ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

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], Sammia Hira[5], Fakhra Parveen[6]

[1]Faculty of Computer Science and Information Technology, Superior University, Lahore, 54000, Pakistan
muqaddassalahuddin60@gmail.com

[2]Department: Digital Technologies, Newcastle College University Centre (NCUC) Rye Hill Campus, Scotswood Rd, Newcastle upon Tyne NE4 7SA, United Kingdom
faheemuzzaman@hotmail.com

[3]Faculty of Computer Science and Information Technology, Superior University, Lahore, 54000, Pakistan
gohar.m@superior.edu.pk

[4]Department of Information Technology, Shaheed Mohtarma Benazir Bhutto Institute of Trauma, Karachi, Pakistan zohaib_khan2017@yahoo.com

[5]Faculty of Computer Science and Information Technology, Superior University, Lahore, 54000, Pakistan
hirac53oort@gmail.com

[6]Faculty of Computer Science and Information Technology, Superior University, Lahore, 54000, Pakistan
parveenfakhra13@gmail.com

* Corresponding Author: Muqaddas Salahuddin (muqaddassalahuddin60@gmail.com)

Abstract:
In today's increasingly interconnected world, cybersecurity threats are more prevalent than ever, making robust intrusion detection systems (IDS) a critical necessity. This research introduces a hybrid IDS model that integrates Fuzzy C-Means clustering with classification methods such as Logistic Regression (LR), K-Nearest Neighbor (KNN), Stochastic Gradient Descent (SGD), and Naïve Bayes (NB). Advanced feature selection techniques are applied to improve detection accuracy and resilience against evolving cyberattacks. Extensive experimentation is used to validate the efficacy of this approach using the NIDS dataset. This study addresses limitations in traditional IDS methods, particularly their vulnerability to novel and complex attacks, and provides insights into leveraging machine learning (ML) to strengthen network security.

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

1. 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.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

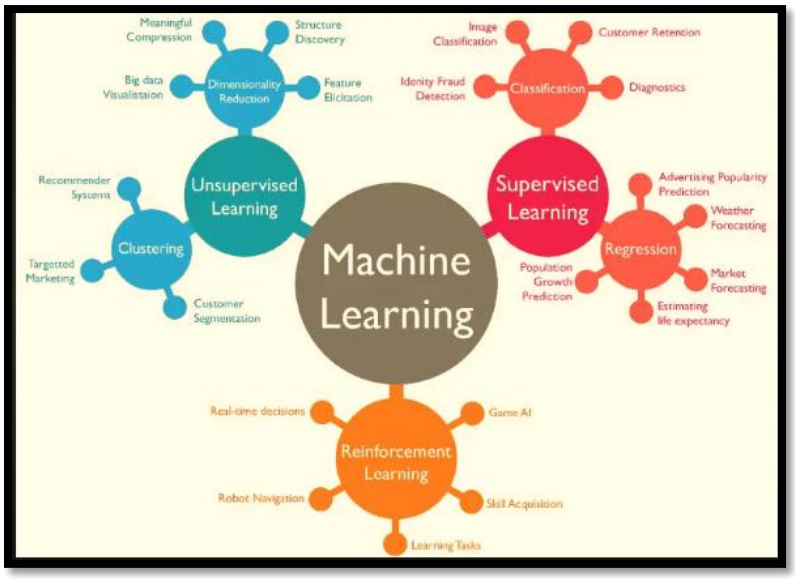


Figure 1: Structuring Machine-Learning Concepts

A key pillar of any organization's cybersecurity framework is intrusion detection, which identifies potential risks and provides solutions to mitigate them. Existing intrusion detection systems, which essentially rely on signature-based methods, are often slow to adapt to newer and more dynamic cyber threats [3]. In recent years, advanced computational methods have emerged as powerful solutions to improve the adaptability and effectiveness of IDS. The three paradigms of supervised learning, unsupervised learning, and reinforcement learning can be used to broadly classify these techniques. By using labeled data to train the model, supervised learning makes it possible to make precise predictions or judgments. While reinforcement learning uses feedback to learn from interactions in the environment. To enhance model performance, semi-supervised learning blends supervised and unsupervised components.

Recent advancements in ML have led to improvements in IDS, including enhanced feature selection and deep learning methods, which handle large datasets more effectively. ML methods, like method of neural network method and Support Vector Machines (SVM) [4]. The application of ML extends beyond simple data analysis, finding use in the security of Internet of Things (IoT) devices and even medical IoT systems [5]. Moreover, emerging technologies like blockchain, federated learning, and Explainable AI (XAI) are being integrated into IDS, offering enhanced transparency and interpretability for network security [6,7].

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

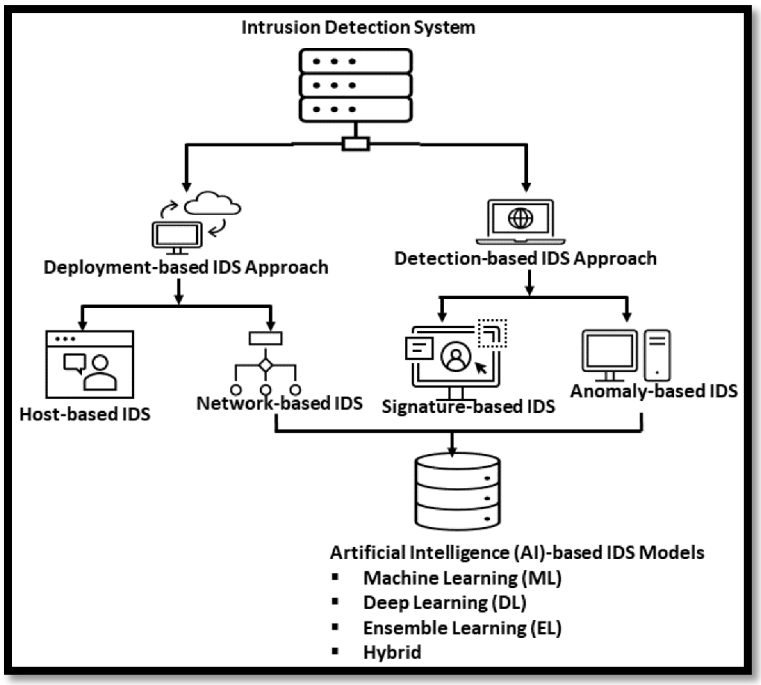


Figure 2: Intrusion Detection Using Explainable Artificial Intelligence (XAI)

This study introduces a hybrid IDS model that integrates Fuzzy C-Means clustering with machine learning classifiers such as Logistic Regression (LR), Naïve Bayes (NB), Stochastic Gradient Descent (SGD), and K-Nearest Neighbor (KNN). By leveraging these advanced techniques, the model aims to improve detection accuracy and scalability in reaction to changing internet risks. The NIDS dataset is used to validate this model. This research addresses key limitations in traditional IDS systems, including their vulnerability to sophisticated attacks, and contributes to the development of a more robust and scalable solution for modern network environments.

There are five sections to this work, beginning with the introduction. A description of the most relevant research was given in Section 2. Furthermore, section 3 outlined the methods and tools. The portion also included modeling and the results. The conversation has been fully completed and concluded in section 5.

2. 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].

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

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 [26]. 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 [27, 28]. 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 [30].

Table 1: Algorithms Comparison Across Various Studies

Authors	Algorithms	Results
Debicha et al. (2023)	SVM	91%
Abdulganiyu et al. (2023)	Support Vector Machine	94%
Hassan et al. (2023)	Gradient Boosting Machine	89%
Alkasassbeh and Al-Haj Baddar (2023)	K-Nearest Neighbors	91%
Azam et al. (2023)	Decision Tree	95%

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.8.No.2 2025

Alotaibi and Rassam (2023)	Neural Network	95%
Heidari and Jamali (2023)	Random Forest	92%
Azar et al. (2023)	Random Forest	93%

3. 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. The first stage in the procedure is to retrieve information from Google Drive.
2. A number of preprocessing procedures were used to the data form, such as data cleaning, Random Forest analysis of important attributes, and Standard Scaler standardization. Variable frequency and amplitudes were converted into standardized values via preprocessing normalization.
3. The standardization includes the process of converting varied frequencies and amplitudes into consistent scales so that the model receives consistent input.
4. Using a machine learning technique called Fuzzy C-Means clustering, the model's performance metrics accuracy, recall, precision, F1-score, sensitivity, and specificity were enhanced.
5. The dataset is divided into main two training and testing goal, with 25% being the testing datasets and 75% of data being the training datasets.
6. The final hybrid model combined Fuzzy C-Means clustering with classification algorithms, including (LR) K-Nearest Neighbor (KNN), Logistic Regression, Stochastic Gradient Descent (SGD) and Naïve Bayes (NB), to detect and classify network intrusions effectively.

Figure 2 provides an overview of the Intrusion Detection System Network Security architecture, summarizing the research project's workflow.

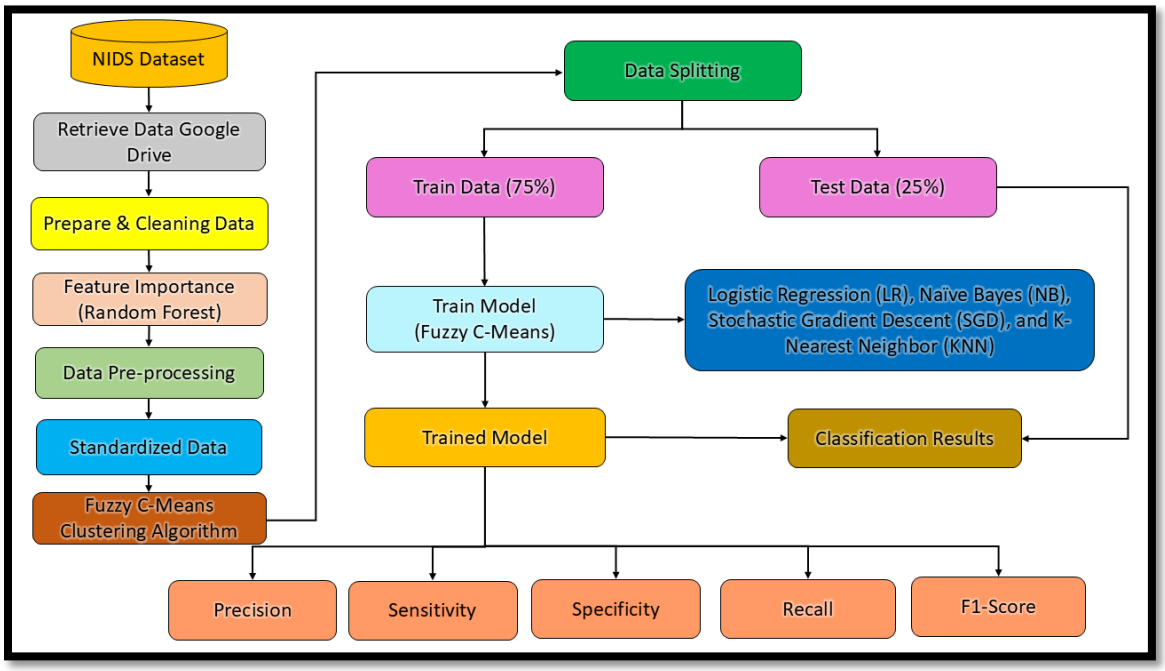


Figure 3: Intrusion Detection System (IDS) Network Security Architecture

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

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 [31].** 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 **[27].** Figure 3 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.

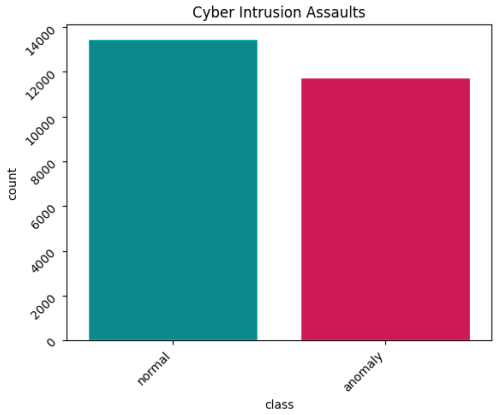


Figure 3: Facts of Cyber-Intrusion Assaults numbers

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. 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 **[32].**

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. This makes the hybrid Intrusion Detection System (IDS) model more efficient for predicting network intrusion attacks.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Table 2 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 4**, emphasizing their importance in identifying network security threats.

Table 2: 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
12	dst_host_srv_diff_host_rate	0.025724
13	srv_count	0.023402
14	dst_host_count	0.018482
15	srv_serror_rate	0.014813

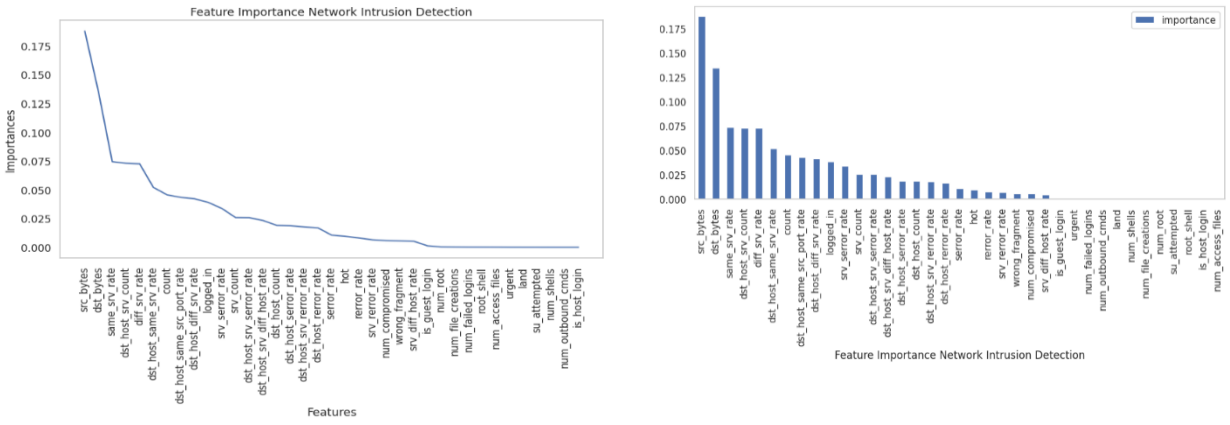


Figure 4: Intrusion Detection Systems (IDS) Feature Importance Dataset

The current study focuses on critical processes involved in predicting cyberattacks, including data normalization, preprocessing, simulation, and feature induction. These steps aim to address

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

challenges such as data complexity, post-processing efficiency, and system performance. The preprocessing phase begins with data collection from the Intrusion Detection Systems (IDS) dataset, followed by standardization and normalization to ensure uniformity and enhance model accuracy. The final preprocessed and cleaned dataset is presented in Table 3. Additionally, sensitive data patterns are analyzed using Fuzzy C-Means clustering, graphically depicted with light-yellow circles representing the Y values and black circles representing the X values.

The processed dataset used to predict network intrusion attacks, which has been standardized for analysis, is shown below in table 3.

Table 3: Processed Intrusion Detection Systems (IDS) Dataset for Network Intrusion Prediction

27-Dimension
[[2.34244932 -0.76718068 -0.08136747 ... -0.13341384 -0.14295606 -0.36060351]
[0.8914446 -0.76718068 -0.08136747 ... -0.1585304 -0.14295606 -0.36060351]
[-0.67826044 -0.76718068 -0.08136747 ... -0.1585304 -0.14295606 0.25073038]
... [-0.67826044 -0.76718068 -0.08136747 ... 0.77640905 0.665054 0.26888741]
[-0.67826044 -0.76718068 -0.08136747 ... -0.1585304 -0.14295606 0.24505025]
[-0.67826044 -0.76718068 -0.08136747 ... -0.1585304 -0.14295606 0.3014431]]

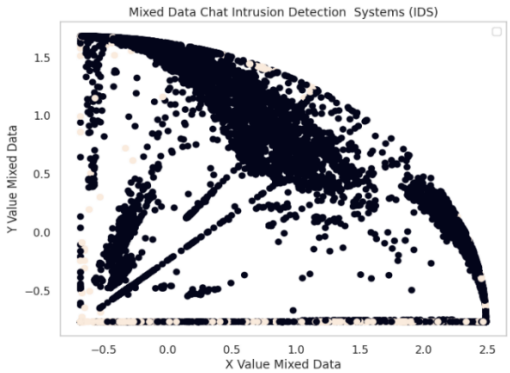


Figure 5: Mixed information Chat Intrusion Detection Systems (IDS)

By employing feature selection techniques and integrating advanced clustering methods like Fuzzy C-Means, the study enhances the system's efficiency in handling complex intrusion detection tasks, ultimately improving its predictive capabilities and performance.

FUZZY C-MEANS CLUSTERING METHOD. Clustering is a prominent form of unsupervised learning widely used in various fields. It involves organizing and analyzing data. Each data point in the FCM technique is given a membership value determined by how far away it is from its cluster's centroid. A data point's membership score increases with its proximity to the centroid. This approach ensures that each data point has a degree of association with the cluster, rather than a strict, single membership **[33, 34].**

Preprocessing techniques are employed to transform raw datasets into more structured forms suitable for clustering. The FCM algorithm plays a crucial role in processing and refining large

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

datasets by removing redundancies. It assigns every data-point a degree of membership, indicating its likelihood of association with a given cluster. This method partitions the data into three distinct clusters, generating a membership matrix that reflects the relationship between each data point and the corresponding cluster. The centroids, or core points of each cluster, reflect the proximity of data points to the centroid, with closer data points having stronger associations.

FCM is particularly effective for datasets with multiple classes, because it allows for the flexible assignment of data points to several clusters with different levels of membership. The IDS dataset used here includes eleven dimensions, with a target attribute that reveals cluster counts and details related to cyberattacks. Figures 6 and 7 illustrate the clustering results through line graphs and cluster distributions, showcasing how the transformation of raw data into a more organized structure enhances interpretability.

Table 4 presents the centroid values for the Fuzzy C-Means classification, showing the central points for each cluster in the IDS dataset. These centroids represent the core attributes of the clusters, indicating the most significant features for each group.

Table 4: Value of the Fuzzy C-Means Classification Centroid

```
array ([[-0.48295092, -0.66646946, -0.01905449, -0.03122024, -0.25182671, -0.01150496, -0.02304871, 0.80522006, 0.00140495, 0.33665781, 0.3365827, 0.00786238, 0.00803129, -0.08707053, 0.11269005, -0.0512385, 0.75305779, -0.31392234, -0.09624955, 0.00870686, -0.05747827, -0.03511175, 0.35142841, 0.35846006, 0.01227972, 0.01073647, 0.19691952],
       [ 0.34724485, 0.74397063, -0.02513277, -0.01975318, 0.23999721, -0.00613617, -0.01817923, -0.64995701, -0.15154976, -0.25534534, -0.25452785, -0.06194141, -0.0621398, -0.02237367, -0.1655957, -0.0153474, -0.69656104, 0.14524786, 0.00186091, -0.05276538, -0.03873757, -0.02298994, -0.26374769, -0.26659601, -0.07422022, -0.06742788, -0.2274263]])
```

Table 5 displays the data points within the two clusters generated by Fuzzy C-Means for the IDS dataset. Each row corresponds to a data point, showing the feature values and the assigned cluster label, with values indicating the clustering structure and the nature of each data point.

Table 5: Two Clusters of Fuzzy C-Means Intrusion Detection Systems (IDS) Pre-processed Predictive Dataset

```
[[ 2.34244932 -0.76718068 -0.08136747 ... -0.13341384 -0.14295606 -0.36060351, 0]
 [ 0.8914446 -0.76718068 -0.08136747 ... -0.1585304 -0.14295606 -0.36060351, 0]
 [-0.67826044 -0.76718068 -0.08136747 ... -0.1585304 -0.14295606 0.25073038, 1]
 ... [-0.67826044 -0.76718068 -0.08136747 ... 0.77640905 0.665054 0.26888741, 0]
 [-0.67826044 -0.76718068 -0.08136747 ... -0.1585304 -0.14295606 0.24505025, 0]
```

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL


```
[-0.67826044 -0.76718068 -0.08136747 ... -0.1585304 -
0.14295606 0.3014431, 0]]
```

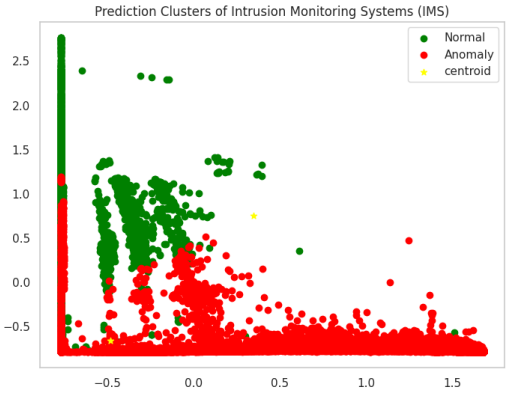


Figure 6: Fuzzy C-Means Three Clusters (IDS)

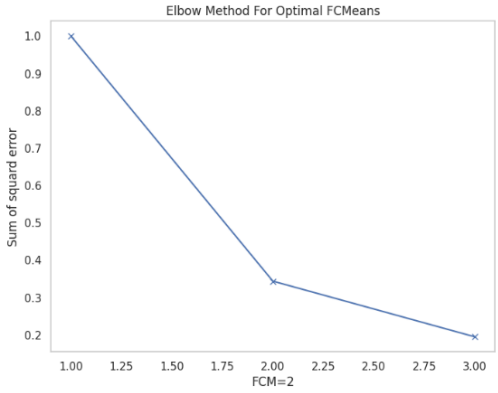


Figure 7: Fuzzy C-Means Sum of Squared Error Line Chart

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.

Stochastic Gradient Descent (SGD) Algorithm. The SGD approach is a highly effective technique for optimizing machine learning models, particularly for large-scale data. It adjusts model parameters by analyzing each individual data point rather than processing the complete dataset, making it computationally efficient. This approach facilitates quicker convergence and reduces computational overhead compared to batch processing methods. Furthermore, SGD introduces a degree of randomness in its updates, which helps to avoid local minima and enhances optimization for complex, non-convex functions. These attributes make SGD a key component in achieving the high accuracy demonstrated by the hybrid IDS framework **[35].**

This confusion matrix is a crucial instrument for assessing classification models since it offers a succinct overview of the model's functionality. To help find possible misclassifications and opportunities for improvement, it separates the predictions as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.8.No.2 2025

Performance Evaluation Metrics:

The categorization system's performance is assessed using the following metrics:

1. **True Positive Rate (TPR):**
   TPR indicates the percentage of successfully identified attacks among all actual attacks. It is calculated as:

$$\textbf{True Positive Rate (TPR)} = \frac{\textbf{TP}}{(\textbf{TP} + \textbf{FN})} \quad \textbf{(1)}$$

2. **False Positive Rate (FPR):**
   FPR calculates the percentage of legitimate cases incorrectly classified as attacks, relative to all legitimate cases. It is computed as:

$$\textbf{False Positive Rate (FPR)} = \frac{\textbf{FP}}{(\textbf{FP} + \textbf{TN})} \quad \textbf{(2)}$$

3. **False Negative Rate (FNR):**
   FNR represents the percentage of actual attacks missed and incorrectly classified as normal. It is determined as:

$$\textbf{False Negative Rate (FNR)} = \frac{\textbf{FN}}{(\textbf{FN} + \textbf{TP})} \quad \textbf{(3)}$$

A confusion matrix for the SGD model's predictions is shown in **Figure 8**. This method is used to forecast categories based on both previously gathered and newly obtained data after training, with the aim of maximizing alignment with the actual labels.

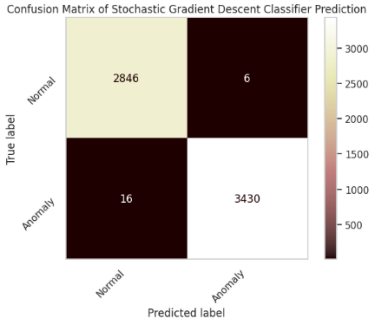


Figure 8: Confusion Matrix Stochastic Gradient Descent (SGD) Algorithm

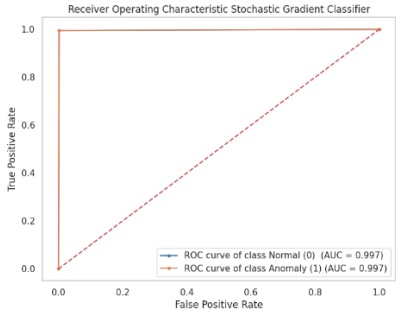


Figure 9: Illustrates the Receiver Operating Characteristic (ROC) Curve for Stochastic Gradient Descent (SGD)

The ROC curve is used to evaluate a classifier's work performance as the classification threshold changes, offering insights into the cost-benefit evaluation and helping to understand the effectiveness of the SGD model.

Logistic Regression (LR) Algorithm. A machine learning-based classification method called logistic regression (LR) uses a probabilistic hypothesis for predictive analysis. The 'Sigmoid function' or' regression model' is a more complicated differential equation used in LR as opposed to Linear Regression, which performs a straightforward linear transformation. In order to provide meaningful interpretations, the LR hypothesis limits the coefficients to the binary values of 0 and 1. Although linear approaches make an effort to approximate these values, the LR hypothesis cannot be supported by them **[36, 37].**

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

These procedures allow the LR algorithm to iteratively modify parameters in order to maximize predictions and reduce expenses. With the aim of modifying predictions to correspond with the labels, LR labels the values of previous data and forecasts the value of fresh data. The outcome of this LR matrix for confusion is shown in Figure 10.

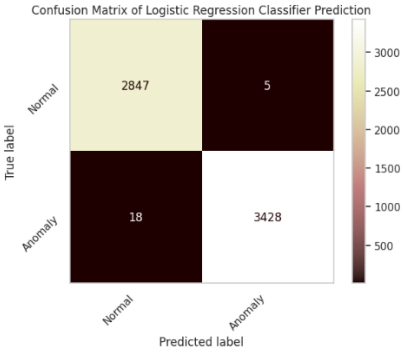


Figure 10: Confusion Matrix Algorithm for Logistic Regression (LR)

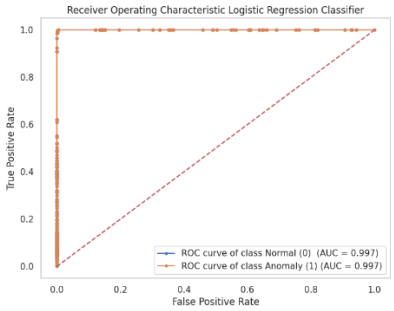
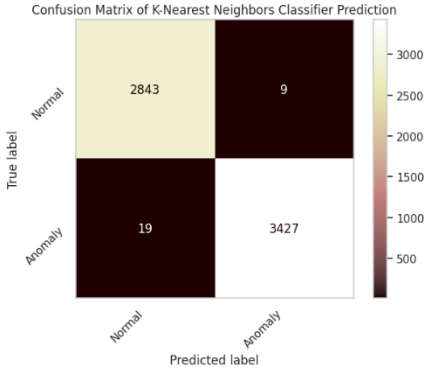


Figure 11: illustrates the Logistic Regression (LR) Receiver Operating Characteristic (ROC) Curve.

The confusion matrix is essential for identifying the expected labels for detection & prediction tasks, provided that the perceptron model—particularly logistic regression—is appropriate for the given situation. The results of using the LR model to a dataset that was generated are illustrated in Figure 11. The LR performance is evaluated with using this Receiver Operating Characteristic curve. The ROC curve in this study aids in evaluating model's predictions accuracy, offering insightful information about prediction trends and enhancing the estimating methodology's overall accuracy.

K-Nearest Neighbors (K-NN) Algorithm. An informal supervised learning classifier, the K-Nearest Neighbors (K-NN) method employs closeness to categorize or forecast how a one data point will be arranged. K-NN is a popular and adaptable machine learning method that is well-known for being straightforward and simple to use. The approach works with a variety of datasets, covering both numerical and categorical data, because it doesn't make any conclusions about the spatial distribution of the data that is underneath. K-NN is less susceptible to outliers than other algorithms since it makes predictions based on how similar the data points within the dataset are **[38, 39].**

Data is precisely mapped onto space with multiple dimensions throughout this process. With the aim of producing predictions that closely match actual labels, the algorithm may predict labels for both new and current data samples after it have been trained. The resulting confusion matrix, which displays the K-NN algorithm's prediction performance, is shown in **Figure 12**.



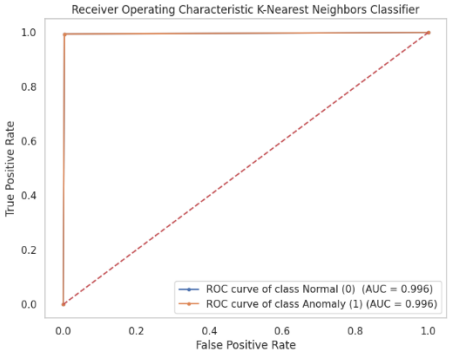


Figure 13: Illustrates the Receiver Operating

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

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

Characteristic (ROC) Curve K-nearest-neighbors (K-NN)

The confusion matrix is essential in identifying the expected labels for our identification and prediction tasks, assuming that the K-NN paradigm is appropriate for our specific situation. The results of applying the K-NN method to a synthetic dataset are illustrated in Figure 13. The K-NN model's performance is shown by the ROC curve. The Receiver Operating Characteristic curve is a good approach for this study to evaluate the quality of our model predictions, which helped us improve the overall precision and refinement of our estimating process while also offering insightful information on prediction trends.

Naive Bayes (NB) Algorithm. Using the theorem of Bayes to predict class membership, the Naive Bayes (NB) method is a popular supervised learning technique for classification tasks. When working with big training datasets, it is very helpful in text categorization challenges. Naive Bayes, which is well-known for its probabilistic methodology, offers a productive method for creating machine learning algorithms that can generate predictions quickly. By applying the ideas of Bayes' Theorem, this method improves the efficiency of different algorithms and increases the accuracy of prediction models **[40, 41].**

These procedures allow the Naïve Bayes classifier to effectively classify data based on the highest probability by calculating the probabilities associated with various class labels **[42, 43].** We labeled past data values using the Naive Bayes method and made predictions about the data values based on those labels. In order to make sure that our forecasts fell into the right categories during the preparation stage, we employed the NB approach. The outcome of this operation is depicted in the resulting matrix, which is displayed in Figure 14.

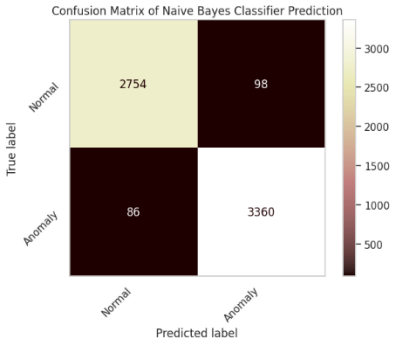


Figure 14: Confusion Matrix Naive Bayes Algorithm

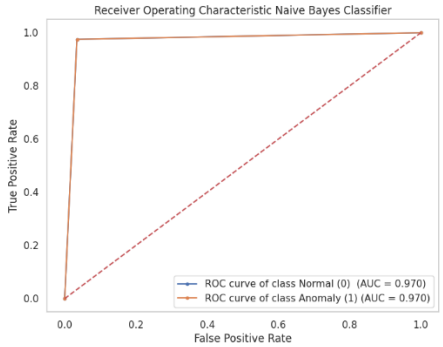


Figure 15: Illustrates the Receiver Operating Characteristic, Curve Naive Bayes

In this study the ROC analysis is using, because it is an essential method for evaluating classifier performance while fine-tuning the discrimination threshold. This analysis is critical for cost-benefit evaluation, aiding in making informed decisions. **Figure 15** provides a graphical representation of the ROC curve, offering a visual interpretation of the classifier's performance and helping to understand the precision of the Naive Bayes algorithm in predicting the class labels.

4. RESULTS AND DISCUSSION:

The research evaluates various hybrid algorithms to determine their effectiveness in predicting and classifying Intrusion Detection Systems (IDS). The following table presents the accuracies of the hybrid algorithms tested.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Table 6: 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 7: Parameter Scores for Intrusion Detection Systems (IDS) using Different Algorithm Combinations

S/No.	Parameter Score	Fuzzy C-Means, Stochastic Gradient Descent (SGD)	Fuzzy C-Means, Logistic Regression (LR)	Fuzzy C-Means, K-Nearest Neighbours (K-NN)	Fuzzy C-Means, Naive Bayes (NB)
1	Precision	0.99457257	0.99613041	0.99537098	0.97068911
2	Recall	0.99520583	0.99651169	0.99566534	0.97034083
3	F1-Score	0.99487676	0.99631670	0.99551558	0.97051135
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

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

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.

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.

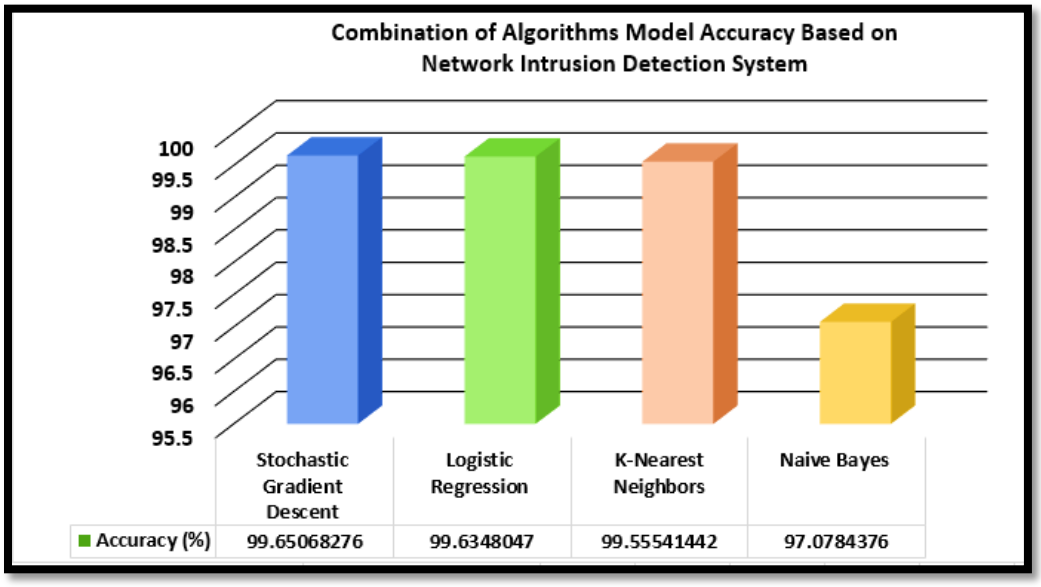


Figure 16: Combination of Algorithms Model Accuracy Based on Network Intrusion Detection System (IDS)

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

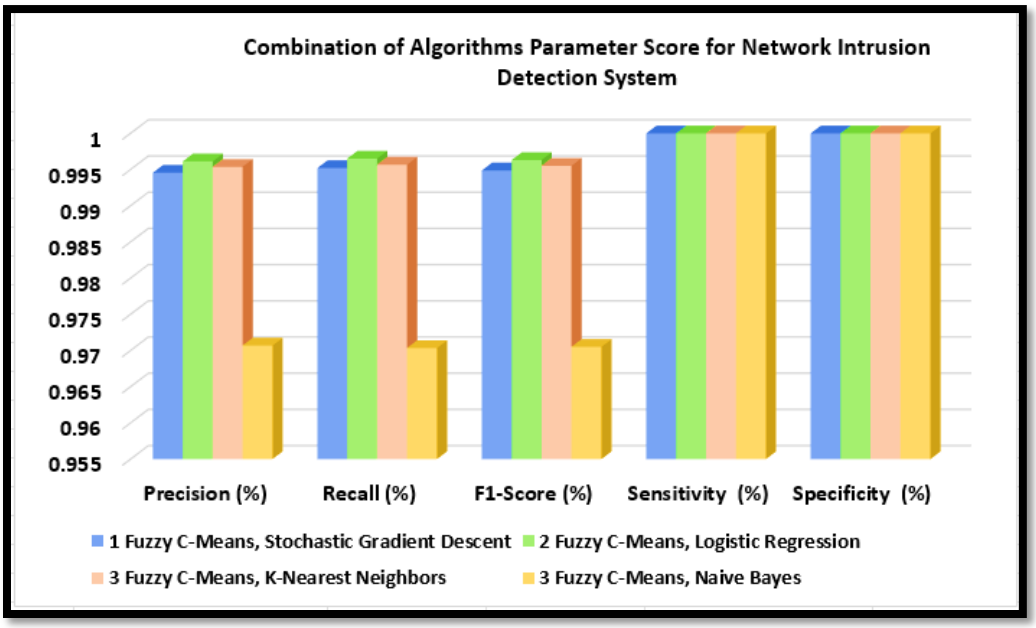


Figure 17: Combination of Algorithms Parameter Score for Network Intrusion Detection System (IDS)

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

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

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. A. Kumar and M. Sinha, "Overview on vehicular ad hoc network and its security issues," in 2014 International Conference on Computing for Sustainable Global Development (INDIACom), Mar. 2014, pp. 792-797.
3. I. H. Sarker, Y. B. Abushark, F. Alsolami, and A. I. Khan, "Intrudtree: a machine learning based cyber security intrusion detection model," Symmetry, vol. 12, no. 5, p. 754, 2020.
4. 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.
5. H. Liu and B. Lang, "Machine learning and deep learning methods for intrusion detection systems: A survey," Applied Sciences, vol. 9, no. 20, p. 4396, 2019.
6. A. V. Turukmane and R. Devendiran, "M-MultiSVM: An efficient feature selection assisted network intrusion detection system using machine learning," Computers & Security, vol. 137, p. 103587, 2024.
7. 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).
8. Parveen, F., Iqbal, S., Mumtaz, G., & Salahuddin, M. (2024). Real-time intrusion detection with deep learning: Analyzing the UNR intrusion detection dataset. Journal of Computing & Biomedical Informatics, 7(02). https://jcbi.org/index.php/Main/article/view/554
9. 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. Bulletin of Business and Economics, 13(2), 723–729. https://doi.org/10.61506/01.00386.
10. 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.
11. 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.
12. Muqaddas, M., Majeed, S., Hira, S., & Mumtaz, G. (2024). A Systematic Literature Review on Performance Evaluation of SQL and NoSQL Database Architectures. Journal of Compu-ting & Biomedical Informatics, 7(02). https://jcbi.org/index.php/Main/article/view/548/502
13. 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.
14. 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.
15. B. Sharma, L. Sharma, C. Lal, and S. Roy, "Explainable artificial intelligence for intrusion detection in IoT networks: A deep learning-based approach," Expert Systems with Applications, vol. 238, p. 121751, 2024.
16. 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.
17. 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.
18. S. Ali, Q. Li, and A. Yousafzai, "Blockchain and federated learning-based intrusion detection approaches for edgeenabled industrial IoT networks: A survey," Ad Hoc Networks, vol. 152, p. 103320, 2024.
19. 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.
20. 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.
21. A. Heidari and M. A. Jabraeil Jamali, "Internet of Things intrusion detection systems: A comprehensive review and future directions," Cluster Computing, vol. 26, no. 6, pp. 3753-3780, 2023.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

22. O. H. Abdulganiyu, T. Ait Tchakoucht, and Y. K. Saheed, "A systematic literature review for network intrusion detection system (IDS)," International Journal of Information Security, pp. 1-38, 2023.
23. Z. Azam, M. M. Islam, and M. N. Huda, "Comparative analysis of intrusion detection systems and machine learning based model analysis through decision tree," IEEE Access, 2023.
24. 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.
25. 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.
26. 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. https://sesjournal.com/index.php/1/article/view/106
27. A. Alotaibi and M. A. Rassam, "Adversarial machine learning attacks against intrusion detection systems: A survey on strategies and defense," Future Internet, vol. 15, no. 2, p. 62, 2023.
28. 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.
29. A. T. Azar, E. Shehab, A. M. Mattar, I. A. Hameed, and S. A. Elsaid, "Deep learning-based hybrid intrusion detection systems to protect satellite networks," Journal of Network and Systems Management, vol. 31, no. 4, p. 82, 2023.
30. Network Intrusion Detection Dataset, Kaggle, https://www.kaggle.com/datasets/sampadab17/network-intrusion-detec.
31. 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.
32. S. Ullah, J. Ahmad, M. A. Khan, et al., "TNN-IDS: Transformer neural network-based intrusion detection system for MQTT-enabled IoT Networks," Computer Networks, vol. 237, p. 110072, 2023.
33. C. Hazman, A. Guezzaz, S. Benkirane, and M. Azrour, "Toward an intrusion detection model for IoT-based smart environments," Multimedia Tools and Applications, pp. 1-2, 2023.
34. 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, and Computer Engineering Department, Altinbas University, İstanbul, Turkey.
35. Feature Importance, Built-In (2011), https://builtin.com/data-science/feature-importance
36. 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. doi: https://doi.org/10.3991/ijoe.v19i03.35331.
37. Data Clustering Algorithms, Fuzzy C-Means Clustering Algorithm, https://sites.google.com/site/dataclusteringalgorithms/fuzzy-c-means-clustering-algorithm
38. Stochastic Gradient Descent (SGD), Tutorials Point, https://www.tutorialspoint.com/scikit_learn/scikit_learn_stochastic_gradient_descent.htm
39. 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. doi: https://doi.org/10.3991/ijet.v18i01.35339
40. Towards Data Science, Logistic-Regression, https://towardsdatascience.com/introduction-to-logistic-regression-66248243c148
41. 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. [online] Journal of Tianjin

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

University Science and Technology. Available at: https://tianjindaxuexuebao.com/details.php?id=DOI:10.17605/OSF.IO/HK3SF [Accessed Apr. 2022].
42. K-Nearest Neighbors (K-NN) Algorithm, IBM (1911), https://www.ibm.com/topics/knn
43. 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. [online] International Journal of Artificial Intelligence & Mathematical Sciences (IJAIMS). Available at: http://ijaims.smiu.edu.pk/ijaims/index.php/AIMS/article/view/47 [Accessed 1 Jul. 2023].
44. Naive-Bayes Algorithm, DataCamp, https://www.datacamp.com/tutorial/naive-bayes-scikit-learn

Authors:

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 B.S. degree in telecommunication engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2013, and the M.S. degree in computer science from the University of Engineering and Technology, Lahore, Pakistan, in 2018. He is completed the Ph.D. degree in computer science with Superior University, Lahore. He worked in diverse kind of environment in the domain of network provisioning, maintenance, and operations at well-known organizations, such as, PTCL and Punjab Safe Cities Authority, Lahore. He is currently a Senior Lecturer with Superior University. His research interests include wireless and wired networks and 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.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

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.