Kasu Journal of Computer Science Vol. 1 No. 4 [December, 2024], pp. 689-701
https://doi.org//10.47514/kjcs/2024.1.4.001

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

An Enhanced Network Intrusion Detection System using Dimensionality Reduction Techniques

Olufemi James Akinola[1], Sulaiman Olaniyi Abdulsalam[2], Michael Favour Edafeajiroke[3], and Kazeem Alagbe Gbolagade[4]

[1]Department of Computer Science, Augustine University, Ilara-Epe, Lagos State, Nigeria.

[2, 4]Department of Computer Science, Faculty of Information and Communication Technology, Kwara State University, Malete, Nigeria.

[3]Department of Computer Science, Faculty of Computing, University of Port Harcourt, Port Harcourt, River State, Nigeria.

olufemi.akinola@augustineuniversity.edu.ng[1], sulaiman.abdulsalam@kwasu.edu.ng[2],

michael.edafeajiroke@uniport.edu.ng[3], kazeem.gbolagade@kwasu.edu.ng[4].

Abstract

Protecting the privacy and confidentiality of data in computer networks is crucial, necessitating reliable intrusion detection methods. The complexity and vast volume of data present challenges to effective intrusion detection, as traditional models struggle with high-dimensional data that reduce accuracy. This study addresses these challenges by integrating Mutual Information (MI) and Linear Discriminant Analysis (LDA) as feature selection and extraction techniques respectively, to enhance the classification performance of Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) classifiers. The Canadian Institute for Cybersecurity Intrusion Detection System 2017 (CIC-IDS 2017) dataset was used to develop an Intrusion Detection System (IDS). The Results findings revealed that integrating MI+LDA with SVM yielded an accuracy of 98.84% while combining MI+LDA with KNN produced an accuracy of 90.17%. The result of the MI+LDA+SVM model based on accuracy, sensitivity, specificity, and F1-score matrixes, outperformed that of MI+LDA with KNN and other existing models. Therefore, the incorporation of MI and LDA dimensionality reduction techniques proved reliable for boosting the performance of an intrusion detection system in computer networks. Future research could apply the result of this study to other types of computer network benchmark datasets and other kinds of machine learning techniques.

Keywords: Intrusion Detection System, Dimensionality Reduction Techniques, Machine Learning, Canadian Institute for Cybersecurity Intrusion Detection System 2017, Classification Performance.

1. Introduction

Computer networks have become ubiquitous in modern society and are used for various purposes, such as communication, data storage, and information exchange (Abdulsalam et al., 2024). However, the increasing popularity of computer networks has also led to increased security risks, such as cyber-attacks, data breaches, and unauthorized access (Rincy and Gupta, 2021). Interconnected devices that share information and resources through the Internet are called Computer networks (Ahmad and Magaji, 2024; Zeyuan, 2022). This computer grows at an exponential rate and this has led to rise in cyber-attacks, posing a significant risk to personal and corporate data stored on network devices (Albulayhi et al., 2022). Current techniques made to enhance computer network security remain inadequate in protecting against these attacks (Yang et al., 2022). Although Intrusion Detection Systems (IDSs) have been developed to address these security issues however traditional IDSs are computationally expensive especially when applied to a rapidly growing computer network (Yang et al., 2022; Zhang et al., 2017). IDS monitors and analyzes data from various sources to detect security breaches and intrusions (Choudhary and Kesswani, 2021). Unlike firewalls that prevent network access, IDS detect intrusions

Kasu Journal of Computer Science Vol. 1 No. 4 [December, 2024], pp. 689-701
https://doi.org//10.47514/kjcs/2024.1.4.001
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

when they occur, alert them, and potentially stop the connection. Network-based IDS (NBIDS) and Host-based IDS (HIDS) are the two main types of IDS (Zhang et al., 2017).

NBIDS monitors network packets on a subnet, while HIDS uses a host-based agent to examine system calls, file system updates, and logs. However, IDSs often generate high-dimensional data, which can lead to computational challenges and reduce the accuracy of intrusion detection systems (Yang et al., 2022). data obtained from network-based IDSs typically includes various features, such as source IP address, destination IP address, source port, destination port, protocol, and the packet payload. The high dimensionality of this data can result in overfitting, high computational time, and reduced accuracy of intrusion detection systems (Yang et al., 2022).

Machine Learning (ML) models have been widely employed in intrusion detection systems (IDS) to enhance accuracy and effectiveness in computer networks. However, it is crucial to consider the reliability of current state-of-the-art models for practical applications in real-world computer networks (Maabreh et al., 2022). The primary focus has been on achieving high performance on a simple dataset, ignoring practical applications (Li et al., 2021). The high cost of errors in IDS and the difficulty in collecting and storing all relevant features in a live computer network feed may have contributed to this trend. (Ahmad et al., 2022; Abdulsalam et al., 2024) affirms the need for further research to enhance the reliability of ML-based intrusion detection systems in computer networks hence this study.

Dimension reduction is a technique used to reduce the complexity of high-dimensional data and keep only the most important information (Jyothsna et al., 2020). It can help improve the accuracy and computational efficiency of intrusion detection system (Zhao et al., 2017). The motivation for this research is to address the challenges faced by IDSs in handling high-dimensional data and to improve the accuracy of intrusion detection systems in computer networks. This study proposes a dimensionality reduction technique based on a mutual information feature selection algorithm with a linear discriminant analysis feature extraction algorithm to obtain only the relevant information from the given data. SVM and KNN are not only simple. efficient but highly flexible when employed for a given task this informed the choice of SVM and KNN to develop IDS and their performance was evaluated in terms of assessment used in this research.

The remaining session was organized into different session. Literature Review, which present a comprehensive review of related literature and establish the research gap. Methodology used employed as well as the system design, including relevant diagrams and the process of data collection. Results obtained were discussed comprehensively. Finally, conclusion and recommendation as well as suggestions for future research direction.

2. Related Works

Several studies have been conducted on using dimension-reduction techniques for intrusion detection. Nabi and Zhou (2024) proposed a model to improve automated intrusion detection by developing a highly accurate classifier with minimal false alarms, using the NSL-KDD dataset and testing classifiers with feature reduction techniques like Random Projection and PCA. Although the achieved improved accuracy and time efficiency from Random Projection, but struggled with potential loss of critical data attributes during dimensionality reduction.

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Shiravani et al. (2023) suggested an enhance network intrusion detection by introducing a novel feature selection method based on fuzzy numbers and correlation-based scoring, tested on the KDD Cup, NSL-KDD, and CICIDS datasets. The proposed method outperforms traditional approaches by reducing the number of features while achieving a high detection accuracy of 99.9%, though further evaluation on real-time data is needed to confirm its practical applicability. also, it is pertinent to include other vital evaluation matrixes. Furthermore, model over-fit due to complexity and improper handling of highly dynamic or evolving network data.

Deore and Bhosale, (2022) proposed intrusion detection system based on the DASO optimization algorithm. A crucial challenge is that it is independent of the research in a fast and sufficient pre-processing phase. Another difficult problem is the increasing increase of zero-day attacks, which highlights the necessity for security systems that can precisely detect previously unknown attacks. Through active feature optimization approaches, an attempt is made to develop a generic meta-heuristic scale for both known and undiscovered assaults with a high detection rate and low false alarm rate. Saheed et al., (2022), utilized machine learning for intrusion detection in IoT networks. They used the UNSW-NB15 dataset and applied feature scaling and dimensionality reduction with PCA. Six machine learning models were then evaluated using metrics such as accuracy, recall, precision, and MCC. The results showed competitive accuracy of 99.9% and MCC of 99.97%.

Kotecha et al. (2021) aims to build an effective network intrusion detection system using the UNSW-NB15 dataset, evaluating multiple models and conducting detailed data analysis. The study identifies the best-performing model through various metrics but highlights potential limitations in generalizing to future, unknown attacks. Another study by Nagpal et al., (2021) suggested an optimization algorithm combined with SVM for intrusion detection systems. The main idea is to utilize Big Bang Big Crunch Optimization to select some features from the NSL-KDD datasets. 41 features, and then utilize SVM Machine to compute the accuracy of these features. Also Kayode-Ajala, (2021) evaluated the effectiveness of various machine learning algorithms for classifying network traffic into normal and anomalous categories, using the NSL-KDD dataset and applying preprocessing, feature scaling, and Principal Component Analysis (PCA) for dimensionality reduction. The results show that the K-Neighbors Classifier achieves the highest accuracy (97.94%), with PCA successfully streamlining computation without significant accuracy loss, but there is a slight reduction in recall, indicating a trade-off between precision and sensitivity.

Alsaadi et al. (2020) proposed an enhance intrusion detection systems by using computational intelligence models, specifically Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO), to select significant features for improved classification. The methodology involves applying these algorithms for feature selection and feeding them into K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naïve Bayes (NB) classifiers, using the KDD Cup and NSL-KDD datasets for validation; results show improved performance over conventional models, but time efficiency in classification requires improvement. In another work Almiani et al., (2020) utilized machine learning techniques for intrusion inherent flow, to better understand the status of machine learning, Almiani analyzed 49 related works from 2009 to 2014 that focused on single, hybrid, and ensemble classifier design architecture. Study showed that dataset in this area is huge hence the need for dimensionality reduction technique for improved detection model hence the study. Yuansheng et al., (2019) used a deep learning approach to identify real-time network intrusion. They found that the approach could effectively recognize intrusion on the network and improve the accuracy of the intrusion detection system.

Kasu Journal of Computer Science Vol. 1 No. 4 [December, 2024], pp. 689-701
https://doi.org//10.47514/kjcs/2024.1.4.001
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Although several dimension reduction techniques have been applied to intrusion detection, limited research has investigated the use of feature selection and feature extraction techniques in combination with traditional and machine learning-based dimension reduction methods for the detection of intrusion in computer networks. Moreover, there is high need for more studies to evaluated the performance of different dimension reduction techniques as researcher are yet to ascertain the most promising technique from the few that have been employed from existing studies (Saheed et al., 2022). Hence, this informed the choice of Mutual Information and Linear Discriminant Analysis algorithms, in combination with SVM and KNN classification methods for this study.

3. Methodology

In this study, machine learning design methodology was employed. The system work flow goes through two stages of development: pre-processing and classification. The input dataset is loaded, and the pre-processing stage involves using Mutual Information to obtain relevant feature subset from the dataset. The selected features are then passed for extraction process, the extracted features are then passed into the classification algorithms, namely SVM and KNN for classification purpose as described in Figure 1.

A. Data collection and Description

To evaluate the effectiveness of the dimension reduction system for intrusion detection, the dataset used was the CIC-IDS2017 dataset, which was collected from the Kaggle repository (*CICIDS2017 | Kaggle*, 2023). The dataset is suitable for the intended purpose as it contains both benign and malicious traffic that closely reflects real-world data. *CICIDS2017* dataset contains attack information of 5days traffic data. Thursday working hour afternoon and Friday data are well suitable for binary classification (Panigrahi and Borah, 2018). The dataset contains 284315 instances and 79 features. During data preparation it was observed that the dataset contains 65410 instances having missing class label and 203 instances having missing information. By removing such missing instances, the dataset instance was reduced to 218702 instances. The remaining instances was then passed for feature selection and extraction. A description of the files contains in the dataset is presented in Table 1

Table 1. Description of contained files in CIC-IDS2017

Name of Files	Attacks Found
Monday-Working_hours	Benign (Normal human activities
Tuesday WorkingHours.pcap	Benign, FTP-Patator, SSH-Patator
WednesdayworkingHours.pcap ISCX.csv	Benign, DOS GoldenEye, Dos Hulk, DOS Slowhttptest, DOS slowloris, Heartbleed
Thursday-WorkingHoursMorning-WebAttacks.pcap_	Benign, Web Attack — Brute Force, Web Attack — Sql Injection, web Attack - XSS
Thursday-WorkingHoursAfternoon-Infilteration.pcap	Benign, Infiltration

Kasu Journal of Computer Science Vol. 1 No. 4 [December, 2024], pp. 689-701
https://doi.org//10.47514/kjcs/2024.1.4.001
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Friday_WorkingHours_Mornin	Benign, Bot
Friday-WorkingHours-AfternoonPortScan	Benign, PortScan
Friday-WorkingHours AfternoonDDos	Benign, DD0S

B. System Development/ Research Design

Figure 1 describes the development of the intrusion detection system, the first process was to select the important features from the dataset by using the mutual information techniques on the CIC-IDS 2017 dataset. Before proceeding to the classification stage, the selected features were further extracted with the LDA technique helped to ensure that the dimensionality of the selected data is reduced to avoid any form of under or over-fitting during the classification stages. These two processes ensure efficient data processing while still preserving important information for intrusion detection.

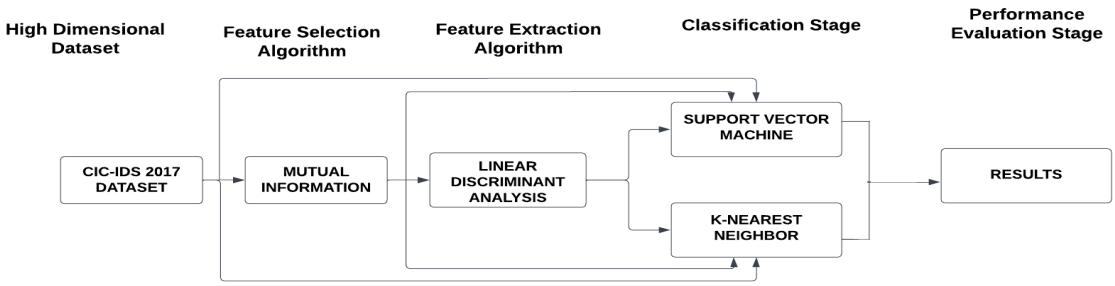

Figure1. Framework of the Developed IDS

C. SVM and KNN Classification

Reduce features from the dimensionality technique was used to develop an SVM and KNN model and performance of the developed was assessed based on various metrics such as accuracy, specificity, sensitivity, precision.

D. Performance Metrics

The final method used in this research is the assessment of the accuracy, specificity, sensitivity, precision, and processing costs of the results. The results of the mutual information with LDA techniques and the SVM and KNN classification performance was compared to those obtained from traditional intrusion detection techniques. The results was analysed to determine the effectiveness of the proposed dimension-reduction system for intrusion detection.

i. Sensitivity

The percentage of actual positives that are correctly identified as positives is measured by sensitivity and is described in equation 1.

$$Sensitivity = TP/(TP + FN)\ \% \quad (1)$$

Kasu Journal of Computer Science Vol. 1 No. 4 [December, 2024], pp. 689-701
https://doi.org//10.47514/kjcs/2024.1.4.001
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

ii. Specificity

The fraction of actual negatives accurately classified as negatives is measured by specificity, also known as selectivity or true negative rate (TNR). Equation 2. Describes the how specificity is been derived mathematically.

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FN})\ \% \quad (2)$$

Sensitivity (true positive fraction) is the likelihood that a diagnostic test will be positive if the person has the condition. While Specificity (true negative fraction) is the probability that a diagnostic test will be negative if the person does not have the condition.

iii. Accuracy

Accuracy is the possibility that a diagnostic test will be done accurately. The number of correct predictions made by the model from all sorts of predictions generated is characterized as accuracy in classification problems. Accuracy is a good measure when the target variable classes in the data are roughly balanced. The equation used in deriving the accurate values for each algorithm model is shown in Equation 3.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})\ \% \quad (3)$$

iv. Precision

Precision is a measure that indicates how accurately the estimates of each model performed on a percentage basis and is illustrated in Equation 4.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (4)$$

Where:

TN (True Negative) = correctly classified negative cases, TP (True Positives) = correctly classified positive cases, FN (False Negative) = incorrectly classified positive cases, and FP (False Positives) = incorrectly classified negative cases.

4. Results and discussion

This section presents the implementation's results and its evaluation. It also contains a full explanation of the findings and results of the proposed model. The evaluation's results support the need for the developed IDS considering this paper's objectives. The raw data were pre-processed. Mutual Information and Linear Discriminant were employed as a dimensionality reduction technique and optimal features were used to build SVM and KNN classifiers.

A. Experimental Analysis

In this research, Python programming language with kernel version 3.7 was used to implement all the techniques and models, the environment used for the execution of each program code was the Jupiter notebook which was built specifically to run Python codes. The experimental machine used for the simulation is a MacBook Pro with a processor speed of 3.6GHz quad-core intel core i7, with a dedicated

Kasu Journal of Computer Science Vol. 1 No. 4 [December, 2024], pp. 689-701
https://doi.org//10.47514/kjcs/2024.1.4.001
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

GPU specification of 2GB and 16GB of system RAM. Data normalization was performed to improve the data's quality and consistency, with the scaling range from 0 to 1.

The dataset was divided into training and testing sets. The model was trained on 70% of the data and tested on 30%. To avoid longer testing time the testing dataset was also divided into two halves with one been the validation dataset. The train data is described in Figure 2. The experiments were designed to evaluate the proposed system's accuracy, specificity, sensitivity, precision, and processing costs. The accuracy of the proposed system was analysed using various classification algorithms, including SVM and KNN. The proposed system's specificity, sensitivity, and precision were also analysed.

```
# Descriptive analysis of the data
X_train.describe()
```

	duration	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	num_failed_logins	logged_in	num_compromised
count	17634.000000	1.763400e+04	1.763400e+04	17634.000000	17634.000000	17634.000000	17634.000000	17634.000000	17634.000000	17634.000000
mean	328.047125	3.019729e+04	2.884210e+03	0.000113	0.024215	0.000057	0.203471	0.001021	0.392254	0.243280
std	2815.155934	2.878748e+06	7.036947e+04	0.010649	0.263108	0.007531	2.162643	0.038386	0.488267	10.756051
min	0.000000	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	0.000000	4.400000e+01	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	0.000000	2.767500e+02	5.110000e+02	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
max	42862.000000	3.817091e+08	5.150836e+06	1.000000	3.000000	1.000000	30.000000	3.000000	1.000000	884.000000

8 rows × 38 columns

Figure 2. Description of the training data after splitting

B. Result of the Developed Intrusion Detection System

The intrusion detection system was developed using SVM and KNN, the result of the classification achieved by both algorithm is described in Figure. 3 and 4. After the system has been developed, the details of each amount of fraudulent transaction against the normal transactions recorded is provided in Figure 5. These results indicate that the developed system was highly influential in detecting network intrusions while maintaining a low false positive rate and false discovery rate. The results provide empirical evidence of the effectiveness of the developed system for intrusion detection in computer networking. Figure 4.2 showed that the result of the confusion matrixes for SVM classification were TP of 3981, TN of 3489, FP of 61 and FN of 27. While Figure 4.3 showed that the result of the confusion matrixes for KNN classification were TP of 3824, TN of 2991, and FP of 218 and FN of 525. Hence it is evident that the result of the experiment shows that the system achieved an accuracy of 98.84%, specificity of 98.28%, precision of 98.49%, a negative predictive value of 99.23%, a false positive rate of 0.17%, and false discovery rate of 0.15% as shown in Table 2.

Kasu Journal of Computer Science Vol. 1 No. 4 [December, 2024], pp. 689-701
https://doi.org//10.47514/kjcs/2024.1.4.001
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

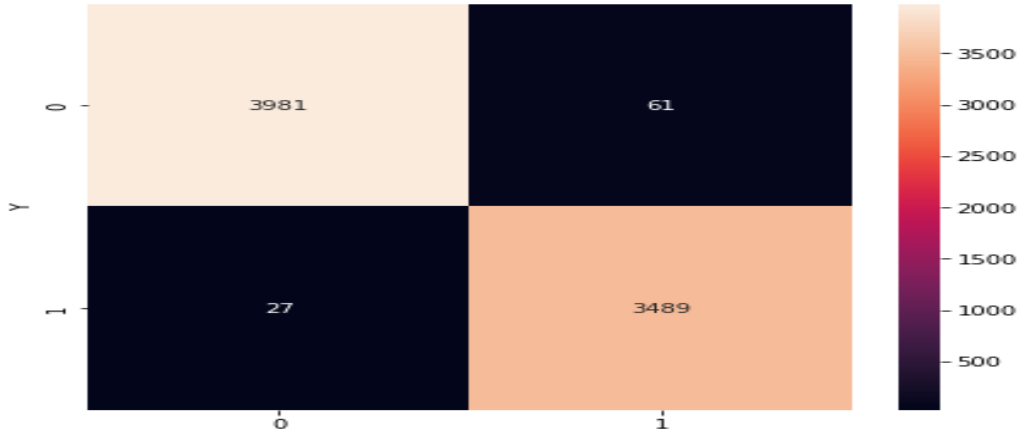

Figure 3. Result of SVM classification

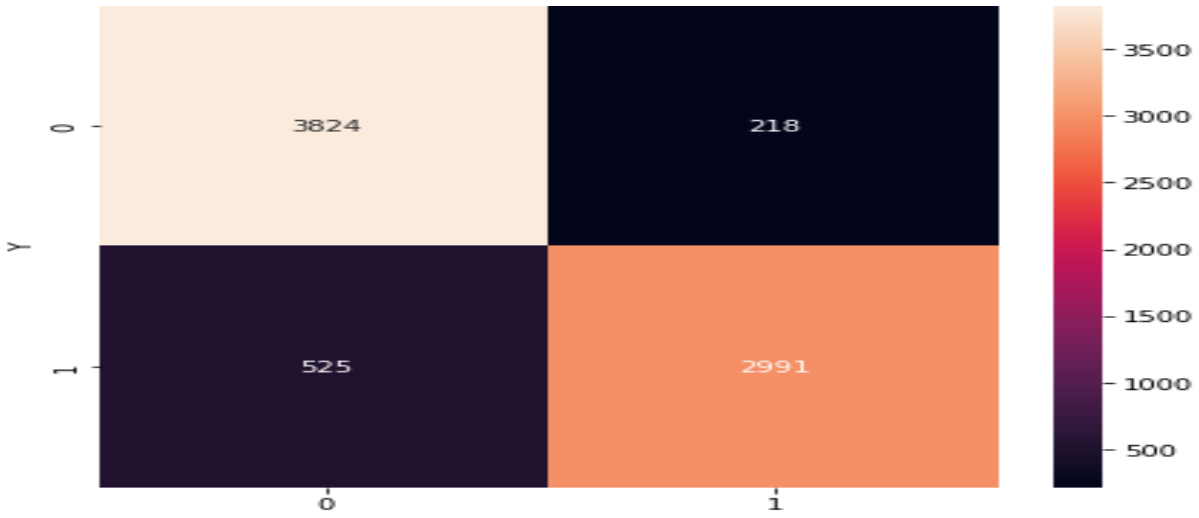

Figure 4. Result of KNN

Print the amount details of fraudulent transaction

```
print('Amount details of fraudulent transaction :-')
fraud.Amount.describe()
```

```
Amount details of fraudulent transaction :-
count     492.000000
mean      122.211321
std       256.683288
min         0.000000
25%         1.000000
50%         9.250000
75%       105.890000
max      2125.870000
Name: Amount, dtype: float64
```

```
print('Amount details of Normal transaction :-')
normal.Amount.describe()
```

```
Amount details of Normal transaction :-
count    284315.000000
mean         88.291022
std         250.105092
min           0.000000
25%           5.650000
50%          22.000000
75%          77.050000
max       25691.160000
Name: Amount, dtype: float64
```

Figure 5. Figure Description of Fraudulent and normal transaction after classification

Kasu Journal of Computer Science Vol. 1 No. 4 [December, 2024], pp. 689-701
https://doi.org//10.47514/kjcs/2024.1.4.001
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Table 2. Result of the Experiment

Performance matrices (%)	**Data+MI+LDA +SVM**	**Data + MI+LDA +KNN**	**Formula**
Accuracy	98.84	90.17	$(TP + TN)/(P + N)$
Specificity	98.28	93.21	$TN/(FP + TN)$
Precision	98.49	94.61	$TP/(TP + FP)$
Negative predictive value	99.23	85.07	$TN/(TN + FN)$
False positive rate	0.17	6.79	$FP/(FP + TN)$
False discovery rate	0.15	0.54	$FP/(FP + TP)$
False negative rate	0.67	12.07	$FN/(FN + TP)$
Sensitivity	99.33	87.93	$TP/(TP + FN)$
F1- score	98.91	91.15	$2TP/(2TP + FP + FN)$

C. Comparative Analysis

In this section, a comparison of the accuracy results was made with previous studies in the field. The accuracy metric was chosen as a measure of comparison due to its widespread use in evaluating model performance. Table 3 compares the accuracy of the proposed MI-LDA combined with SVM and KNN with the results from other studies that employed machine learning algorithms for intrusion detection. This comparison was limited to studies that used similar approaches to ensure that the criteria for comparison were aligned with the scope of this study. The results indicated that the proposed model in this research showed superior accuracy compared to previous studies, as reflected in Table 3.

Kasu Journal of Computer Science Vol. 1 No. 4 [December, 2024], pp. 689-701
https://doi.org//10.47514/kjcs/2024.1.4.001
Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Table 3 Comparative analysis of results of this study with related works

Authors	Methods	Accuracy
(Rincy and Gupta, 2021)	Decision trees	80.77 %
(Shin et al., 2020)	Decision trees and random forests	88.89%
(Saranya et al., 2020)	LDA algorithm and Random Forest	88.1%.
(Wani et al., 2019)	SVM and CART algorithm	87.60%
Present research	**MI+LDA +SVM**	**98.84%**
	MI+LDA +KNN	**90.17%**

The experiment results showed that the proposed dimension reduction system was highly influential in enhancing intrusion detecting rate on the network while maintaining a low false positive rate and false discovery rate. The system's accuracy, specificity, precision, negative predictive value, false positive rate, and false discovery rate were 98.84%, 98.28%, 98.49%, 99.23%, 0.17%, and 0.15%, respectively. These results indicate that the proposed system effectively reduced the dimensionality of the data while preserving the relevant information for intrusion detection. The high accuracy, specificity, and precision demonstrate the system's ability to effectively distinguish between normal and malicious network traffic, which is essential for intrusion detection. Additionally, the low false positive rate and false discovery rate suggest that the system has a low rate of incorrect detections, which can reduce the risk of false alarms in network security.

The results also showed that the proposed system achieved high processing efficiency by using mutual information and LDA techniques for feature selection and extraction and SVM and KNN for classification. The high processing efficiency of the proposed system makes it suitable for real-time intrusion detection in computer networking. From the above statements, it clear that the results of the proposed dimension reduction system for intrusion detection showed that the system could provide better accuracy, specificity, sensitivity, precision, and processing costs than other existing methods that was reviewed. Hence the results of this study confirmed the effectiveness and efficiency of the proposed system for intrusion detection in computer networking. However, employing other vital techniques such as ensemble techniques, deep learning techniques, and regularisation of parameter, optimisation of parameter and transfer techniques can play vital role in enhancing model performance.

5. Conclusion and Future Work

Intrusion detection is critical to computer network security, as it helps organizations protect their networks from cyber threats. Hence Combination of Mutual Information and Linear Discriminant Analysis dimensionality reduction algorithms as a pre-processing technique as well as SVM and KNN at the

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

classification have helped in the development of a robust model for intrusion detection purpose. The results of this study make a significant contribution to the field of intrusion detection in computer networking by providing a dimension reduction system capable of accurately detecting common attacks in computer networks. Hence the developed system can be recommended to be valuable for researchers and practitioners working in the field of computer network security. The increasing use of computer networks and the increasing sophistication of cyber-attacks have made intrusion detection an important area of study. In this research a dimension-reduction system for intrusion detection was developed to address the challenges of detecting intrusions in high-dimensional and complex computer networks. The system uses Mutual Information and LDA techniques to reduce the dimensionality of the network data, followed by classification using SVM and KNN to identify intrusions.

The study also evaluated the performance of the proposed system in terms of accuracy, specificity, sensitivity, precision, and processing costs and compared its results with other methods. The CIC-ID2017 intrusion dataset, which included a representation of both benign and malicious network traffic, was sourced from the Kaggle repository, and used in this research. The dataset was pre-processed and split into training (70%) and testing (30%) sets. The KNN and SVM classifiers were applied to the training data and achieved 98.84% and 90.17% accuracy, respectively. The model was evaluated using performance metrics such as accuracy, sensitivity, precision, specificity, F1-score. The results were also compared with previous research work. Despite the promising results of this study, some limitations need to be addressed. Future works could be extended to include other feature selection and extraction techniques, such as feature importance and correlation, to improve the system's performance, also suitable data balancing and ensemble techniques could be utilized for better performance. Additionally, the study could be extended to include other classification algorithms, such as Random Forest and Neural Networks, methods to reduce the false positive rate of the system can be developed, as this is a common challenge in intrusion detection.

References

Abdulsalam, S. O., Ayofe, R. A., Edafeajiroke, M. F., Ajao, J. F., and Babatunde, R. S. (2024). Development of an intrusion detection system using mayfly feature selection and artificial neural network algorithms. LAUTECH Journal of Engineering and Technology, 8(2), 148–160. https://doi.org/10.36108/laujet/4202.81.0241

Ahmad B. and Magaji Y. M. (2024). Securing IoT Networks: An Intrusion Detection Framework Using Convolutional Neural Networks. Kasu Journal of Computer Science Vol. 1 No. 2 [June, 2024], pp. 218-233. Online ISSN: 1597-2216 Print ISSN: 1597-2178 www.kjcs.edu.ng. https://doi.org//10.47514/kjcs/2024.1.2.006

Ahmad, K., Maabreh, M., Ghaly, M., Khan, K., Qadir, J., and Al-Fuqaha, A. (2022). Developing future human-centered smart cities: Critical analysis of smart city security, Data management, and Ethical challenges. *Computer Science Review*, *43*, 100452. https://doi.org/10.1016/j.cosrev.2021.100452

Albulayhi, K., Abu Al-Haija, Q., Alsuhibany, S. A., Jillepalli, A. A., Ashrafuzzaman, M., and Sheldon, F. T. (2022). IoT Intrusion Detection Using Machine Learning with a Novel High Performing Feature Selection Method. *Applied Sciences*, *12*(10), 5015. https://doi.org/10.3390/app12105015

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Almiani, M., AbuGhazleh, A., Al-Rahayfeh, A., Atiewi, S., Abdul Razaque, and Razaque, A. (2020). Deep recurrent neural network for IoT intrusion detection system. *Simulation Modelling Practice and Theory*, *101*, 102031. https://doi.org/10.1016/j.simpat.2019.102031

Alsaadi, H. I., Almuttairi, R. M., Bayat, O., and Uçan, O. N. (2020). Computational intelligence algorithms to handle dimensionality reduction for enhancing intrusion detection system. Journal of Information Science and Engineering, 36, 293–308. https://doi.org/10.6688/jise.202003_36(2).0009

Choudhary, S., and Kesswani, N. (2021). A Hybrid Classification Approach for Intrusion Detection in IoT Network. In *Journal of Scientific & Industrial Research* (Vol. 80).

Deore, B., and Bhosale, S. (2022). Intrusion Detection System with a Modified DASO Optimization Algorithm. *International Journal on Engineering, Science and Technology*, *4*(1), 54–63. https://doi.org/10.46328/ijonest.66

Jyothsna, V., Sreedhar, A. N., Mukesh, D., and Ragini, A. (2020). *A Network Intrusion Detection System with Hybrid Dimensionality Reduction and Neural Network Based Classifier*. 187–196. https://doi.org/10.1007/978-981-15-0936-0_19

Kayode-Ajala, O.,(2021). Anomaly Detection in Network Intrusion Detection Systems using and Machine Learning and Dimensionality Reduction SSRAML SageScience, 4(1), 12–26.

Kotecha, K., Verma, R., Rao, P. V., Prasad, P., Mishra, V. K., Badal, T., Jain, D., Garg, D., and Sharma, S. (2021). Enhanced Network Intrusion Detection System. Sensors, 21(23), 7835. https://doi.org/10.3390/s21237835

Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., Liu, X., and He, B. (2021). A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection. *IEEE Transactions on Knowledge and Data Engineering*. https://doi.org/10.1109/TKDE.2021.3124599

Maabreh, M., Obeidat, I., Elsoud, E. A., Alnajjar, A., Alzyoud, R., and Darwish, O. (2022). Towards Data-Driven Network Intrusion Detection Systems: Features Dimensionality Reduction and Machine Learning. *International Journal of Interactive Mobile Technologies*, *16*(14), 123–135. https://doi.org/10.3991/ijim.v16i14.30197

Nabi, F., and Zhou, X. (2024). Enhancing intrusion detection systems through dimensionality reduction: A comparative study of machine learning techniques for cyber security. Cyber Security and Applications, 2, 100033. https://doi.org/10.1016/j.csa.2023.100033

Nagpal, M., Kaushal, M., and Sharma, A. (2021). A Feature Reduced Intrusion Detection System with Optimized SVM Using Big Bang Big Crunch Optimization. *Wireless Personal Communications*, 1–27. https://doi.org/10.1007/s11277-021-08975-2

Panigrahi, R., and Borah, S. (2018). A detailed analysis of CICIDS2017 dataset for designing Intrusion Detection Systems. In *Article in International Journal of Engineering & Technology* (Vol. 7, Issue 3). https://www.researchgate.net/publication/329045441

Online ISSN: 1597-2216
Print ISSN: 1597-2178
www.kjcs.edu.ng

Rincy, T., and Gupta, R. (2021). Design and Development of an Efficient Network Intrusion Detection System Using Machine Learning Techniques. *Wireless Communications and Mobile Computing*, *2021*, 1–35. https://doi.org/10.1155/2021/9974270

Saheed, K. Y., Idris, A. A., Misra, S., Kristiansen, H. M., and Colomo-Palacios, R. (2022). A machine learning-based intrusion detection for detecting internet of things network attacks. *Alexandria Engineering Journal*, *61*(12), 9395–9409. https://doi.org/10.1016/j.aej.2022.02.063

Shiravani A., Sadreddini, M. H., and Nahook, H. N.(2023). Network intrusion detection using data dimensions reduction techniques.Journal of Big Data (2023) 10:27 https://doi.org/10.1186/s405 37-023-00697-5

Wani, A. R., Rana, Q. P., Saxena, U., and Pandey, N. (2019). Analysis and Detection of DDoS Attacks on Cloud Computing Environment using Machine Learning Techniques. *Proceedings - 2019 Amity International Conference on Artificial Intelligence, AICAI 2019*, 870–875. https://doi.org/10.1109/AICAI.2019.8701238

Yang, X., Peng, G., Zhang, D., and Lv, Y. (2022). An Enhanced Intrusion Detection System for IoT Networks Based on Deep Learning and Knowledge Graph. *Security and Communication Networks*, *2022*, 1–21. https://doi.org/10.1155/2022/4748528

Yuansheng, D., Wang, R., and He, J. (2019). Real-Time Network Intrusion Detection System Based on Deep Learning. *2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)*. https://doi.org/10.1109/icsess47205.2019.9040718

Zeyuan F. (2022). Computer Network Intrusion Anomaly Detection with Recurrent Neural Network. *Mobile Information Systems*, *2022*, 1–11. https://doi.org/10.1155/2022/6576023

Zhang, X., Ding, S., and Xue, Y. (2017). An improved multiple birth support vector machine for pattern classification. *Neurocomputing*, *225*, 119–128. https://doi.org/10.1016/j.neucom.2016.11.006

Zhao, S., Li, W., Zia, T., and Zomaya, A. Y. (2017). A Dimension Reduction Model and Classifier for Anomaly-Based Intrusion Detection in Internet of Things. *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, 836–843. https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2017.141