**Faculty of Natural and Applied Sciences Journal of Mathematical and Statistical Computing**
**Print ISSN: 3026-8605, e-ISSN 3043-632X**
www.fnasjournals.com
**Volume 2; Issue 3; May 2025; Page No. 44-50.**
DOI: https://doi.org/10.63561/jmsc.v2i3.856

# A Fusion-Based Data Mining Model for Intrusion Detection in Distributed Environments

***[1]Olojido, J.B., & [2]Oriola, O.**
[1]Department of Computer Science, Rufus Giwa Polytechnic Owo
[2]Department of Computer Science, Adekunle Ajasin University Akungba Akoko

***Corresponding author email:** olojidoj@gmail.com

**Abstract**
As distributed systems become more prevalent, the frequency of distributed attacks—such as distributed denial-of-service (DDoS) and worms—is increasing. Traditional intrusion detection systems are struggling to efficiently identify and report threats in a timely manner. Consequently, various Distributed Intrusion Detection Systems (DIDS) utilizing machine learning algorithms have been implemented. However, their effectiveness has been limited due to high computational costs and suboptimal accuracy levels. This paper aims to enhance the Fusion-Based Data Mining Model for Intrusion Detection in Distributed Environments. Two well-known distributed attack datasets, NSL-KDD'15 and UNSW-NB'15, were utilized in this study. (Fusion-Based Data Mining Model (FBDMM)) was chosen as the evaluation framework due to its widespread use. To minimize computational costs, both Principal Component Analysis (PCA) and Information Gain Ratio (IGR) were employed to extract the five most significant features from each dataset. Classifiers such as Support Vector Machines (SVM), Naïve Bayes (NB), and Multilayer Perceptron (MLP) were The hybridized using a Voting Classification technique to boost accuracy. The The hybridized Data Mining Model (Fusion-Based Data Mining Model (FBDMM)) comprised six classifiers: PCA+SVM, IGR+SVM, PCA+NB, IGR+NB, PCA+MLP, and IGR+MLP. The evaluation results were compared across these six classifiers based on accuracy (ACC), detection rate (DR), and false alarm rate (FAR). Computational costs, measured in System Running Time (SRT), were compared between five-feature and full-feature sets: forty-one features for NSL-KDD'15 and forty-nine features for UNSW-NB'15. The Fusion-Based Data Mining Model (FBDMM) achieved ACC, DR, and FAR values of 77.78, 96.98, and 2.55, respectively, while the highest performance among individual classifiers for NSL-KDD'15 was 72.17, 92.29, and 2.71. For UNSW-NB'15, the Fusion-Based Data Mining Model (FBDMM) recorded ACC, DR, and FAR values of 85.58, 95.98, and 3.35, respectively, with the best performance from individual classifiers being 82.88, 97.23, and 4.66. The SRT for NSL-KDD'15 was 10 seconds with five features and 5,200 seconds with forty-one features, while for UNSW-NB'15, it was 9 seconds with five features and 68,000 seconds with forty-nine features. The findings indicate that fusion-based Data Mining Model outperforms existing data mining models used in Distributed Intrusion Detection Systems in terms of both accuracy and computational cost. Therefore, fusion-based Data Mining Model is recommended for use in Distributed Intrusion Detection.

**Keywords:** Intrusion Detection System, Distributed Attacks, Data Mining, Hybrid Model, Accuracy, Computational Cost

## Introduction

Intrusion detection is the process of monitoring activities in computer system or network, analyzing them to recognize the signs of attacks, and then acknowledge as attempts harmful to the confidentiality, integrity and availability of the security mechanisms of computer or network (Abdurrazaq et al., 2014). It is a proactive defense technology which is the focus of the attention of computer and network security today. It can timely detect network vulnerabilities and rapidly respond to, and enhance network security. Preventing unauthorized access to these system resources and data requires the design of robust security mechanisms. However, completely preventing breaches of security appears currently unrealistic but these intrusions can be detected in an attempt to take action before the damage happens (Du et al., 2004). The essence of IDS is to detect computers break-ins, penetrations, denial of service attacks, ports scans and other forms of computer abuse that exploit security vulnerabilities or


*Cite this article as*:
Olojido, J. B., & Oriola, O. (2025). A fusion-based data mining model for intrusion detection in distributed environments. *FNAS Journal of Mathematical and Statistical Computing*, *2*(3), *44-50*. https://doi.org/10.63561/jmsc.v2i3.856

flows in systems. Intrusion Detection Systems (IDS) nowadays represent a well-known solution to identify unauthorized and malicious use of the network resources. The relevance of the tasks carried out by an IDS necessarily poses several requirements to be met which includes the accuracy of the detection process. Accuracy is related to the capability of the system to both correctly classify ongoing attacks and avoid misclassifications of normal traffic as attacks (Antonio et al., 2010).

Intrusion detection implements two basic approaches; signature-based and anomaly based. Ability of signatures-based detection rely on the known information about attack patterns in database. This method does not have ability to detect a new pattern that has not been defined in the database. Anomaly detection rely on a history of normal pattern, and detect deviations then classify as an anomaly. An IDS can run in real-time and offline within two main architectures, centralized and distributed. Generally, IDS developed using a centralized architecture identify attacks on a single monitored system. The trend of distributed and coordinated attacks in today's information systems use many machines as an attacker or victims (Abdurrazaq et al., 2014).

Detecting intrusions on a single host or information system has been achievable by designers of different IDS because few patterns of intrusions are perpetrated on these systems. Intruders have devised several means of attacking different systems hosted in a distributed network environment which is proving challenging to tracked and detect intrusions due to their variations. Therefore, distributed intrusion detection system (Fusion-Based Data Mining Model (FBDMM)) provides a better solution to solve these challenges. DIDS is an intrusion Detection System capable of detecting multiple hosts, multiple segment data and information associated with that information (correlation) and comprehensive analysis to detect possible distributed network attacks and process the response (Hui, 2011). Distributed Intrusion Detection System is an important supplement of the traditional reactive approach of network security detection. Traditional intrusion detection technology in the network implements a centralized architecture where network information is collected by placing multiple detectors (sensors) in the network, and send the information to a central console for analysis. This model appears to be inadequate in the face of large-scale, heterogeneous network environment, and the case of distributed coordinated attacks. This is because the load of the center console is too large, thus it becomes the bottleneck of system operation and also the network transmission delays which allow untimely delivery of information probe sent to the center console network. It is unable to detect distributed attack accurately as it lacks the ability to link attack information from various sub-network. However, in a distributed environment, the detection of intrusion from variety of sources represents a complex task and searching for intrusion attributes has a non-deterministic character, where multiple numbers of intrusions are generated and are challenging to detect. To tackle the distributed nature of this attack. The objectives of the research are to, design a fusion-based data mining model for intrusion detection system using data mining technique and to implement the model, implement a model a fusion base data mining for distributed intrusion detection.

**Review of Related Works**

Aladesote et al. (2016) proposed a feature extraction approach for intrusion detection systems using Gain Ratio and Principal Component Analysis (PCA). Motivated by the low detection rates in existing systems, the study aimed to enhance performance by selecting relevant features from the KDD'99 dataset through these techniques.

Cepheli et al. (2016) proposed a Hybrid Intrusion Detection System (H-IDS) that combines signature-based and anomaly-based techniques to improve intrusion detection. The system integrates a Gaussian Mixture Model (GMM)-based anomaly detector and a Snort-based signature detector, with a hybrid engine managing their outputs. A decision combiner then evaluates the results and triggers alarms based on a sensitivity parameter.

Alkasassbeh et al. (2016) proposed a data mining-based model for detecting Distributed Denial of Service (DDoS) attacks, aiming to identify previously unknown attack types. The study employed Multilayer Perceptron, Naïve Bayes, and Random Forest classifiers on a newly collected dataset containing modern DDoS attacks to effectively classify and detect them. Qin (2017) presented a study on intrusion detection using a distributed collaborative structure, aiming to improve Internet information security. Motivated by issues such as low efficiency, complex configuration, and slow response in existing collaborative IDS, the research proposed a Distributed Intrusion Detection System framework based on a ring structure to enhance coordination and performance.

Ibrahim and Zainal (2018) proposed an adaptive and distributed intrusion detection model for cloud computing, aiming to enhance security and quality of service. Motivated by the need to detect coordinated attacks and address virtual machine migration challenges, the model includes components for feature selection, change point detection, adaptive detection, and aggregation.. Al-Dabbagh (2017), presented a research on an Intrusion Detection System for Cyber Attacks in


*Cite this article as*:
Olojido, J. B., & Oriola, O. (2025). A fusion-based data mining model for intrusion detection in distributed environments. *FNAS Journal of Mathematical and Statistical Computing, 2*(3), *44-50*. https://doi.org/10.63561/jmsc.v2i3.856

Wireless Networked Control Systems. The research was motivated by the need to study a wireless control topology and identify the existence of attacks. The objective was to present a modelling framework for the closed-loop control system with IDS, and a computational procedure to design and compute the IDS

Ibrahim and Zainal (2018) proposed an adaptive and distributed intrusion detection model for cloud computing to enhance security and quality of service. The model aims to detect coordinated attacks and address challenges posed by virtual machine migration. It consists of four key components: feature selection, change point detection, adaptive detection, and aggregation.

**Materials and Methods**

This presents the analysis and design and evaluation of the fusion base data mining system for distributed intrusion detection which combines Principal Component Analysis (PCA) and Information Gain Ratio (IGR) as feature selection techniques and Support Vector Machine (SVM), Multilayer Perceptron (MLP) and Naïve Bayes (NB) as classifiers. Fegure 1: described architecture features a data mining-based intrusion detection system using feature selection methods—Principal Component Analysis (PCA) and Information Gain Ratio (IGR)—alongside classification algorithms: Support Vector Machine, Multilayer Perceptron, and Naïve Bayes. These classifiers are integrated through a voting mechanism. The system comprises two independently operated IDS setups, each with three core modules: data preprocessing, classification learning, and classification result.


*Cite this article as*:
Olojido, J. B., & Oriola, O. (2025). A fusion-based data mining model for intrusion detection in distributed environments. *FNAS Journal of Mathematical and Statistical Computing*, *2*(3), *44-50*. https://doi.org/10.63561/jmsc.v2i3.856

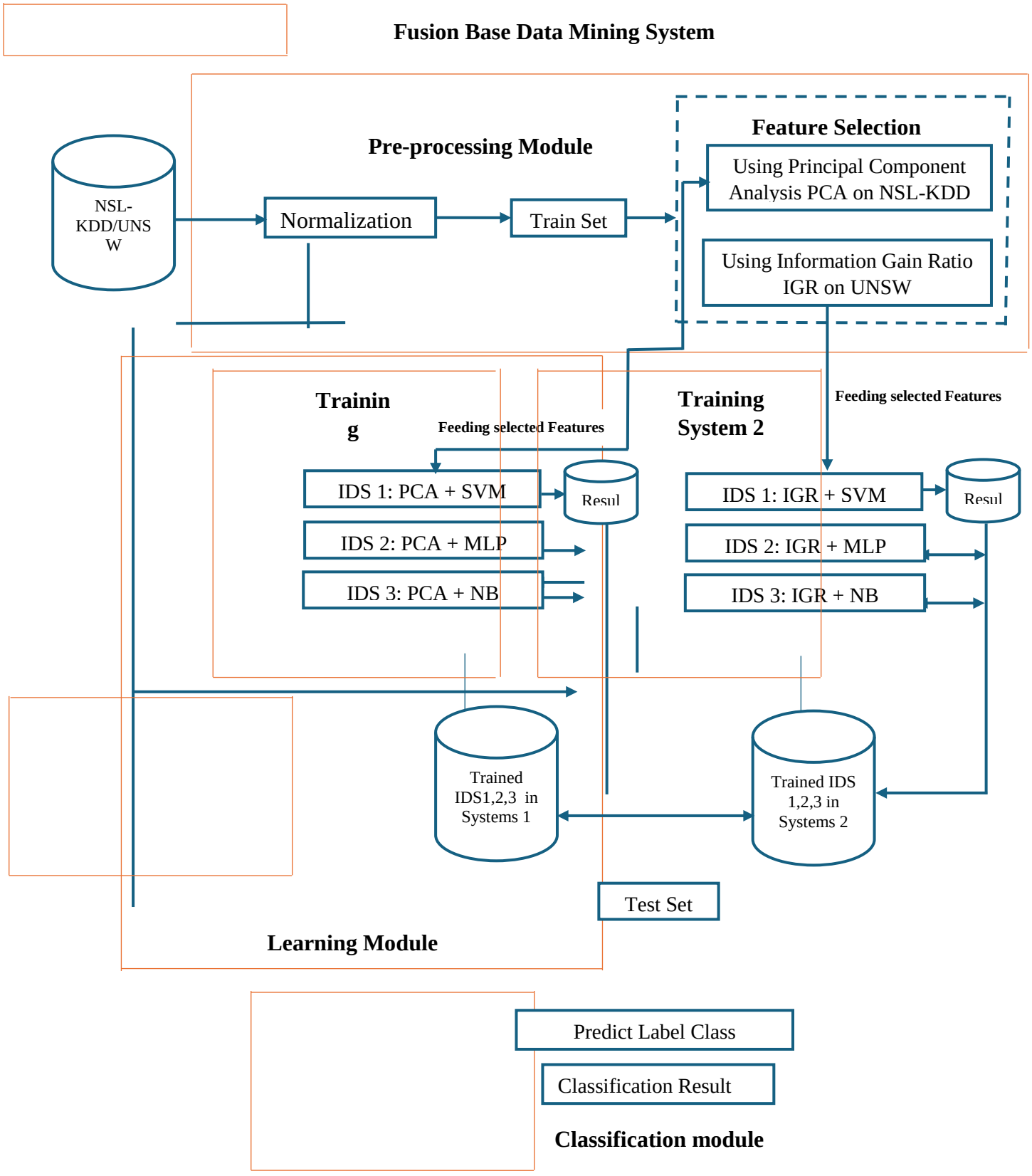


Figure 1: Architecture of Fusion-Based Data Mining Model


*Cite this article as*:
Olojido, J. B., & Oriola, O. (2025). A fusion-based data mining model for intrusion detection in distributed environments. *FNAS Journal of Mathematical and Statistical Computing*, *2*(3), *44-50*. https://doi.org/10.63561/jmsc.v2i3.856

The described architecture features a data mining-based intrusion detection system using feature selection methods—Principal Component Analysis (PCA) and Information Gain Ratio (IGR)—alongside classification algorithms: Support Vector Machine, Multilayer Perceptron, and Naïve Bayes. These classifiers are integrated through a voting mechanism. The system comprises two independently operated IDS setups, each with three core modules: data preprocessing, classification learning, and classification result.

In preprocessing, datasets are normalized using min-max normalization and split into training and testing sets. PCA and IGR are applied to the training set for feature selection. Evaluation metrics include Accuracy (ACC), Detection Rate (DR), and False Alarm Rate (FAR), calculated based on true/false positives and negatives.

For implementation, the KDDTrain+.txt and KDDTest.txt datasets were used, filtered to include only Normal and DoS intrusion types. The training set contained 67,343 Normal and 45,927 DoS samples, while the test set had 17,169 samples with a similar distribution. PCA was implemented using the Scikit-learn library to extract the top 10 principal components for classification.

**Classification Methods**

The hybridization of Support Vector Machine (SVM), MultiLayer Perceptron (MLP), and Naïve Bayes (NB) was used as classification algorithm for the proposed model. Voting Classification was applied to combine the predictions of the three algorithms where the final class label is the class label that has been predicted most frequently by the classification models. Assuming the three classifiers (SVM, MLP, and NB) classify a training sample as follows:

i. SVM ⟶ class 0
ii. MLP ⟶ class 0
iii. NB ⟶ class 1

The hybridized model will predict/ classify the sample as class 0 using the concept of majority vote as class 0 was most classified.

**Performance Evaluation**

The hybridization of Support Vector Machine (SVM), MultiLayer Perceptron (MLP), and Naïve Bayes (NB) was used as classification algorithm(s) for the proposed model. Voting Classification was applied to combine the predictions of the three algorithms where the final class label is the class label that has been predicted most frequently by the classification models. Assuming the three classifiers (SVM, MLP, and NB) classify a training sample as follows:

**Table 1: Confusion matrix of the models on NSL-KDD Dataset with Gain Ratio**

| Models | TP | FP | TN | FN |
|---|---|---|---|---|
| SVM | 7842 | 816 | 8894 | 4990 |
| MLP | 8914 | 667 | 9043 | 3918 |
| NB | 6822 | 263 | 9447 | 6010 |
| Hybridized | 7637 | 385 | 9325 | 5195 |

**Table 2: Evaluation of results of the Models**

| Models | Detection Rate (%) | FAR (%) | Accuracy (%) |
|---|---|---|---|
| SVM | 90.57 | 8.40 | 74.24 |
| MLP | 93.04 | 6.86 | 74.89 |
| NB | 96.29 | 2.71 | 72.17 |
| Hybridized | 95.20 | 3.96 | 75.25 |


*Cite this article as*:
Olojido, J. B., & Oriola, O. (2025). A fusion-based data mining model for intrusion detection in distributed environments. *FNAS Journal of Mathematical and Statistical Computing*, *2*(3), *44-50*. https://doi.org/10.63561/jmsc.v2i3.856

**Table 3. Confusion matrix of the models on UNSW-NB15 Dataset with information Gain**

| Models | TP | FP | TN | FN |
|---|---|---|---|---|
| **SVM** | 118534 | 16097 | 39903 | 807 |
| **MLP** | 93896 | 2500 | 53500 | 25445 |
| **NB** | 32393 | 2152 | 53848 | 86948 |
| **Hybridized** | 98170 | 4117 | 51883 | 21171 |

**Table 4: Evaluation of results of the Models**

| Models | Detection Rate (%) | FAR (%) | Accuracy (%) |
|---|---|---|---|
| **SVM** | 88.04 | 28.75 | 78.67 |
| **MLP** | 94.41 | 4.46 | 84.06 |
| **NB** | 93.77 | 3.84 | 49.18 |
| **Hybridized** | 95.98 | 3.35 | 85.58 |

| Models | TP | FP | TN | FN |
|---|---|---|---|---|
| **SVM** | 7888 | 571 | 9139 | 4944 |
| **MLP** | 8114 | 304 | 9406 | 4718 |
| **NB** | 7878 | 298 | 9412 | 4954 |
| **Hybridized** | 7964 | 248 | 9462 | 4868 |

**Table 5: Evaluation of results of the Models**

| Models | Detection Rate (%) | FAR (%) | Accuracy (%) |
|---|---|---|---|
| **SVM** | 93.25 | 5.88 | 75.53 |
| **MLP** | 96.38 | 3.13 | 77.72 |
| **NB** | 96.36 | 3.07 | 76.70 |
| **Hybridized** | 96.98 | 2.55 | 77. 78 |

| Models | TP | FP | TN | FN |
|---|---|---|---|---|
| **SVM** | 91291 | 3550 | 52450 | 28050 |
| **MLP** | 88348 | 625 | 55375 | 30993 |
| **NB** | 88065 | 5732 | 50268 | 31276 |
| **Hybridized** | 91932 | 2612 | 53388 | 27409 |

**Table 6: Evaluation of results of the Models**

| Models | Detection Rate (%) | FAR (%) | Accuracy (%) |
|---|---|---|---|
| **SVM** | 96.26 | 6..33 | 81.97 |
| **MLP** | 96.89 | 4.57 | 81.97 |
| **NB** | 93.89 | 5.332 | 78.89 |
| **Hybridized** | 97.23 | 4.664 | 82.88 |


*Cite this article as*:
Olojido, J. B., & Oriola, O. (2025). A fusion-based data mining model for intrusion detection in distributed environments. *FNAS Journal of Mathematical and Statistical Computing*, *2*(3), *44-50*. https://doi.org/10.63561/jmsc.v2i3.856

**Discussion**

Conclusively, a The hybridized intrusion detection system that combines the predictions of SVM, MLP, and NB using Majority Voting classification has been established. The developed model was built and validated using NSL-KDD and UNSW-NB15 Datasets. PCA and IG were used as feature selection techniques, and 10 relevant features were selected from the datasets before they were fed to the developed model for classification of network instances as either an attack or a normal connection. The result obtained from the developed model was compared with the individual models. From the result, it was evident that the hybridized model slightly outperformed the individual models with an accuracy of 85% which shows that the hybridized model is effective in detecting intrusions over a network. This research proposes a solution using data mining algorithms (Support Vector Machine, Multilayer Perceptron, and Naïve Bayes) The hybridized through a voting classification mechanism to improve detection accuracy in distributed environments. The results demonstrate the effectiveness of this hybrid model in achieving higher detection accuracy and minimizing errors compared to individual algorithms

The research also demonstrates that the model effectively detects intrusions in distributed information systems. By using a data mining-based, the hybridized intrusion detection system, it offers a feasible solution for protecting system resources like files, data, and communication. The model is recommended for use by public and private enterprises, IT companies, and e-commerce providers to enhance security. Future work will focus on implementing an ensemble approach to improve prediction accuracy by combining the decisions of the data mining algorithms.

**References**

Abdurrazaq, M. N., Bambang, R. T., & Rahardjo, B. (2014). *Distributed intrusion detection system using cooperative agent based on ant colony clustering*. 2014 IEEE International Conference on Electrical Engineering and Computer Science, 109–114.

Aladesote, O. I., Olutola, A., & Olayemi, O. (2016). Feature or attribute extraction for intrusion detection system using gain ratio and principal component analysis (PCA). *Communications on Applied Electronics (CAE), 4*(3), 1–4.

Al-Dabbagh, A. W. (2017). An intrusion detection system for cyber attacks in wireless networked control systems. *IEEE Transactions on Circuits and Systems*. https://doi.org/10.1109/TCSII.2017.2690843

Aljumah, A. (2017). Detection of distributed denial of service attacks using artificial neural networks. *International Journal of Advanced Computer Science and Applications, 8*(8), 306–318.

Anand, A., & Patel, B. (2012). An overview on intrusion detection system and types of attacks it can detect considering different protocols. *International Journal of Advanced Research in Computer Science and Software Engineering, 2*(8).

Antonio, S. D., Formicola, V., Mazzariello, C., Oliviero, F., & Romano, S. P. (2010). Performance assessment of a distributed intrusion detection system in a real network scenario. *IEEE*. https://doi.org/10.1109/ICSNC.2010.23

Cepheli, Ö., Büyükçorak, S., & Karabulut, K. G. (2016). Hybrid intrusion detection system for DDoS attacks. *Journal of Electrical and Computer Engineering*, Article ID 1075648. https://doi.org/10.1155/2016/1075648

Dhanabal, L., & Shantharajah, S. P. (2015). A study of NSL-KDD dataset for intrusion detection systems based on classification. *International Journal of Advanced Research in Computer and Communication Engineering, 4*(6), 446–452. https://doi.org/10.17148/IJARCCE.2015.4696

Du, Y., Hui-qiang, W., & Pang, Y. G. (2004). Design of a distributed intrusion detection system based on independent agents. *Proceedings of IClSlP 2004*. https://doi.org/10.1109/ICISP.2004.135

Hui, Z. (2011). A design of distributed collaborative intrusion detection model. *International Conference on Computer Science and Education (ICCSE)*, 99–101.

Ibrahim, N. M., & Zainal, A. (2018). A model for adaptive and distributed intrusion detection for cloud computing. *2018 Seventh ICT International Student Project Conference (ICT-ISPC)*. https://doi.org/10.1109/ICT-ISPC.2018.123456



*Cite this article as*:
Olojido, J. B., & Oriola, O. (2025). A fusion-based data mining model for intrusion detection in distributed environments. *FNAS Journal of Mathematical and Statistical Computing, 2*(3), *44-50*. https://doi.org/10.63561/jmsc.v2i3.856