Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 4, 2025

INTEGRATING NETWORK INTRUSION DETECTION WITH MACHINE LEARNING TECHNIQUES FOR ENHANCED NETWORK SECURITY

Muqaddas Salahuddin[*1], Fahim Uz Zaman[2], Gohar Mumtaz[3], Muhammad Zohaib Khan[4], Maria Kainat[5], Sammia Hira[6], Fakhra Parveen[7], Rashid Mahmood[8]

[*1,3,6,7,8]Faculty of Computer Science and Information Technology, Superior University, Lahore, 54000, Pakistan
[2]Department: Digital Technologies, Newcastle College University Centre (NCUC) Rye Hill Campus, Scotswood Rd, Newcastle upon Tyne NE4 7SA, United Kingdom
[4]Department of Information Technology, Shaheed Mohtarma Benazir Bhutto Institute of Trauma, Karachi, Pakistan
[5]Department of Computer Science, Faculty of Sciences, University of Agriculture, Faisalabad, 38000, Punjab, Pakistan

[*1]muqaddassalahuddin60@gmail.com, [2]faheemuzzaman@hotmail.com, [3]gohar.m@superior.edu.pk, [4]zohaib_khan2017@yahoo.com, [5]mariakainat33@gmail.com, [6]hirac53oort@gmail.com , [7]parveenfakhra13@gmail.com, [8]rmbhatti88@gmail.com

DOI: https://doi.org/10.5281/zenodo.15253983

Keywords
Intrusion Detection Systems (IDS), Network Security, Fuzzy C-Means Clustering, K-Nearest Neighbor (KNN), Naïve Bayes (NB), Logistic Regression (LR), and Stochastic Gradient Descent (SGD).

Article History
Received on 12 March 2025
Accepted on 12 April 2025
Published on 21 April 2025

Copyright @Author
Corresponding Author: *
Muqaddas Salahuddin[*1]

Abstract

Cybersecurity attacks are more common than ever in today's globally networked society, which makes having strong intrusion detection systems (IDS) is essential. In this study, a hybrid IDS model is presented that combines fuzzy C-Means clustered with classification techniques including Naïve Bayes (NB), K-Nearest Neighbor (KNN), Logistic Regression (LR), and Stochastic Gradient Descent (SGD). To increase the accuracy of detection and resilience to changing cyberattacks, sophisticated feature selection approaches are used. The effectiveness of this method is confirmed through extensive testing with the NIDS dataset. By overcoming the limitations of conventional IDS, this study strengthens defenses against sophisticated assaults and uses machine learning to increase network security.

INTRODUCTION

In this constantly changing digital landscape, network security has become paramount due to the proliferation of interconnected devices and systems. These networked environments introduce vulnerabilities, potentially resulting in a toxic combination of the negative aspects of cyberspace. Among the layers of network structure, the data link layer is the most vulnerable, and if compromised, it can lead to a total breach of network security and integrity [1]. This vulnerability is even more pronounced in wireless networks, such as ad-hoc vehicle networks and wireless sensor networks, due to their reliance on shared communication channels, making them particularly susceptible to external cyber threats [2]. The structural ideas of these approaches are depicted in Figure 1.

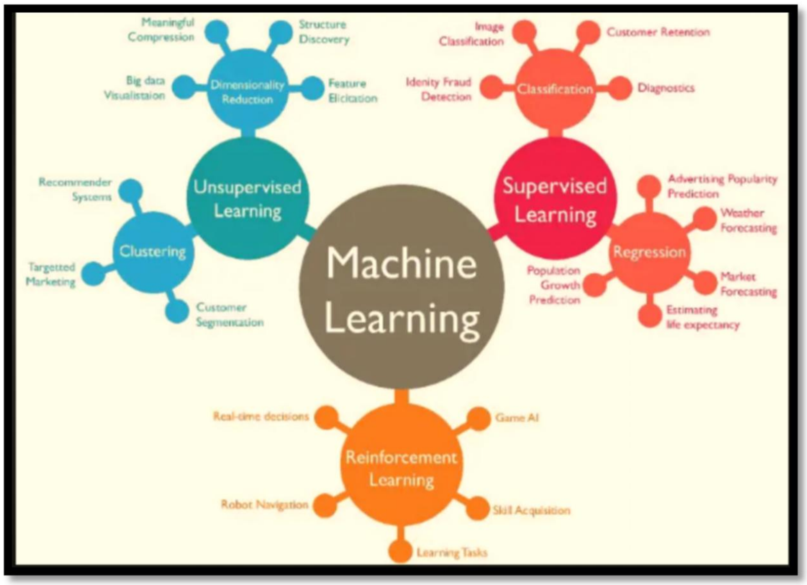

Figure 1: Structuring Machine-Learning Concepts [2]

These networks face significant external cyber threats. Intrusion detection is crucial for cybersecurity, identifying risks and mitigating threats. Traditional systems rely on signature-based methods but struggle with evolving cyber threats [3]. Advanced ML techniques, including supervised, unsupervised, and reinforcement learning, improve IDS adaptability. Methods like neural networks and SVM enhance feature selection and handle large datasets effectively [4]. ML is also applied in IoT and medical IoT security [5]. Emerging technologies like blockchain, federated learning, and XAI improve IDS transparency and security [6,7].

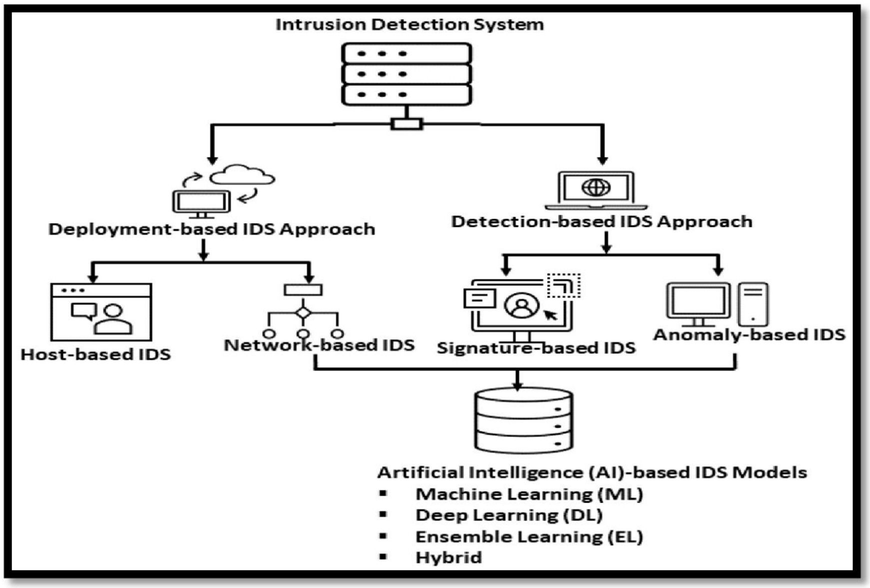

Figure 2: Intrusion Detection Using Explainable Artificial Intelligence (XAI)

This study proposes a hybrid IDS model integrating Fuzzy C-Means with ML classifiers (LR, NB, SGD, KNN) to improve scalability and detection accuracy against changing threats. Using the NIDS dataset, it addresses IDS vulnerabilities to sophisticated attacks. The study includes five sections: introduction, related research (Section 2), methods and tools (Section 3), modeling and results, and conclusions (Section 5).

1. SYSTEMETIC LITERATURE REVIEW (SLR):

Intrusion Detection Systems (IDS) are an integral component of cybersecurity strategies, combining various mechanisms to detect and prevent unauthorized activities across networks. IDS can be categorized into several types, including signature-based IDS, which rely on a database of known threats to detect malicious activities by matching system activity patterns [8]. Behavior-based IDS, also referred to as anomaly-based systems, detect anomalies by comparing network behavior to established norms [9]. Host-based Intrusion Detection Systems (HIDS) monitor individual endpoints or host-based systems, with a particular focus on detecting anomalous behavior or suspicious activities, which is especially useful in financial institutions [10]. Network-based Intrusion Detection Systems (NIDS) monitor traffic at key points in the network, identifying suspicious activity by analyzing network flow patterns [11]. Protocol-based IDS analyze network protocols and headers to detect unusual activity, while statistical IDS rely on statistical analysis to identify deviations from expected behavior [12, 13].

Wireless Intrusion Detection Systems (WIDS) safeguard wireless networks by addressing the specific security challenges unique to such environments [14]. Hybrid IDS combine multiple detection techniques to improve accuracy, while signature-less IDS leverage advanced methods like machine learning and behavior profiling to identify novel threats without relying on predefined signatures [15, 16]. In cloud environments, Intrusion Detection and Prevention Systems (IDPS) protect cloud resources by detecting and mitigating threats targeting cloud infrastructure [17]. Various ML techniques have been applied to IDS. Linear regression is a straightforward statistical method used to analyze relationships between variables in network traffic and operational irregularities. Logistic regression is effective for binary classification tasks, making it suitable for distinguishing between malicious and normal network activity in IDS [18]. Despite its simplicity, decision trees provide an intuitive classification structure, though they may require real-time parameter tuning to avoid overfitting [19].

Support Vector Machines (SVMs) generate hyperplanes in high-dimensional spaces for binary classification, offering robust generalization against overfitting and handling both linear and non-linear data efficiently [20]. Naïve Bayes classifiers, based on Bayes' theorem, offer probabilistic classification but assume feature independence, which may limit accuracy in some IDS applications [21]. The K-Nearest Neighbors (KNN) algorithm, while computationally intensive, excels in detecting anomalies and classifying network traffic based on similarity metrics [22]. Ensemble methods like Random Forest combine predictions from multiple decision trees, providing improved accuracy and robustness in handling high-dimensional data [23]. The integration of machine-learning methods with IDS marks a significant advancement in cybersecurity by enhancing the detection of security threats [24]. Such models combine the strengths of multiple machine-learning techniques, improving detection rates and reducing false positives [25]. However, challenges remain, particularly in the scalability and performance of machine learning-based IDS in large-scale networks. Moreover, adversarial machine learning techniques, which automate evasion strategies, pose a growing threat to machine learning-based IDS, emphasizing the need for robust defenses. Scalability issues, real-time processing demands, and the need for continuous updates to combat evolving threats highlight the complexities of integrating IDS with machine learning [26]. Nevertheless, continued research and innovation in this area hold promise for the development of more effective and efficient intrusion detection systems, as highlighted by recent studies and surveys in the field [29]. By addressing these challenges and leveraging advancements in both machine learning and cybersecurity, researchers can develop IDS solutions that enhance security across a wide range of network environments.

2. PROPOSED METHODOLOGY:

The creation of the breakdown structure described in this section is known as the Intrusion Detection System (IDS). The following are the steps involved:

1. Data is retrieved from Google Drive.
2. Preprocessing includes data cleaning, Random Forest analysis, and Standard Scaler normalization, converting variable frequencies into standardized values.
3. Standardization ensures consistent input for the model.
4. Fuzzy C-Means clustering enhances accuracy, recall, precision, F1-score, sensitivity, and specificity.
5. The dataset is divided into 25% testing and 75% training. The final hybrid model integrates Fuzzy C-Means clustering with LR, KNN, SGD, and NB for intrusion detection.

Figure 3 illustrates the IDS Network Security architecture.

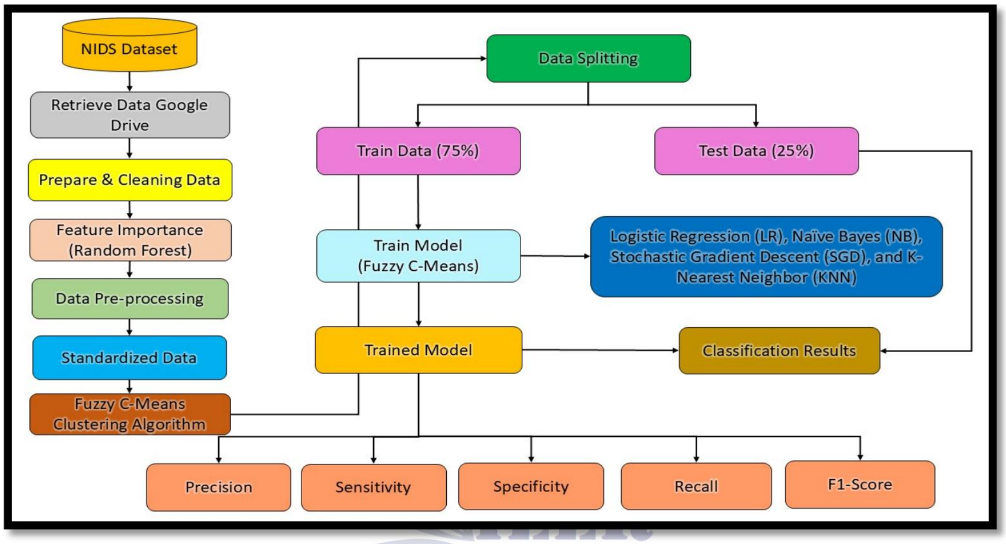
NIDS Dataset
Retrieve Data Google Drive
Prepare & Cleaning Data
Feature Importance (Random Forest)
Data Pre-processing
Standardized Data
Fuzzy C-Means Clustering Algorithm
Data Splitting
Train Data (75%)
Test Data (25%)
Train Model (Fuzzy C-Means)
Logistic Regression (LR), Naïve Bayes (NB), Stochastic Gradient Descent (SGD), and K-Nearest Neighbor (KNN)
Trained Model
Classification Results
Precision
Sensitivity
Specificity
Recall
F1-Score

Figure 3: Intrusion Detection System (IDS) Network Security Architecture

3.1. PREPROCESSING

The preprocessing stage is crucial for cleaning and clustering data to extract relevant and meaningful information. This stage ensures the quality of the input data by applying data cleaning techniques, standardization, and normalization. The clustering process involves techniques like fuzzy c-means to prepare the data for further classification.

Data Collection

The dataset used in this study is the Intrusion Detection Systems (IDS) dataset, which was sourced from Kaggle, a widely recognized platform for machine learning datasets. This dataset was initially utilized in a study conducted by **Elzaridi et al [27].** It comprises internet traffic logs collected by a cyber-intrusion detection system, capturing traces of network traffic influenced by cyberattacks, although only partial evidence is available to confirm the occurrence of these attacks. The dataset includes 42 attributes related to Internet of Things (IoT) devices and consists of 25,192 entries. Among these, 13,449 entries represent normal network activity, while 11,743 entries correspond to anomalies linked to computer virus-related defects or other cyber-intrusion events **[25].**

Figure 4 provides a comparative analysis of the frequency and distribution of cyber-intrusion events within the dataset. Normal activity is represented by a value of 0, while anomalies indicating cyber intrusions are represented by a value of 1. This comparative analysis offers valuable insights into the patterns and prevalence of security breaches, highlighting the dataset's significance for developing an effective intrusion detection system.

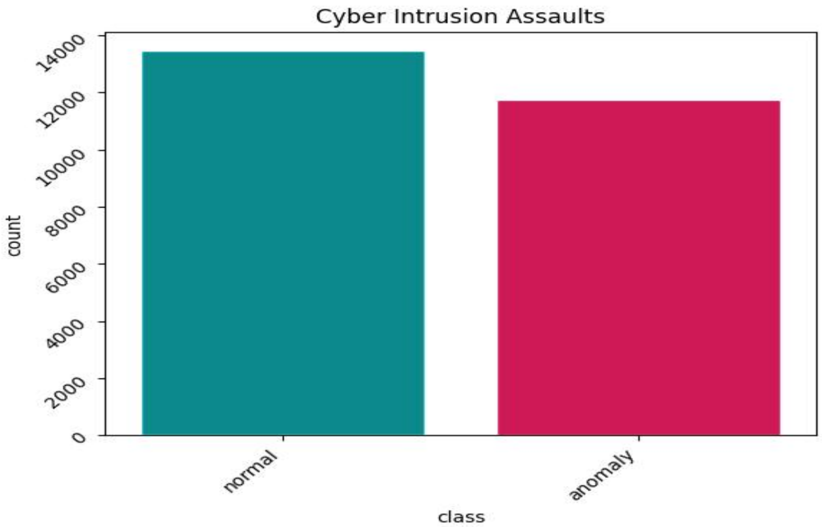

Figure 4: Facts of Cyber-Intrusion Assaults numbers

FEATURE SELECTION: In a machine learning model, each input feature is assessed and assigned a score based on its significance, reflecting how much influence it has on the model's predictions. The more vital the feature is to the outcome, the higher the score it receives [32,33]. By retaining the features with the highest scores and discarding those with lower scores, which are deemed less influential, this process helps reduce the model's dimensionality. This reduction not only simplifies the model but also enhances its overall performance and efficiency**.**

In this study, Recursive Feature Elimination (RFE) was employed as the primary feature selection method to identify the most critical attributes from the Intrusion Detection Systems (IDS) dataset. RFE iteratively eliminates the least important features, leveraging advanced ML algorithms to evaluate the applicable of every feature to the target variable. Unlike traditional RFE methods, this approach integrates predictive algorithms, improving both the accuracy and interpretability of the model while reducing computational overhead [35,36]. This makes the hybrid Intrusion Detection System (IDS) model more efficient for predicting network intrusion attacks.

Table 1 presents the comprehensive scores for the key IDS features, providing insights into their relevance for intrusion detection. These critical features are visually represented in **Figure 5**, emphasizing their importance in identifying network security threats.

Table 1: Intrusion Detection Systems (IDS) Feature Importance Dataset

Serial No.	Features	Features Importance Scores
1	dst_bytes	0.193906
2	src_bytes	0.151329
3	same_srv_rate	0.088118
4	diff_srv_rate	0.073638
5	dst_host_srv_count	0.068845
6	logged_in	0.055464
7	dst_host_same_srv_rate	0.052811
8	dst_host_same_src_port_rate	0.044682
9	dst_host_diff_srv_rate	0.042435
10	count	0.041681
11	dst_host_srv_serror_rate	0.026889

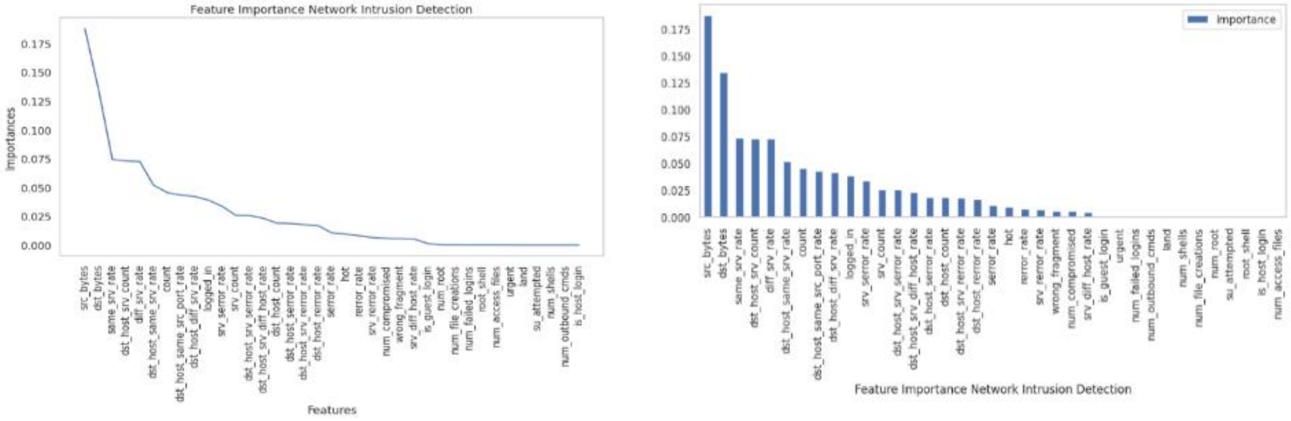

Figure 5: Intrusion Detection Systems (IDS) Feature Importance Dataset

FUZZY C-MEANS CLUSTERING METHOD. Clustering assigns data based on proximity to centroids. FCM refines datasets, creating a membership matrix for three clusters, enhancing IDS analysis. Figures 6 and 7 show results, IDS data points, feature values, and cluster labels in two Fuzzy C-Means clusters.

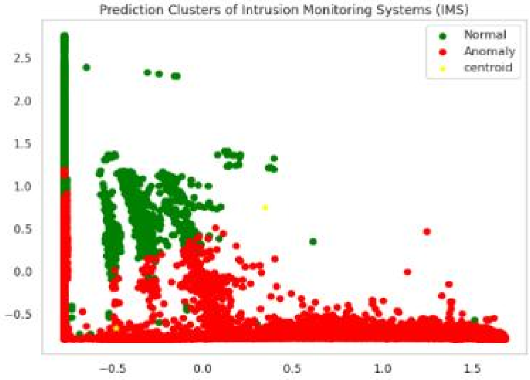

Figure 6: Three Clusters of Fuzzy C-Means (IDS)

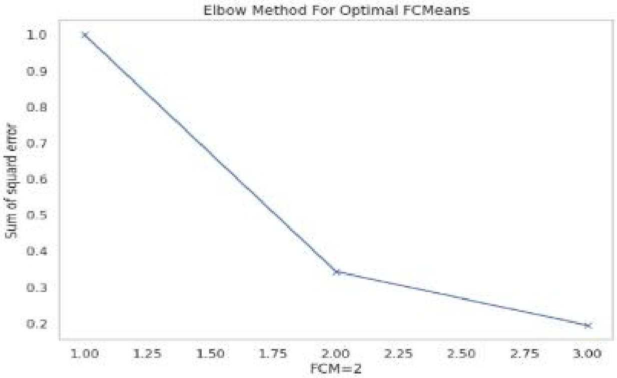

Figure 7: Average of Squared Fault Line Chart with Fuzzy C-Means

Clustering feature detection is evaluated using a variety of clustering techniques, with performance dignified by key metrics such as F-measure, recall, and precision. The system's deviation from predefined criteria is assessed, categorizing results as legitimate, suspicious, or unlawful. This approach classifies the data based on established standards, offering a comprehensive evaluation of the system's capacity to distinguish and identify different types of outcomes. By leveraging these metrics and classification techniques, the system's effectiveness in detecting and categorizing anomalies is thoroughly assessed.

3.2 CLASSIFICATION

Data classification involves using training data to categorize information into distinct groups or classes. The optimal classifier for a given dataset is determined through supervised learning techniques. Multiple tests are conducted to compare and assess the work performance of several categorization algorithms. This approach ensures the preference of the most effective classifier for the dataset, enabling accurate and reliable predictions. By evaluating different algorithms, the process helps identify the algorithms that provides the good results in terms of classification efficiency and accuracy [38, 39].

Stochastic Gradient Descent (SGD) Algorithm. SGD optimizes ML models efficiently, preventing local minima and reducing overhead. Figure 8 shows its confusion matrix, while the Receiver Operating Characteristic (ROC) curve enhances performance evaluation [40,41].

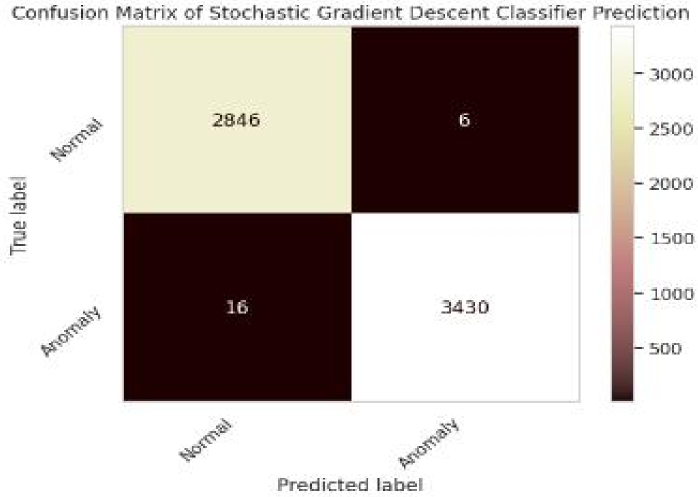

Figure 8: Confusion Matrix Stochastic Gradient Descent (SGD) Algorithm

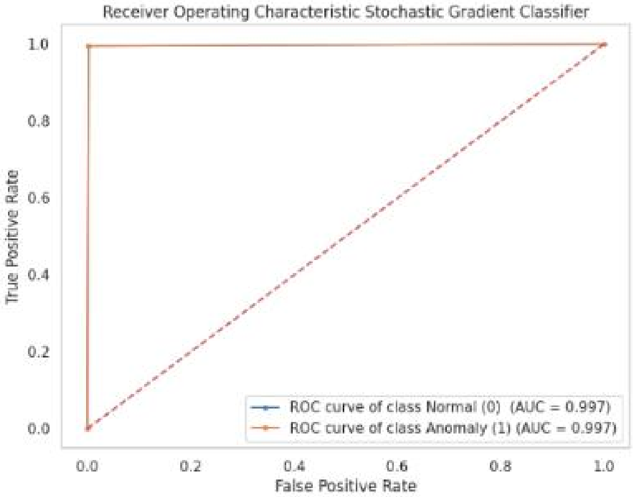

Figure 9: Shows ROC Curve for SGD

Logistic Regression (LR) Algorithm. Logistic Regression (LR) uses the Sigmoid function for classification, optimizing predictions iteratively. Figures 10 and 11 show its confusion matrix and ROC accuracy.

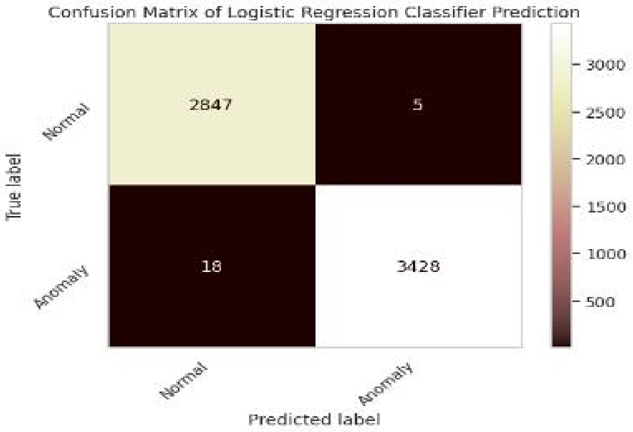

Figure 10: Confusion Matrix Algorithm for Logistic Regression (LR)

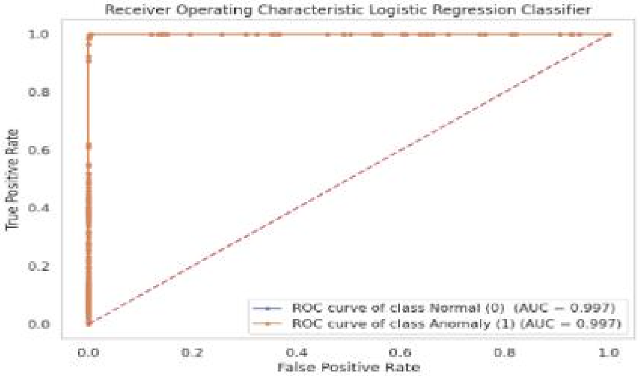

Figure 11: illustrates the Logistic Regression (LR) Receiver Operating Characteristic (ROC) Curve.

K-Nearest Neighbors (K-NN) Algorithm. K-NN categorizes data according to proximity, handling numerical and categorical data without spatial assumptions. It maps data in multidimensional space, predicting labels effectively. Figure 12 shows its confusion matrix.

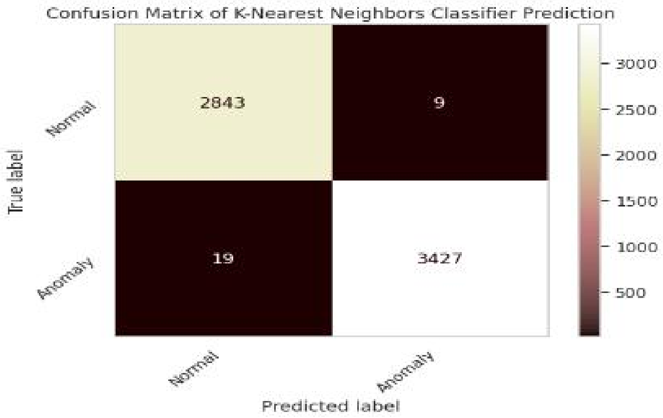

Figure 12: Confusion Matrix Algorithm for K-nearest-neighbors (K-NN)

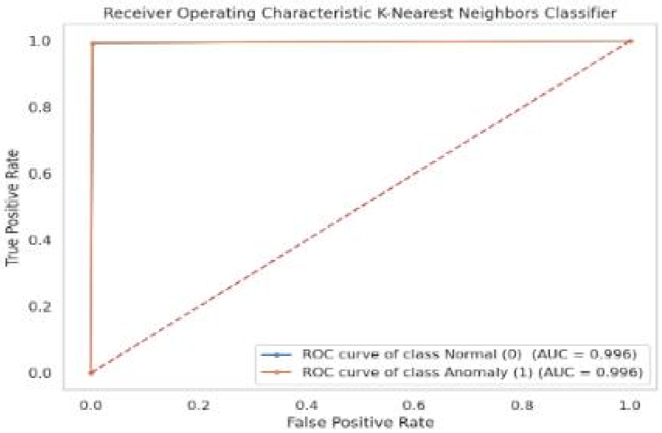

Figure 13: Shows ROC Curve for K-NN

The confusion matrix and ROC curve K-NN performance, shows prediction in (Figure 13).

Naive Bayes (NB) Algorithm. Naïve Bayes is a probabilistic classifier based on Bayes' Theorem, efficient for large datasets and text classification [34]. It assigns class labels by calculating probabilities, ensuring accurate predictions. Figure 14 illustrates the classification results.

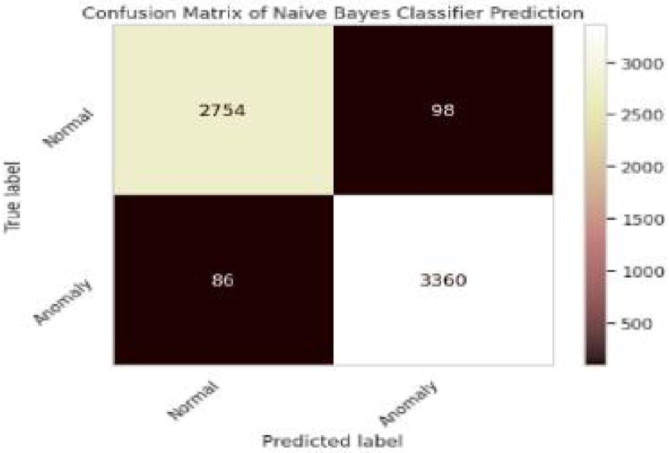

Figure 14: Confusion Matrix Algorithm for Naive Bayes

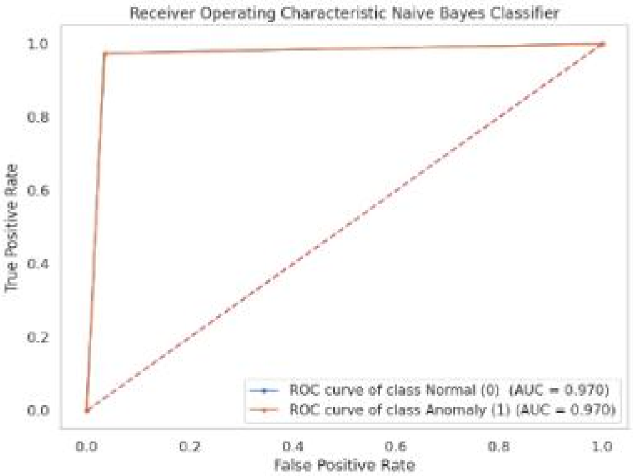

Figure 15: Shows ROC Curve for Naive Bayes

The Receiver Operating Characteristic (ROC) analysis optimizes classifier thresholds. Figure 15 shows NB precision in class prediction.

3. RESULTS AND DISCUSSION:

The study assesses hybrid algorithms for IDS prediction and classification. The table below displays the accuracy of the tested hybrid models.

Table 2: Accuracy of Hybrid Algorithms for Intrusion Detection Systems (IDS)

Hybrid Algorithm	Accuracy of Algorithms
Stochastic Gradient Descent (SGD) **Proposed Method**	99.6506 %
Logistic Regression (LR) **Proposed Method**	99.6348 %
K-Nearest Neighbours (K-NN) **Proposed Method**	99.5554 %
Naive Bayes (NB) **Proposed Method**	97.0784 %

The results indicate that hybrid algorithms exhibit varying effectiveness in predicting network intrusion attacks. Among the combinations tested, the fusion of Fuzzy-C-Means and Stochastic Gradient Descent (SGD) achieved the highest accuracy at an impressive 99.6506%. This outcome demonstrates the power of combining these algorithms, resulting in a highly accurate model for forecasting network intrusions. Following closely, the combination of Logistic Regression (LR) and Fuzzy-C-Means secured the second-highest accuracy at 99.6348%, while the pairing of Fuzzy-C-Means and K-Nearest Neighbors ranked third with an accuracy of 99.5554%. The Fuzzy-C-Means and Naive Bayes (NB) hybrid achieved the fourth-highest accuracy at 97.0784%. These findings emphasize the potential of hybrid algorithms in significantly enhancing the performance and efficiency of IDS, highlighting the importance of integrating multiple algorithms to improve predictive proficiency.

The performance of combined algorithms in forecasting cyberattacks is evaluated using several key parameters.

Table 3: IDS Parameter Score with Various Algorithm Combinations

S/No.	Parameter Scores	Fuzzy C-Means, SGD	Fuzzy C-Means, LR	Fuzzy C-Means, KNN	Fuzzy C-Means, NB
1	Precision	0.99457257	0.99613041	0.99537098	0.97068911
3	F1-Score	0.99487676	0.99631670	0.99551558	0.97051135
2	Recall	0.99520583	0.99651169	0.99566534	0.97034083
4	Sensitivity	1.0	1.0	1.0	1.0
5	Specificity	1.0	1.0	1.0	1.0

The performance of combined algorithms in forecasting cyberattacks is evaluated using several key parameters. When Naive Bayes (NB) and Fuzzy C-Means were paired, the model achieved a precision of 0.97068911, perfect specificity of 1.0, F1 score of 0.97051135, and a recall of 0.97034083 with sensitivity of 1.0. This demonstrates the balanced performance of this combination in detecting cyber threats. The pairing of Fuzzy C-Means and K-Nearest Neighbors outperformed others with a recall value of 0.99566534, F1 score of 0.99551558 and precision of 0.99566534. It also achieved a specificity of 1.0 and perfect sensitivity, showcasing its robust detection capabilities. Similarly, the combination of Logistic Regression (LR) and Fuzzy C-Means yielded a recall of 0.99651169, specificity of 1.0, F1 score of 0.99631670 and precision of 0.99613041. This indicates high reliability and accuracy in cyberattack prediction. Lastly, the Stochastic Gradient Descent (SGD) and Fuzzy C-Means hybrid model achieved a recall of 0.99520583, specificity of 1.0, precision of 0.99457257, and an F1 score of 0.99487676, further emphasizing the effectiveness of combining these algorithms. These parameter values highlight the powerful capabilities of hybrid models in improving the accuracy and reliability of cyberattack detection, reinforcing their potential for robust threat forecasting.

4. COMPARATIVE ANALYSIS

The comparative study in Table 4 shows the results of the suggested approach as well as the performance of different algorithms as documented in earlier research. While Hassan et al. [19] reported an accuracy of 89% with Logistic Regression (LR), Heidari et al. [18] obtained a 92% accuracy with the Random Forest approach. Alkasassbeh et al. [20] used the K-Nearest Neighbors (K-NN) algorithm and showed a 91% accuracy rate. On the other hand, the suggested approach significantly outperforms all tested algorithms. In particular, the Stochastic Gradient Descent (SGD) attained a remarkable accuracy of 99.6506%, closely followed by K-NN at 99.5554% and Logistic Regression at 99.6348%. In the suggested method, even Naive Bayes, which is usually regarded as a simpler model, achieved a high accuracy of 97.0784%. These results clearly indicate that the proposed method is highly efficient and accurate, significantly outperforming existing techniques in the literature.

Table 4: Algorithms Comparison Across Various Studies

Hybrid Algorithm	Accuracy of Algorithms
Heidari **et al. [18]**	Random Forest 92%
Hassan **et al. [19]**	Logistic Regression (LR) 89%
Alkasassbeh **et al. [20]**	K-Nearest Neighbors (K-NN) 91%
Stochastic Gradient Descent (SGD) **Proposed Method**	99.6506 %
Logistic Regression (LR) Proposed **Method**	99.6348 %
K-Nearest Neighbours (K-NN) **Proposed Method**	99.5554 %
Naive Bayes (NB) **Proposed Method**	97.0784 %

Figures 16 and 17 showcase the accuracy levels of various algorithm pairings within a hybrid framework for cyberattack detection. The pairings display a remarkable range of accuracy from 97.07% to 99.6506%, with the highest accuracy of 99.6506% achieved by combining Fuzzy C-Means with SGD. The combination of Logistic Regression (LR) and Fuzzy C-Means secured an accuracy of 99.6348%, while the Fuzzy C-Means and K-Nearest Neighbors (K-NN) combination reached 99.5554%. Lastly, the Naive Bayes (NB) and Fuzzy C-Means hybrid achieved an accuracy of 97.0784%. These figures clearly illustrate how hybrid algorithms significantly enhance the effectiveness of cyberattack detection systems, offering a powerful tool for improving accuracy in predictive models.

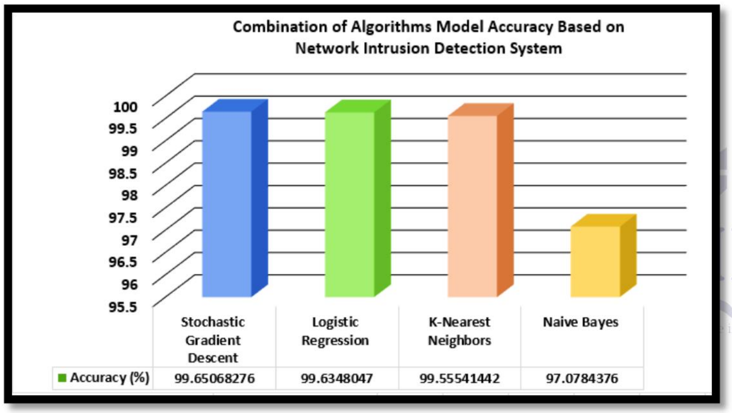

Figure 16: Combination of Algorithms Model Accuracy Based on Network IDS

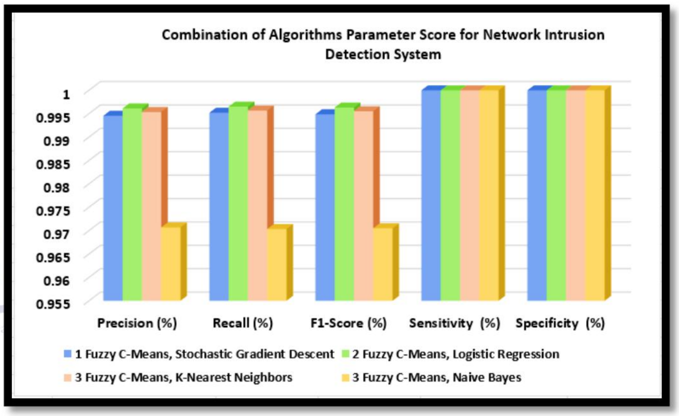

Figure 17: Combination of Algorithms Parameter Score for Network IDS

This study compared the accuracy of our Intrusion Detection Systems (IDS) with a hybrid model using the NIDS dataset against various accuracy levels from reference publications. The hybrid model achieved accuracy surpassing all these benchmarks. To completely comprehend each model's performance, additional examination of the experimental setup, feature selection techniques, and evaluation methodologies is required.

The efficiency of an IDS in detecting and minimizing network invasions is demonstrated by the study's experimental results. The IDS has the ability to greatly enhance security measures, as evidenced by its excellent performance across a number of evaluation criteria. The study does, however, admit its shortcomings in simulating real-world situations, inherent biases in evaluation processes, and difficulties with generalizability and scalability. To increase the precision and effectiveness of IDS, future studies should concentrate on refining ML techniques, such as reinforcement learning and deep learning. In order to produce creative ideas and solutions to growing cyberthreats, the study also highlights the significance of interdisciplinary collaboration among experts in network engineering, cybersecurity, and information science.

5. CONCLUSION:

This study develops a robust Intrusion Detection System (IDS) utilizing machine learning algorithms, with an emphasis on feature selection and ensemble learning. The experiments replicated real-world cyber-attacks, assessing the IDS's effectiveness in detecting various network intrusions. Models like Logistic Regression (LR), K-Nearest Neighbors (KNN), Stochastic Gradient Descent (SGD), and Naïve Bayes (NB) were explored, providing insights into their strengths and limitations. A hybrid model combining Fuzzy C-Means clustering with these algorithms achieved remarkable performance, with SGD and Fuzzy C-Means reaching 99.6506% accuracy. The study highlights the importance of integrating algorithms to enhance IDS reliability. Future studies should concentrate on improving feature selection, real-time detection, scalability, and adversarial robustness. Additionally, integrating IDS with security orchestration platforms can improve overall security. These advancements will help advance the field of intrusion detection, contributing to a safer digital landscape.

REFERENCES

1. S. Mahmood, S. M. Mohsin, and S. M. A. Akber, "Network security issues of data link layer: An overview," in 2020 3rd International Conference on Computing Mathematics and Engineering Technologies (iCoMET), Jan. 2020, pp. 1-6.
2. Y. Hamid, M. Sugumaran, and L. Journaux, "Machine learning techniques for intrusion detection: a comparative analysis," in Proceedings of the International Conference on Informatics and Analytics, Aug. 2016, pp. 1-6.
3. A. V. Turukmane and R. Devendiran, "M-MultiSVM: An efficient feature selection assisted network intrusion detection system using machine learning," 2024.
4. Moeez, M., Mahmood, R., Asif, H., Iqbal, M. W., Hamid, K., Ali, U., & Khan, N. (2024). Comprehensive Analysis of DevOps: Integration, Automation, Collaboration, and Continuous Delivery. Bulletin of Business and Economics (BBE), 13(1).
5. Parveen, F., Iqbal, S., Mumtaz, G., & Salahuddin, M. (2024). Real-time intrusion detection with deep learning: Analyzing the UNR intrusion detection dataset.
6. Afzal, M., Salahuddin, M., Hira, S., Sultan, M. F., Ahmad, S. Z., & Iqbal, M. W. (2024). A systematic literature review of understanding the human-computer interaction collaboration with user experience design.
7. Y. Lu, S. Chai, Y. Suo, F. Yao, and C. Zhang, "Intrusion detection for Industrial Internet of Things based on deep learning," Neurocomputing, vol. 564, p. 126886, 2024.
8. R. Devendiran and A. V. Turukmane, "Dugat-LSTM: Deep learning-based network intrusion detection system using chaotic optimization strategy," Expert Systems with Applications, vol. 245, p. 123027, 2024.
9. Muqaddas, M., Majeed, S., Hira, S., & Mumtaz, G. (2024). A Systematic Literature Review on Performance Evaluation of SQL and NoSQL Database Architectures. https://jcbi.org/index.php/Main/article/view/548/502
10. K. Psychogyios, A. Papadakis, S. Bourou, et al., "Deep Learning for Intrusion Detection Systems (IDSs) in Time Series Data," Future Internet, vol. 16, no. 3, p. 73, 2024.
11. N. O. Aljehane, H. A. Mengash, M. M. Eltahir, et al., "Golden jackal optimization algorithm with deep learning assisted intrusion detection system for network security," Alexandria Engineering Journal, vol. 86, pp. 415-424, 2024.
12. B. Sharma, L. Sharma, C. Lal, and S. Roy, "Explainable artificial intelligence for intrusion detection in IoT networks: A deep learning-based approach," 2024.
13. A. Almotairi, S. Atawneh, O. A. Khashan, and N. M. Khafajah, "Enhancing intrusion detection in IoT networks using machine learning-based feature selection and ensemble models," Systems Science & Control Engineering, vol. 12, no. 1, p. 2321381, 2024.
14. Z. Sun, G. An, Y. Yang, and Y. Liu, "Optimized machine learning enabled intrusion detection system for internet of medical things," Franklin Open, vol. 6, p. 100056, 2024.

15. S. Ali, Q. Li, and A. Yousafzai, "Blockchain and federated learning-based intrusion detection approaches for edgeenabled industrial IoT networks: A survey," 2024.

16. R. Kimanzi, P. Kimanga, D. Cherori, and P. K. Gikunda, "Deep Learning Algorithms Used in Intrusion Detection Systems—A Review," arXiv preprint arXiv:2402.17020, 2024.

17. C. EL Asry, I. Benchaji, S. Douzi, and B. EL Ouahidi, "A robust intrusion detection system based on a shallow learning model and feature extraction techniques," Plos One, vol. 19, no. 1, p. e0295801, 2024.

18. A. Heidari and M. A. Jabraeil Jamali, "Internet of Things intrusion detection systems: A comprehensive review and future directions," Cluster Computing, vol. 26, 2023.

19. H. A. Hassan, E. E. Hemdan, W. El-Shafai, M. Shokair, and F. E. A. El-Samie, "Intrusion Detection Systems for the Internet of Thing: A Survey Study," Wireless Personal Communications, vol. 128, no. 4, pp. 2753-2778, 2023.

20. M. Alkasassbeh and S. Al-Haj Baddar, "Intrusion detection systems: A state-of-the-art taxonomy and survey," Arabian Journal for Science and Engineering, vol. 48, no. 8, pp. 10021-10064, 2023.

21. M. Alkasassbeh and S. Al-Haj Baddar, "Intrusion detection systems: A state-of-the-art taxonomy and survey," Arabian Journal for Science and Engineering, vol. 48, no. 8, 2023.

22. Ramzan, M. S., Nasim, F., Ahmed, H. N., Farooq, U., & Khan, H. (2025). An innovative machine learning-based end-to-end data security framework in emerging cloud computing databases and integrated paradigms: Analysis on taxonomy, challenges, and opportunities. Spectrum of Engineering Sciences, 3(2), 90–125.

23. A. Alotaibi and M. A. Rassam, "Adversarial machine learning attacks against intrusion detection systems: A survey on strategies and defense," Future Internet, vol. 15,p. 62, 2023.

24. I. Debicha, R. Bauwens, T. Debatty, et al., "TAD: Transfer learning-based multi-adversarial detection of evasion attacks against network intrusion detection systems," Future Generation Computer Systems, vol. 138, pp. 185-197, 2023.

25. Network Intrusion Detection Dataset, Kaggle, https://www.kaggle.com/datasets/sampadab17/network-intrusion-detec.

26. A. H. Balla, M. H. Habaebi, E. A. Elsheikh, M. R. Islam, and F. M. Suliman, "The Effect of Dataset Imbalance on the Performance of SCADA Intrusion Detection Systems," Sensors, vol. 23, no. 2, p. 758, 2023.

27. Elzaridi, K., & Kurnaz, S. (n.d.). Integration between network intrusion detection and machine learning techniques to optimizing network security. Research Article. Information Technologies Department, Altinbas University, İstanbul, Turkey.

28. Data Clustering Algorithms, Fuzzy C-Means Clustering Algorithm, https://sites.google.com/site/dataclusteringalgorithms/fuzzy-c-means-clustering-algorithm

29. Stochastic Gradient Descent (SGD), Tutorials Point, https://www.tutorialspoint.com/scikit_learn/scikit_learn_stochastic_gradient_descent.htm

30. K-Nearest Neighbors (K-NN) Algorithm, IBM (1911), https://www.ibm.com/topics/knn

31. Naive-Bayes Algorithm, DataCamp, https://www.datacamp.com/tutorial/naive-bayes-scikit-learn

32. Uz Zaman, F., Khuhro, M.A., Kumar, K., Mirbahar, N., Khan, M.Z. and Kalhoro, A. (2021). Comparative Case Study Difference Between Azure Cloud SQL and Atlas MongoDB NoSQL Database.

33. Khan, M.Z., Khan, A.A., Laghari, A.A., Shaikh, Z.A., Khani, M.A.K., Morkovkin, D., Gavel, O., Shkodinsky, S., Makar, S. and Taburov, D. (2022). Comparative case study: an evaluation of performance computation between support vector machine, k-nearest neighbors, k-mean, and principal component analysis.

34. Mirza Azam Baig, Sarmad Ahmed Shaikh, Kamlesh Kumar Khatri, Muneer Ahmed Shaikh, Muhammad Zohaib Khan and Rauf, A. (2023). Prediction of Students Performance Level Using Integrated Approach of ML Algorithms. International Journal of Emerging Technologies in Learning (iJET), 18(01), pp.216–234.

35. Khan, M.Z., Shaikh, S.A., Shaikh, M.A., Khatri, K.K., Mahira Abdul Rauf, Kalhoro, A. and Muhammad Adnan (2023). The Performance Analysis of Machine Learning Algorithms for Credit Card Fraud Detection. International Journal of Online and Biomedical Engineering (iJOE), 19(03), pp.82–98.

36. Khan, M.Z., Uz Zaman, F., Adnan, M., Imroz, A., Abdul Rauf, M. and Phul, Z. (2023). Comparative Case Study: An Evaluation of Performance Computation Between SQL And NoSQL Database. [online] Sindh Journal of Headways in Software Engineering (SJHSE).

37. H. Farahbakhsh, Qureshi, N., Muhammad Zohaib Khan, Muhammad Ali Khan, Abdul Salam Soomro, Imroz, A. and Hussain Bux Marri (2023). Performance evolution for sentiment classification using machine learning algorithm. Journal of applied research in technology & engineering, 4(2), pp.97–110.

38. Siddiqui, M., Kalwar, H.A., Khan, M.Z., Khan, M.A., Imroz, A., Kalwar, M.A. and Marri, H.B. (2023). Performance Analysis for the Diagnosis of COVID-19 Prediction by Mathematical Modeling & Simulation.

39. Shaikh, S.Y., Qureshi, N.A., Khan, M.Z., Khan, M.A., Imroz, A. and Kalwar, M.A. (2023). Performance Analysis of Classification Algorithms for Software Defects Prediction by Mathematical Modelling & Simulations.

40. Zahra, W. U., Zaman, F. U., Mumtaz, G., Salahuddin, M., Khan, M. Z., Sultan, S. A., Hira, S., & Parveen, F. (2025). Approaches to predict cardiovascular issue using machine learning method. Spectrum of Engineering Sciences, 3(4), 417–429.

41. Mahmood, M. S. R., Mustafa, S., Asif, H., & Raza, A. (2025). Leveraging artificial intelligence to optimize software project management: Enhancing efficiency, risk mitigation, and decision-making. Contemporary Journal of Social Science Review, 3(2), 142–157.

BIBLIOGRAPHY:

Muqaddas Salahuddin received the B.Sc. degree in computer science and the M.Sc. degree in information technology from the Faculty of Computer Science and Information Technology, Superior University, Lahore, Punjab, Pakistan from 2019 to 2021. She is currently doing her MS emerging from the Faculty of Computer Science and Information Technology, Superior University, Lahore, Punjab, Pakistan from 2023 to 2025. She has attended various national & international conferences. She has also participated in many professional seminars, workshops, symposia and trainings. Her research interests include Data Science, Artificial Intelligence, Database Architectures, Machine Learning, Deep Learning, Cyber Security, Intrusion Detection and Internet of Things. She has authored and presented research papers at the national & international conferences and journals.

Fahim Uz Zaman holds a Master of Science in Cyber Security Management from the University of Law, Birmingham, United Kingdom, and a Master of Science in Computer Science from Sindh Madressatul Islam University, Karachi, Pakistan. He earned his Bachelor's degree in Computer Science from the same institution, graduating as a Gold Medalist. He has served in several roles, including DevOps Engineer at Premier Systems Ltd (Audi Pakistan), IoT Lead Trainer at the Presidential Initiative for Artificial Intelligence and Computing (PIAIC), and Computer Science Lecturer at Sindh Madressatul Islam University. He is currently working as an Advanced Technical Lecturer (Web and DevOps Engineering) at Newcastle College, United Kingdom. He is also a Microsoft Certified Trainer and has delivered numerous technical training sessions in DevOps, networking, and IoT. He has authored and published several research papers in reputable journals. His research interests include Serverless Computing, Cloud Infrastructure, DevOps Automation, Internet of Things (IoT), and Database Systems.

Gohar Mumtaz received the BS degree in Telecommunication Engineering from University of Engineering and Technology, Taxila, Pakistan, in 2013, the MS degree in Computer Science from University of Engineering and Technology, Lahore, Pakistan in 2018 and the PhD degree in Computer Science from The Superior University, Lahore, Pakistan in 2024. From 2014 to 2018, He worked in diverse kind of environment in the domain of Network provisioning, maintenance, and operations at well-known organizations like PTCL and Punjab Safe Cities Authority Lahore. Currently, he is working as Assistant Professor in the Faculty of CS and IT at "The Superior University" Lahore. His research areas include Wireless and Wired Networks, Network Administration and Security. Further, he focuses in 4G and 5G spectrum management and cyber-attacks prediction using machine learning.

Muhammad Zohaib Khan was born in Pakistan. He has received Master degree in Computer Science from Sindh Madressatul Islam University, Karachi, Pakistan and Bachelor degree in Computer Science from the University of Sindh, Jamshoro, Pakistan. He has worked as an IT Engineer in the Department of IT, Sindh Public Procurement Regulatory Authority from 2017 to 2019. He currently works as Software and Data Engineer, in the Department of IT, Shaheed Mohtarma Benazir Bhutto Institute of Trauma. He has authored and presented various research papers at the national & international conferences and journals. His research interests include Data Science, Artificial Intelligence, Machine Learning, Deep Learning, and the Internet of Things.

Maria Kainat was born in Pakistan. She has completed her graduation from the Department of Computer Science, Faculty of Sciences, University of Agriculture, Faisalabad, Punjab, Pakistan. She has recently completed her MS in Computer Science from the same department. She currently works as the CEO of MK-Professional in Faisalabad. She has attended various national & international conferences. She has also participated in many professional seminars, workshops, symposia and trainings. Her research interests include Data Science, Artificial Intelligence, Machine Learning, Deep Learning, Cyber Security, Internet of Things and Cloud Computing. She has authored and presented research papers at the national & international conferences and journals.

Sammia Hira is a research analyst from Lahore, Punjab, Pakistan. she has completed an MPhil in Information Technology at Superior University. she is AWS certified Solution Architect-Associate. she is doing job as a system engineer in Information Technology in Pakistan based datacenter moreover working on different research projects related Network and Cloud Computing Security, CISCO Networking etc. she has hands on experience in research in CISCO Networking, Cloud Computing (ESXi, vSphere, VCenter, Horizon Connection Server, IaaS service, PaaS service, and Serverless), Cyber Security.

Fakhra Parveen was born in Pakistan. She is currently doing her MS emerging from the Faculty of Computer Science and Information Technology, Superior University, Lahore, Punjab, Pakistan. Her research interests include Data Science, Artificial Intelligence, Database Architectures, Machine Learning, Deep Learning, Cyber Security and Internet of Things. She is currently a junior Lecturer with Superior University.

Rashid Mahmood received his MSc degree in Information Technology from the University of Education, Multan, Pakistan, in 2012. He is currently pursuing an MS in Software Project Management from Superior University, Lahore. Over the years, he has gained extensive experience in custom web development and technical support, working in diverse environments at reputable organizations such as The Tower Technologies, Lahore. Currently, he serves as a Senior Software Engineer at The Tower Technologies. His areas of expertise lie in software project management, with a strong focus on managing large-scale IT implementations, streamlining development processes, and leveraging emerging technologies—such as machine learning—for enhanced project planning and risk mitigation.