evaluation using accuracy, precision, recall, and F1-score demonstrates effective detection of anomalies and reliable classification, making the system scalable and suitable for modern network security applications.

Key words: Network Intrusion Detection, Anomaly Detection, Network Security, Real-time Monitoring.

1. INTRODUCTION

In the rapidly evolving technological landscape, network monitoring systems generate large volumes of traffic data that are essential for security analysis and decision-making. Ensuring the reliability and integrity of this data is critical, as inconsistencies may lead to incorrect threat detection and system vulnerabilities [1]. This challenge becomes more significant in modern networked environments and distributed systems, where traffic flows are dynamic and complex shown in figure 1. The difficulty of maintaining data integrity is further amplified due to varying traffic patterns and potential cyber-attacks in real-world conditions [2]. Traditional anomaly detection approaches, such as signature-based Intrusion Detection Systems (IDS), are effective for known threats but fail to detect new and evolving attack patterns. As network threats continue to advance, there is a strong need for adaptive and intelligent detection mechanisms capable of identifying anomalies and ensuring robust system security [3]. This research addresses the identified gap by proposing a wavelet-enhanced machine learning approach for effective network anomaly detection and intrusion

analysis [4]. Discrete Wavelet Transform (DWT) is utilized to extract meaningful features from network traffic data by decomposing it into different frequency components and capturing localized variations.

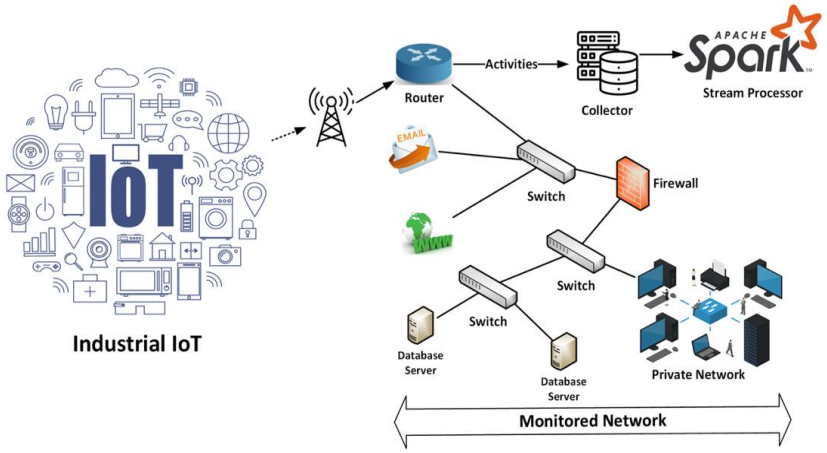

Figure 1: Monitoring of IoT-based network traffic flows

The Haar wavelet was selected due to its simplicity and efficiency in handling non-stationary patterns, enabling the detection of both sudden anomalies and persistent irregular behaviours in network activity. Haar wavelets provide strong time-frequency localization, which is essential for identifying subtle deviations in complex traffic patterns. Their low computational complexity makes them suitable for scalable and real-time analytical systems [5]. The extracted wavelet features are integrated with machine learning algorithms to enhance classification performance. Additionally, distance-based evaluation techniques support the identification of deviations from normal behaviour, improving detection accuracy. This combined approach enables efficient, adaptive, and reliable intrusion detection in modern network environments.

2.LITERATURE SURVEY

Rekha Gangula et al. [6] proposed a stacked autoencoder-based intrusion detection framework for Internet of Things (IoT) environments. The model extracted hierarchical deep features from network traffic data. A weighted loss function improved class imbalance handling and enhanced detection accuracy.

Paracha, et al. [7] proposed an investigation model based on AI/ML techniques that can analyse network traffic and behavioural patterns to identify any prior or potential cyberattacks. The proposed AI-based network forensics model speeds up the investigation process, boosting network monitoring without human intervention. This also aimed to provide timely and accurate information to network administrators for quick and effective decisions, enabling them to avoid and circumvent future cyberattacks. Chiriac, et al. [8] presented an approach for implementing a generic model of a network-based intrusion detection system for Industry 4.0 by integrating the computational advantages of the Nvidia Morpheus open-source AI framework. The solution is modularly built with two pipelines for data analysis. The pipelines used a pre-trained XGBoost (eXtreme Gradient Boosting) model that achieved an accuracy score of up to 90%. The proposed IDS has a fast rate of analysis, managing more than 500,000 inputs in almost 10 s, due to the application of the federated learning methodology.

Abuali, et al. [9] aimed to propose a SVM based deep learning system that will classify the data extracted from servers to determine the intrusion incidents on social media. To implement deep learning-based IDSs

for multiclass classification, the CSE-CIC-IDS 2018 dataset has been used for system evaluation. The CSE-CIC-IDS 2018 dataset was subjected to several preprocessing techniques to prepare it for the training phase. The proposed model has been implemented in 100,000 instances of a sample dataset. Toldinas, et al. [10]. proposed a novel approach for network intrusion detection using multistage deep learning image recognition. The network features are transformed into four-channel (Red, Green, Blue, and Alpha) images. The images then are used for classification to train and test the pre-trained deep learning model ResNet50. The proposed approach is evaluated using two publicly available benchmark datasets, UNSW-NB15 and BOUN Ddos. On the UNSW-NB15 dataset, the proposed approach achieves 99.8% accuracy in the detection of the generic attack. On the BOUN DDos dataset, the suggested approach achieves 99.7% accuracy in the detection of the DDos attack and 99.7% accuracy in the detection of the normal traffic.

Zhang, et al. [11] proposed a novel algorithm based on both wavelet leader multifractal analysis (WLM) and machine learning (ML) principles. In earlier research on unmanned aerial systems (UAS), intrusion detection systems (IDS) based on multifractal (MF) spectral analysis have been used to provide accurate MF spectrum estimations of network traffic. Such an estimation is then used to detect and characterize flooding anomalies that can be observed in an unmanned aerial vehicle (UAV) network. However, the previous contributions have lacked the consideration of other types of network intrusions commonly observed in UAS networks, such as the man in the middle attack (MITM). In this work, this promising methodology has been accommodated to detect a spoofing attack within a UAS. Ali, et, al. [12] implemented various deep learning models, including multilayer perceptron (MLP), convolutional neural network (CNN), and long short-term memory (LSTM), alongside traditional machine learning algorithms such as logistic regression, naive Bayes, random forest, K-nearest neighbors, and decision trees.

Mao, et, al. [13] employed a dynamic routing mechanism to map sample feature vectors into robust class vector representations, achieving superior generalization when detecting unseen attack types. Compared to existing FCN–Transformer models, MFEI-IDS incorporates inductive learning to handle data imbalance and small-sample scenarios. Experiments on ISCX 2012 and CIC-IDS 2017 datasets show that MFEI-IDS outperforms mainstream IDS methods in accuracy, precision, recall, and F1-score, excelling in cross-dataset validation and demonstrating strong generalization capabilities. Mari, et al. [14] demonstrated a way to create adversarial instances of network traffic that can be used to evade detection by a machine learning-based IDS. Moreover, this traffic can be used for training in order to improve performance in the case of new attacks. Thus, a generative adversarial network (GAN) i.e., an architecture based on a deep-learning algorithm capable of creating generative models was implemented.

Rekha Gangula et al. [15] proposed an IoT intrusion detection system using an Enhanced Flower Pollination Algorithm with an ensemble classifier. The optimization technique selected optimal features and parameters. The ensemble model improved classification robustness and reduced false alarms.

3. PROPOSED SYSTEM

The proposed system architecture processes raw network traffic by applying preprocessing and wavelet-based feature extraction to generate informative and discriminative inputs for analysis. Discrete Wavelet Transform (DWT) is used to capture variations and hidden patterns within network data, improving feature quality. These processed features are then fed into machine learning models that perform anomaly detection and authentication failure classification with high accuracy. The system is designed to identify subtle deviations in network behavior, including unusual traffic patterns and irregular login activities. By integrating wavelet-based features with classification algorithms, the framework enhances detection

capability and reduces false positives. Data balancing techniques further improve model reliability when handling imbalanced datasets. The architecture supports both training and real-time prediction through an efficient pipeline. The system provides a scalable and effective solution for network intrusion detection, as illustrated in Figure 2.

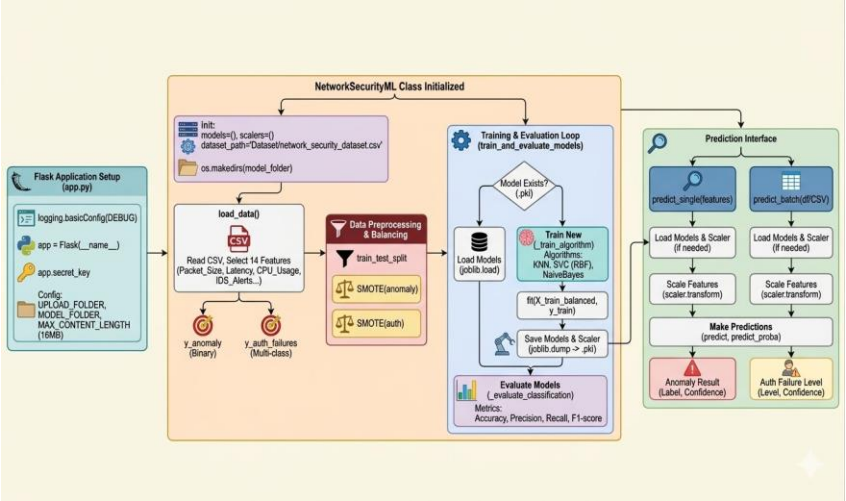

Figure 2: Proposed system architecture

Raw Network Data Acquisition: The operational flow begins with the continuous acquisition of raw network traffic from distributed security endpoints such as routers, gateways, authentication servers, and firewall logs. This stage is crucial because long-range intrusion patterns require diverse multi-source data to identify deviations and subtle attack footprints. The system collects packet headers, connection metadata, frequency of requests, session durations, and access logs that capture both successful and failed authentication attempts. These inputs serve as the foundational evidence for detecting anomalies and unusual traffic bursts that may not be visible when using single-source monitoring systems.

Data Preprocessing and Cleaning: Once data enters the pipeline, the preprocessing layer transforms it into a standardized and noise-free structure suitable for advanced analytics. This step removes corrupted packets, duplicate entries, and incomplete authentication logs, ensuring that only meaningful and validated data flows into further processing. Normalization techniques are applied to maintain uniform scaling across diverse numerical attributes such as packet size, request count, and login intervals. Temporal alignment is also performed so that long-range dependencies such as slow attacks or progressive intrusion attempts—can be captured more accurately across chronological event sequences.

Multi-Scale Wavelet Decomposition: The next stage introduces the core innovation of the proposed system: multi-scale wavelet decomposition. Unlike traditional filters or time-domain transformations, wavelets analyze network signals at multiple resolutions simultaneously, allowing the system to detect both sudden anomalies (burst attacks) and long-term deviations (slow-paced intrusions). During this process, the raw data is decomposed into approximate (low-frequency) and detailed (high-frequency) components, revealing hidden fluctuations that attackers often exploit to bypass classic IDS models. This multi-resolution representation significantly enhances the system's ability to sense micro-level variations in packet flow and macro-level authentication irregularities.

Feature Engineering and Selection: After wavelet transformation, the system begins engineering features that capture the intelligence embedded within the decomposed signals. This step extracts statistical

indicators such as entropy, variance, spike density, packet symmetry ratios, and login deviation scores, which represent refined patterns of normal and abnormal user behavior. High-dimensional features are then filtered using selection techniques to remove redundant or weak predictors, ensuring that the final dataset contains only the most relevant attributes. This optimization enhances computational efficiency and prevents model overfitting, especially when analyzing large-scale enterprise network traffic.

Model Training Using Proposed Machine Learning Algorithms: The processed dataset is then fed into a multi-model learning environment where algorithms like NBC, KNN, and SVC are rigorously trained. Each model is exposed to thousands of labeled samples representing normal behavior, anomalies, and authentication outcomes categorized as positive, negative, or neutral. Cross-validation ensures that the model generalizes well to unseen traffic patterns, while hyperparameter tuning enhances decision boundaries, probabilistic estimations, and nearest-neighbor distances. This training phase equips the proposed system to differentiate between benign fluctuations and genuine attacks with improved precision and reduced false alarms.

Real-Time Prediction and Intrusion Detection: During live deployment, the system performs high-speed inference by transforming incoming traffic into wavelet-based features and applying the trained models to classify each event. Every network request is rapidly evaluated to determine whether it represents normal behavior, suspicious activity, or an authentication failure category. Real-time alerts are triggered for anomalies such as brute-force attacks, session hijacking attempts, excessive login failures, or unusual traffic bursts originating from unknown sources. By processing data instantly and intelligently, the system ensures robust early detection of intrusions before they escalate into major security breaches.

Continuous Learning and Adaptive Feedback Loop: The final stage introduces adaptive intelligence by feeding system outputs back into the learning layer. Detected anomalies, new attack signatures, and evolving traffic behaviors are stored to update the model's knowledge base over time. This feedback loop ensures resilience against emerging threats, supporting continuous retraining and self-improvement without requiring manual reconfiguration. As attack patterns evolve—whether through AI-based evasion or long-range stealth techniques—the system dynamically adjusts thresholds, improves classification boundaries, and maintains high detection performance in complex defense-network environments.

4. RESULTS ANALYSIS

The results demonstrate the effectiveness of the proposed wavelet-based network intrusion detection framework in analyzing and classifying network traffic. The system successfully integrates data preprocessing, feature extraction, and machine learning models within a unified dashboard for comprehensive analysis. Exploratory data analysis highlights the distribution and characteristics of normal and anomalous network activities, providing insights into dataset composition. Performance evaluation shows that the Naive Bayes classifier achieves superior accuracy compared to KNN and SVC models. Visualizations such as pie charts, histograms, and scatter plots help in understanding feature behavior and relationships.

Figure 3 illustrates the distribution of anomalous and normal network traffic instances using a pie chart visualization. The chart shows that approximately 72.9 percent of the dataset corresponds to normal network activity, while 27.1 percent represents anomalous network events. This graphical representation highlights the proportion of security events present in the dataset. Understanding this class distribution is important when designing anomaly detection models since class imbalance may affect classification

performance. The visualization therefore provides an overview of how network events are distributed across normal and anomalous categories.

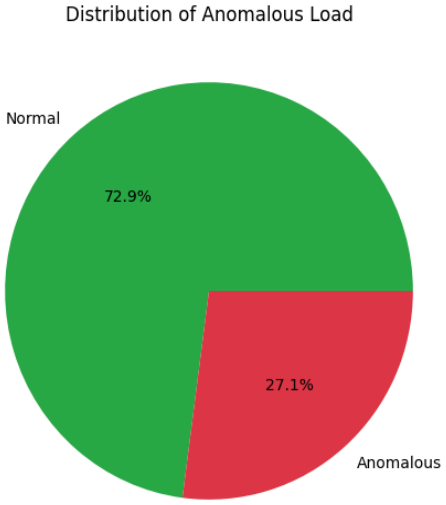

Figure 3: Visualizing anamoly distribution using pie chart.

Figure 4 depicts the distribution of important network traffic features using histogram based visualizations. The figure includes histograms for packet size, transmission rate, latency, and CPU usage. Packet size values range approximately between 500 and 1500 units, while transmission rate values range between about 50 and 200 units. Latency values are distributed between roughly 10 and 100 units, and CPU usage ranges between approximately 5 and 90 percent. These distributions help researchers understand how network characteristics vary across the dataset. Such analysis is useful for identifying feature patterns and detecting potential outliers before model training.

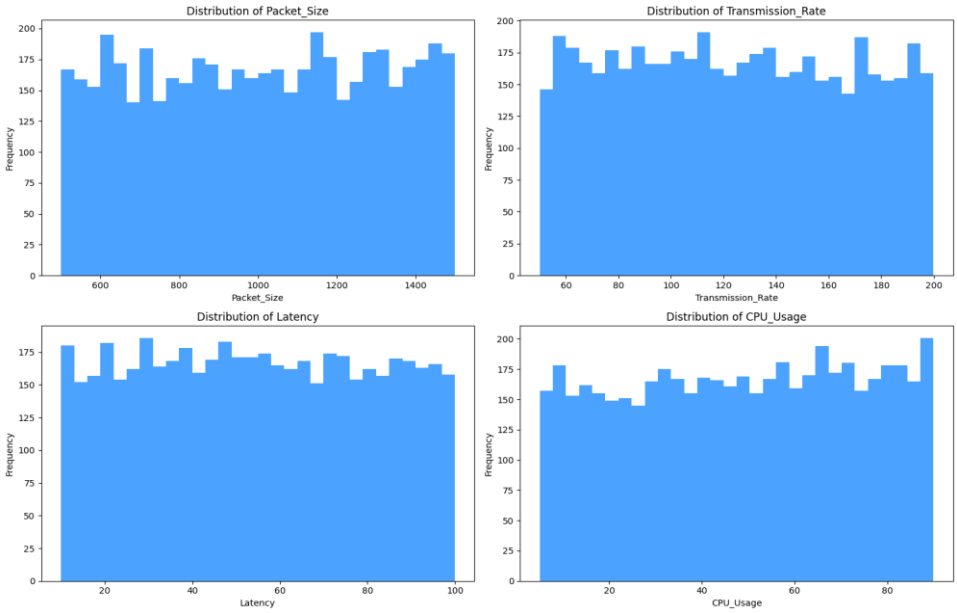

Figure 4: Visualizing Key Feature Distributions using Histogram.

Figure 5 presents the evaluation results obtained after training the selected machine learning models. For anomaly detection classification, Naive Bayes achieved an accuracy of 95.40 percent, precision of 95.89 percent, recall of 95.40 percent, and F1 score of 95.49 percent. The KNN model achieved an accuracy of 82.40 percent with an F1 score of 83.23 percent, while the SVC model achieved an accuracy of 72.30 percent with an F1 score of 73.82 percent. In authentication failure classification, Naive Bayes achieved 100 percent accuracy, precision, recall, and F1 score.

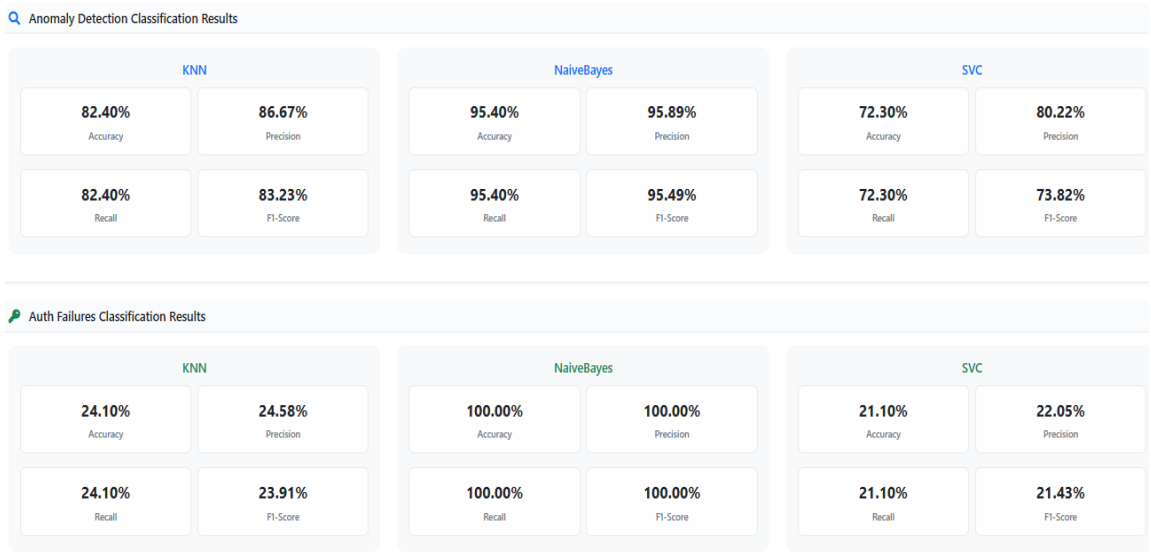

Figure 5: Model Performance Evaluation Interface after Training the models.

Figure 6 shows the prediction results generated by the system after processing the provided network traffic data. The anomaly detection module predicts the network activity as normal with a confidence level of approximately 97.72 percent. In addition, the authentication failure classification module predicts level 9 with 100.00 percent confidence. These prediction results demonstrate the ability of the trained machine learning model to analyze network features and provide security related insights.

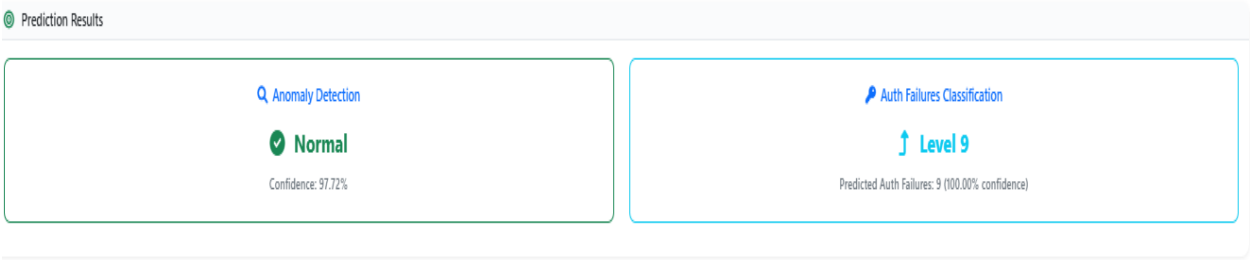

Figure 6: Output of the Test Data.

5. CONCLUSION

The research presents a robust Flask-based web application that seamlessly integrates machine learning for intelligent network security analysis. The system leverages a network dataset enriched with wavelet-derived attributes, particularly DWT_Feature_1, to strengthen anomaly detection and authentication failure classification. Multiple classifiers, including KNN, SVC, and NBC, are evaluated, with NBC achieving superior performance such as 95.40% accuracy, precision, recall, and F1-score for anomaly detection, and

a perfect 100% across all metrics for authentication failure classification. In contrast, KNN and SVC demonstrate comparatively lower performance, achieving 82.40% and 72.30% anomaly detection accuracy, and only 24.10% and 21.10% accuracy for authentication failure prediction. Performance optimization strategies include SMOTE-based class balancing to address dataset skew (1,354 anomalous vs. 3,646 normal samples), StandardScaler-driven feature normalization for stable learning, and joblib-based model persistence for efficient deployment. These enhancements collectively improve predictive reliability, computational efficiency, and real-time operational readiness of the intrusion detection framework.

REFERENCES

[1]. Li, D.; Wang, Y.; Wang, J.; Wang, C.; Duan, Y. Recent Advances in Sensor Fault Diagnosis: A Review. Sens. Actuators A Phys. 2020, 309, 111990.

[2]. Fatima, N.; Riaz, S.; Ali, S.; Khan, R.; Ullah, M.; Kwak, D. Sensors Faults Classification and Faulty Signals Reconstruction Using Deep Learning. IEEE Access 2024, 12, 100544–100558.

[3]. Yang, J.W.; Lee, Y.D.; Koo, I.S. Convolutional Autoencoder-Based Sensor Fault Classification. Int. Conf. Ubiquitous Future Netw. 2018, 2018, 865–867

[4]. Poojari, R. (2025). A Comparative Analysis of Fine-Tuning Versus Retrieval-Augmented Approaches for Enhancing Healthcare-Centric Large Language Models..

[5]. Purmani, S. S. R. (2025). Streamlining IT operations and service management with agile frameworks. European Journal of Advances in Engineering and Technology, 12(4), 76–81.

[6]. Rekha Gangula, V. Murali Mohan, Ranjeeth Kumar. M, "Stacked autoencoder with weighted loss function for intrusion detection in IoT application". Multimed Tools Appl (2024). https://doi.org/10.1007/s11042-024-19962-7.

[7]. Paracha, M.A.; Jamil, S.U.; Shahzad, K.; Khan, M.A.; Rasheed, A. Leveraging AI for Network Threat Detection—A Conceptual Overview. Electronics 2024, 13, 4611. https://doi.org/10.3390/electronics13234611

[8]. Chiriac, B.-N.; Anton, F.-D.; Ioniță, A.-D.; Vasilică, B.-V. A Modular AI-Driven Intrusion Detection System for Network Traffic Monitoring in Industry 4.0, Using Nvidia Morpheus and Generative Adversarial Networks. Sensors 2025, 25, 130. https://doi.org/10.3390/s25010130

[9]. Kalae, U. K. (2021). Enhancing data analytics and reporting efficiency using Power BI and SQL in cloud computing environments. Journal of Computational Analysis and Applications, 29(6), 2021. https://doi.org/10.48047/jocaaa.2021.29.06.48

[10]. Toldinas, J.; Venčkauskas, A.; Damaševičius, R.; Grigaliūnas, Š.; Morkevičius, N.; Baranauskas, E. A Novel Approach for Network Intrusion Detection Using Multistage Deep Learning Image Recognition. Electronics 2021, 10, 1854. https://doi.org/10.3390/electronics10151854

[11]. Zhang, R.; Condomines, J.-P.; Lochin, E. A Multifractal Analysis and Machine Learning Based Intrusion Detection System with an Application in a UAS/RADAR System. Drones 2022, 6, 21. https://doi.org/10.3390/drones6010021

[12]. Ali, M.L.; Thakur, K.; Schmeelk, S.; Debello, J.; Dragos, D. Deep Learning vs. Machine Learning for Intrusion Detection in Computer Networks: A Comparative Study. Appl. Sci. 2025, 15, 1903. https://doi.org/10.3390/app15041903

[13]. Mao, J.; Yang, X.; Hu, B.; Lu, Y.; Yin, G. Intrusion Detection System Based on Multi-Level Feature Extraction and Inductive Network. Electronics 2025, 14, 189. https://doi.org/10.3390/electronics14010189

[14]. Mari, A.-G.; Zinca, D.; Dobrota, V. Development of a Machine-Learning Intrusion Detection System and Testing of Its Performance Using a Generative Adversarial Network. Sensors 2023, 23, 1315. https://doi.org/10.3390/s23031315

[15]. Rekha Gangula, V. Murali Mohan, Ranjeeth Kumar. M, "Network Intrusion Detection System for Internet of Things based on Enhanced Flower Pollination Algorithm and Ensemble Classifier", Concurrency and Computation: Practice and Experience – Wiley Online Library, Accepted on 05 April 2022, ISSN: 1532-0634 (online), ISSN: 1532-0626(print), Impact Factor: 1.536. Standard age of Journal - 21 Yrs. (since 2001).