ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a10

Al-Noor Journal for Information Technology and Cybersecurity

https://jncs.alnoor.edu.iq/

A Hybrid Intelligence Framework for Enhanced Network Intrusion Detection and Classification

[1]Karam Muhammed Mahdi Salih, [2]Shahba Ibrahim Khalil, [1]Ali Othman Mohammed, [3]Lubna Thanoon Alkahla, [1]Abdulmajeed Sulaiman

[1]Department of Computer Networks and Internet, College of Information Technology, Ninevah University, Mosul, Iraq
[2] Department of Software Engineering, College of computer sciences and Mathematics, University of Mosul, Mosul, Iraq
[3] Department of Artificial intelligence, College of Information Technology, Ninevah University, Mosul, Iraq

Article information

Article history:
Received: October, 15, 2025
Revised: November, 19, 2025
Accepted: December, 20, 2025

Keywords:
Network Intrusion Detection
Hybrid Intelligence
K-means Clustering
Genetic Algorithm (GA)
Particle Swarm Optimization (PSO)

Correspondence:

Karam Muhammed Mahdi
karam.mahdi@uoninevah.edu.iq

Abstract

Securing contemporary computer networks has become increasingly difficult as cyber-attacks continue to grow in complexity and sophistication. Conventional Intrusion Detection Systems (IDS) often fall short in recognizing emerging threats because they depend heavily on predefined attack signatures. To overcome this limitation, hybrid intelligent methodologies that merge clustering with optimization strategies have gained attention as effective tools for improving intrusion detection and classification. This study introduces an enhanced hybrid model that combines K-means clustering with both Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) to strengthen anomaly detection and misuse detection within IDS environments. The approach was tested on the KDD CUP 99 dataset, a standard benchmark in intrusion detection research. The developed Hybrid Clustering Algorithm II (HCAII) refines the detection process by lowering false-positive rates and achieving high accuracy across major attack categories, including Denial of Service (DoS), Probe, User-to-Root (U2R), and Remote-to-Local (R2L). Comparative evaluations indicate that HCAII surpasses traditional clustering and optimization methods by offering superior detection performance and more reliable classification outcomes. Overall, the proposed framework addresses critical limitations in existing IDS techniques and provides a resilient, adaptable solution capable of defending network infrastructures against continuously evolving cyber threats.

DOI: https://doi.org/10.69513/jncs.v2.i2.a10 ©Authors, 2025, Alnoor University.

1. INTRODUCTION

The rapid expansion of Internet technologies and global connectivity has significantly improved the speed and convenience of communication and information exchange. However, this growth has also increased system exposure to sophisticated attacks that exploit software weaknesses[1], system vulnerabilities, and malicious code that firewalls alone cannot detect. Intrusion detection refers to the process of monitoring system activity to identify attempts that may compromise the confidentiality, integrity, or availability of resources. Systems designed to recognize such attempts are known as Intrusion Detection Systems (IDS)[2]. A wide range of intrusion detection methods has been proposed in recent years, and many of them employ Artificial Intelligence to enhance their analytical capability[3]. In this study, we introduce a hybrid approach based on K-means clustering, Genetic Algorithms, and Particle-Swarm-Optimization. In this model, network packets are analyzed to detect abnormal activity. These approaches can be applied to both anomaly-based and misuse-based intrusion detection. IDS solutions may operate as anomaly-based or misuse-based detectors and can be deployed in host-based or network-based configurations.

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a10

Hybrid IDS integrate both detection modes to improve detection coverage. In recent years, artificial intelligence-driven techniques have been increasingly employed to enhance both anomaly and misuse detection capabilities..

Most commercial intrusion detection systems rely heavily on misuse detection, which limits their effectiveness to attacks that already have defined signatures. As a result, they often fail to identify new or evolving threats[4]. This limitation is due to several factors, including the long time required to update attack signatures, the lack of sufficient attack samples, and the unavailability of complete attack patterns. Furthermore, these systems generally lack strong forensic analysis capabilities, making it difficult to trace attack origins, analyze attacker behavior, or enhance future defense strategies. Existing hybrid IDS models suffer from fixed centroid evolution and limited swarm interaction, leading to reduced generalization and sensitivity to minority attack classes. Therefore our objectives in this study are:

- Design an adaptive hybrid clustering framework
- Improve detection under class imbalance
- Reduce false positives without supervision

This study is structured to guide the reader progressively from the problem context to the proposed solution and its validation. First, the limitations of traditional misuse-based and anomaly-based intrusion detection systems are outlined, highlighting their inability to effectively address evolving and previously unseen attacks. Next, recent advances in hybrid intelligence-based intrusion detection are reviewed to establish the current research landscape. The proposed Hybrid Clustering Algorithm II (HCAII) is then introduced, emphasizing its design rationale and methodological integration. Finally, experimental evaluation and analytical discussion are presented to demonstrate the effectiveness and limitations of the proposed approach.

Despite the extensive use of machine learning and hybrid optimization techniques in intrusion detection systems, existing IDS models still suffer from several critical limitations. In particular, many clustering-based and hybrid IDS approaches exhibit premature convergence, limited adaptability to highly imbalanced attack distributions, and unstable centroid evolution when exposed to complex and heterogeneous network traffic. These limitations reduce detection reliability, especially for low-frequency attack classes such as User-to-Root (U2R) and Remote-to-Local (R2L) attacks. Furthermore, several existing hybrid models rely on loosely coupled optimization stages, which restrict their ability to jointly optimize clustering structure and detection performance in an unsupervised manner. Consequently, there remains a clear need for a unified hybrid intrusion detection framework that enhances clustering stability, improves robustness under class imbalance, and maintains high detection accuracy without relying on labeled training data.

To address the above-mentioned research problem, this study aims to achieve the following objectives: (i) to design a unified hybrid clustering framework that integrates K-means clustering, Genetic Algorithms, and Particle Swarm Optimization within a single optimization process; (ii) to improve centroid adaptation and convergence stability for unsupervised intrusion detection in highly imbalanced network traffic (iii) to enhance detection accuracy and reduce false-alarm rates across multiple attack categories; and (iv) to evaluate the effectiveness of the proposed HCAII framework using a benchmark intrusion detection dataset through comprehensive experimental analysis.

2. Related Work

This section reviews prior intrusion detection studies with a focus on clustering-based, evolutionary, and swarm-intelligence approaches, highlighting their strengths and limitations in order to clearly position the proposed HCAII framework. Intrusion detection has been a key focus in computer network security. In 2002, M. Sabhnani and G. Serpen developed an intrusion detection system using intelligent techniques, applying nine algorithms—including neural networks, fuzzy logic, and decision trees—on the 99KDD dataset. Each algorithm excelled in detecting specific attack types, such as MLP for "Probe" and K-means for "DOS."

A. Kien et al. [5] used a genetic algorithm to optimize features for a C4.5 decision tree classifier, enhancing detection rates and reducing false alarms on the same dataset. M. Omran applied the Particle Swarm Optimization (PSO) algorithm to unsupervised image classification, outperforming traditional clustering methods like K-means. Y. Liu [6] combined the Radial Basis Function (RBF) network with PSO for network anomaly detection. The integration improved the RBF network's performance on the 99KDD dataset, demonstrating the potential of hybrid approaches in intrusion detection. Recent research has continued to explore and enhance intrusion detection systems (IDS) using genetic algorithms (GA), particle swarm optimization (PSO), and clustering techniques. A study [7] proposed a network IDS employing a combination of the Whale Optimization Algorithm (WOA) and GA for feature selection, integrated with the K-Nearest Neighbors (KNN) classifier. This approach demonstrated improved detection accuracy on the KDDCUP1999 dataset. Researchers in [8] introduced an IDS that integrates the Harris Hawks Optimization (HHO) algorithm with a Multilayer Perceptron (MLP), achieving superior performance metrics, including an accuracy rate of 93.17%, when evaluated on the KDD dataset. Another study [9] presented an intelligent IDS utilizing fuzzy logic based on the PSO algorithm, which effectively detected intelligent attacks with high stability and convergence. Additionally, research in [10] proposed a hybrid IDS combining K-Means clustering with an Artificial Neural Network (ANN) optimized by PSO, resulting in enhanced efficiency and detection capabilities when tested with the NSL-KDD dataset [11]. These studies collectively highlight the ongoing advancements in IDS methodologies, particularly through the application of GA, PSO, and clustering techniques to bolster network security. In recent years, intrusion detection research has increasingly focused on hybrid intelligence and metaheuristic-driven models to address the limitations of traditional learning-based IDS, particularly in handling high-dimensional data and class imbalance. Recent research has increasingly focused on hybrid and optimization-driven intrusion detection systems to overcome the limitations of traditional IDS models, particularly in handling high-dimensional features, class imbalance, and evolving attack patterns. Several studies have demonstrated that integrating swarm intelligence, evolutionary optimization, and machine learning techniques can significantly enhance detection accuracy and robustness. For instance, hybrid intrusion detection frameworks that combine Particle Swarm Optimization, Genetic Algorithms, and ensemble or deep learning models have reported improved detection performance and reduced false-alarm rates [12], [13], [14]. Moreover, recent optimization-based IDS approaches have shown strong adaptability to modern network environments and cyberattack scenarios [15], [16]. In addition, contemporary surveys and empirical evaluations emphasize that hybrid intelligence-based IDS models remain essential for achieving better generalization and robustness in realistic network traffic conditions [17], [18], [19]. These recent advances confirm that hybrid intrusion detection remains an active research area and provide strong motivation for the proposed HCAII framework.

3. Improving Intrusion Detection with Hybrid Clustering and Optimization

Clustering and optimization algorithms play a critical role in improving the performance of various tasks, including data classification and intrusion detection. The K-means algorithm, a popular clustering technique, uses the Euclidean distance to distribute data points and calculate centroids to divide objects into k groups. Until convergence, cluster centroids are iteratively optimized by recalculating them as the mean of their assigned data points. In a similar vein, Genetic Algorithms (GA) are a reliable method for resolving optimization and search issues.

If we let **p** denote the probability of a successful event and **(1 − p)** represent the probability of failure, then the cumulative distribution function can be expressed as follows [11]:

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - p, & 0 \le x < 1, \\ 1, & \mathrm{x} \ge 1 \end{cases} \quad ....(1)$$

In the current study, **Binary Mutation** and **Arithmetic Crossover** operators were applied. The Arithmetic Crossover operator creates new chromosomes by forming a weighted linear combination of two parent chromosomes. The resulting offspring are generated according to the following formulations [20]:

$$Offspring1 = a.parent1 + (a - 1).parent2 \quad ...(2)$$

$$Offspring2 = (a - 1).parent1 + a.parent2 \quad ...(3)$$

It starts with a population of chromosomes that have been randomly initialized; each chromosome represents a possible solution that has been encoded as bits, characters, or numbers. Genetic operators like crossover and mutation mimic natural evolution, producing better solutions through successive generations, and each chromosome's fitness is assessed using an objective function. By utilizing genetic operators to optimize cluster centroids and classifying packets according to their fitness and Euclidean distance to centroids, the Hybrid Clustering Algorithm (HCA) combines these methods to exploit the advantages of both K-means and GA for intrusion detection. The behavior of particles navigating a search space is simulated by Particle Swarm Optimization (PSO) [21], a population-based optimization technique. Based on its past performance and the best solutions found by the swarm as a whole, each particle dynamically modifies its position and velocity to converge toward ideal solutions. Due to their distinctive mechanisms, these algorithms significantly enhance data analysis and problem-solving in complex domains. Each particle is characterized by the following features [22]:

- Xi: the current particle location.
- Vi: the current particle velocity.
- Yi: the best position of the particle.

The particle's velocity at the next iteration is updated according to the following formulation:
v(i,j)(t+1) = w * v(i,j)(t) + c1 * rand1,j * (pbest(i,j)(t) – x(i,j)(t)) + c2 * rand2,j * (gbest(j)(t) – x(i,j)(t)) …. (4)

Where *i* is the particle's index and X_i =($x_{i1}, x_{i2}, .., x_{in}$) is *i*th particle's position, V_i =($v_{i1}, v_{i2}, .., v_{in}$) is *i*th particle's velocity. $pbest_i$ =($pbest_{i1}, pbest_{i2}, .., pbest_{in}$), is the *i*th particle's best position which get the best fitness function. *gbest* =($gbest_1, gbest_2, .., gbest_n$) is the best particle which gets the best fitness in the whole swarm. *w* is the weight of inertia value between {1,0} and c1, c2 represent accelerate constants c1 be responsible for controlling the spin Local Search The c2 will be responsible for controlling the spin comprehensive search. And $r_{1,j}, r_{2,j}$ random numbers between {1,0}, t is the number of iteration. Finally, the particle location x_i will be updating according to the following equation [23]:

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad ......(5)$$

4. METHODOLOGY

After examining and experimentally implementing the Particle Swarm Optimization (PSO) technique, its advantages were combined with those of the Hybrid Clustering Algorithm (HCA) and the Genetic Algorithm (GA), leading to the development of a unified approach that integrates the strengths of all three methods. In traditional HCA, the elements of each generation are treated as training samples, while the algorithm maintains only a single candidate solution that is continuously refined using GA operations such as selection and mutation. In contrast, PSO treats every particle as an independent candidate solution, each associated with its own centroid. The proposed method adopts this PSO-like principle: all individuals within the generation are regarded as candidate solutions, and each one is assigned its own centroid.

The process begins by generating an initial population with randomly selected centroids. The dataset is then assigned to clusters based on the Euclidean distance between each data point and the candidate centroids. A fitness value is computed by taking the square root of the total cluster error. After that, the probability of each individual is calculated, and the best individuals are selected using the roulette-wheel mechanism. This allows the algorithm to retain high-quality solutions from each generation while discarding inferior ones. The discarded individuals are replaced with new members produced through pairwise crossover among the top solutions. A mutation operator is also applied across clusters of elite individuals by selecting a random packet and then choosing a random feature from its 41 attributes to mutate. These evolutionary steps continue iteratively until the termination criterion is met. Experimental results show that the proposed method achieves superior performance compared to both HCA and PSO. As shown in Fig. 1, the input to this module consists of the KDD99 training file, which contains 494,020 instances, and the KDD99 testing file, which includes 311,029 instances. The testing set contains previously unseen attacks, enabling an accurate assessment of the system's detection capability.

The selected parameter values summarized in Table 1 were chosen to ensure stable convergence and efficient exploration of the search space during the optimization process. The mutation and crossover rates were configured to preserve population diversity while guiding the evolutionary search toward high-quality solutions. Similarly, the PSO inertia weight and acceleration coefficients were selected to balance global exploration and local exploitation, which is critical for optimizing clustering centroids in high-dimensional and imbalanced intrusion detection data. Overall, the parameter configuration presented in Table 1 follows commonly adopted practices in optimization-based intrusion detection studies and was found to provide consistent and reliable performance in the proposed HCAII framework.

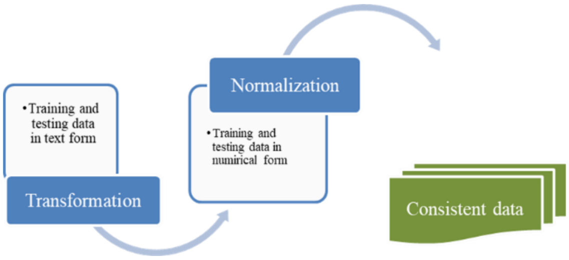
•Training and testing data in text form
Transformation
Normalization
•Training and testing data in numirical form
Consistent data

Figure (1) Dataset Preprocessing Unit

Table (1) Parameter settings for the GA–PSO components of the proposed HCAII framework

Parameter	**Value**	**Justification**
Mutation rate	0.01	Prevents premature convergence
Crossover rate	0.8	Encourages exploration
PSO (w)	0.7	Stability–exploration trade-off
c1, c2	1.5	Standard swarm dynamics

The proposed approach was evaluated using the KDD CUP99 Intrusion Detection dataset [11], which was originally derived from the DARPA 1998 cyber-security evaluation program. This dataset is organized into two main subsets: a training set and a testing set. The training portion contains 494,021 network packets, of which 97,280 are labeled as normal traffic. The testing subset includes 311,029 packets, with 60,593 designated as normal instances. Both the training and testing records comprise 41 distinct features for each packet [20]. The overall processing steps applied to this dataset are illustrated in the flowchart presented in Fig. 2.

The **KDD CUP99** dataset includes a variety of both known and previously unseen attack types, as summarized in **Table 2** [20]. These attacks are grouped into four major categories [22][23]:

- **Probe:**
  In this class, the attacker attempts to gather information about the target system prior to launching the actual intrusion.
- **Denial of Service (DoS):**
  The attacker overwhelms system resources, causing them to become overloaded or unresponsive, which results in service disruption for legitimate users connected to the network.
- **User to Root (U2R):**
  In these attacks, an intruder with normal

user privileges attempts to escalate to root or administrative privileges.

- **Remote to Local (R2L):** The attacker sends a series of packets from a remote machine in an attempt to gain unauthorized access to a local system.

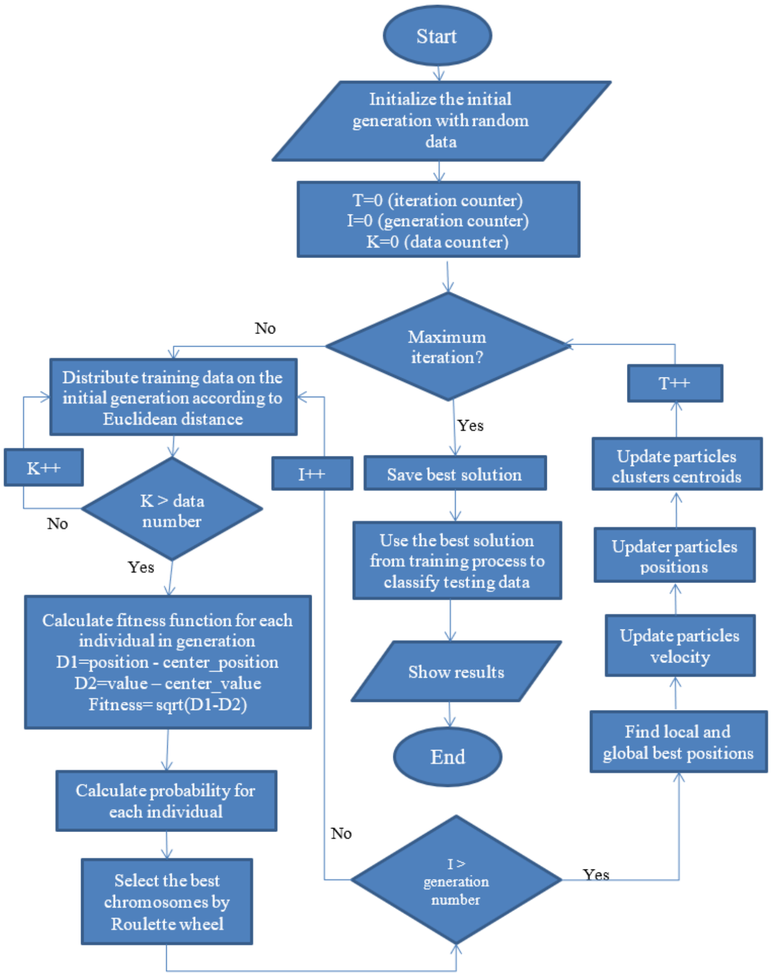
Start
Initialize the initial generation with random data
T=0 (iteration counter)
I=0 (generation counter)
K=0 (data counter)
Maximum iteration?
No
Yes
Distribute training data on the initial generation according to Euclidean distance
K++
K > data number
No
Yes
Calculate fitness function for each individual in generation
D1=position - center_position
D2=value – center_value
Fitness= sqrt(D1-D2)
Calculate probability for each individual
Select the best chromosomes by Roulette wheel
I++
No
I > generation number
Yes
Save best solution
Use the best solution from training process to classify testing data
Show results
End
T++
Update particles clusters centroids
Updater particles positions
Update particles velocity
Find local and global best positions

Figure (2) Overall architecture of the proposed HCAII framework integrating K-means, GA, and PSO mechanisms.

A comparison of these attack categories and their distribution in both the training and testing datasets is presented in **Fig. 3**.

Table (2) Examples of Known and Unknown Attacks

	DOS	PROBE	R2L	U2R
Known attack	Back, land, Neptune, Pod, smurf, teardrop	ipsweep, satan, nmap, portsweep	ftp_write, guess_passwd, warezmaster, warezclient, imap, phf, spy, multihop	Rootkit, loadmodule, buffer_overflow, perl
Unknown attack	Apache2, udpstorm, processtable, mailbomb	Saint, mscan	Named, xlock, sendmail, xsnoop, worm, snmpgeattack, snmpguess	Xterm, ps, sqlattack, httptunnel

The processing unit consists of two fundamental operations:

1. **Transformation:** The raw text-based KDD99 records are converted into numerical form. Each line in the dataset is decomposed into **41 numerical features**, representing the attributes of the corresponding KDD99 packet.
2. **Normalization:** All numerical values are normalized to fall within the range **0 to 1** to enhance convergence and stabilize the learning

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a10

process. Each feature has its own normalization formula.

For example, the *dst_Bytes* feature is normalized by applying:

Normalized dst_Bytes = (dst_Bytes – min(dst_Bytes)) / (max(dst_Bytes) – min(dst_Bytes))…(6)

This procedure is applied to all other features as well. The output of this stage is a fully numerical and normalized version of the KDD99 dataset, with consistent feature values and ready for use in the training and testing phases of the proposed system.

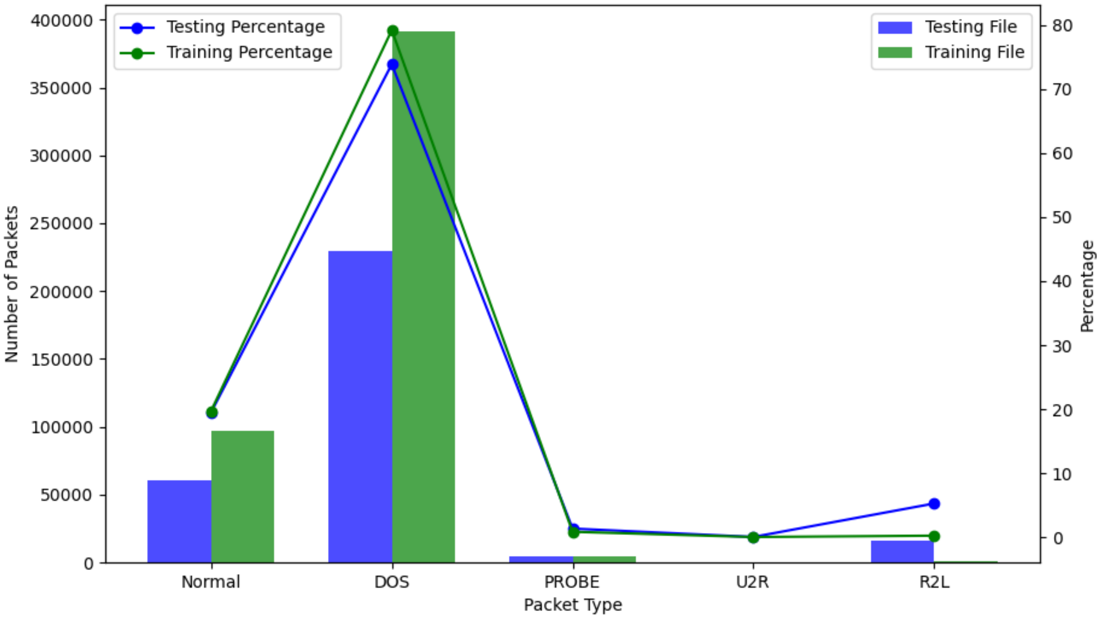

Figure (3) Distribution of attack categories in the KDDCUP99 training and testing datasets used for anomaly and misuse detection evaluation.

Table 3 show the content of training file which use to test the consistency of the system for anomaly and misuse detection.

Table (3) the content of anomaly detection training and testing file

	training file		testing file	
Packet Category	Record Percentage	Record Count	Record Percentage	Record Count
Normal Packets	19.69%	97277	19.48%	60593
DOS	79.23%	391458	73.90%	229853
PROBE	0.83%	4107	1.33%	4166
U2R	0.01%	52	0.02%	70
R2L	0.22%	1126	5.25%	16347
Total number	100%	494020	100%	311029

The training and testing datasets for anomaly detection contain a majority of DOS attack packets, making up approximately 79.23% and 73.90% of their respective files. Normal traffic accounts for around 19.69% in training and 19.48% in testing, maintaining a balanced proportion. PROBE and U2R attacks are relatively rare, with each comprising less than 1.5% of both datasets. R2L attacks show a significant disparity, making up 5.25% of the testing data but only 0.22% of the training data, which could impact model performance. Although clustering and optimization techniques such as K-means, Genetic Algorithms, and Particle Swarm Optimization have been explored individually and in various hybrid forms for intrusion detection, the novelty of the proposed HCAII framework lies in the manner in which these techniques are integrated and operationalized. Unlike conventional hybrid IDS models that apply optimization algorithms sequentially or as external tuning mechanisms, HCAII adopts a unified evolutionary–swarm strategy in which each individual is treated as an adaptive centroid-bearing solution. This design enables simultaneous centroid evolution, population-based

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a10

exploration, and swarm-guided convergence within a single optimization process. As a result, the proposed framework mitigates premature convergence, enhances cluster stability, and improves detection robustness under highly imbalanced attack distributions. Therefore, the contribution of this work is not the introduction of new algorithms, but the development of a structured hybrid optimization framework that enhances intrusion detection effectiveness and classification consistency.

5. EXPERIMENTAL RESULTS OF THE HCA-II TRAINING AND TESTING UNIT

In this study, the **confusion matrix** was adopted as the primary evaluation tool for the proposed method. The confusion matrix is a two-by-two table that summarizes the number of **True Positives (TP)**, **True Negatives (TN)**, **False Positives (FP)**, and **False Negatives (FN)** [8]. Figures **4** and **5** illustrate the confusion matrices used in the evaluation phase, as they represent one of the most important performance indicators in intrusion detection research.

In addition to the confusion matrix, the system was assessed using **Detection Rate (DR)**, **Precision**, and **Accuracy**. Accuracy expresses the proportion of correctly identified instances and reflects the overall reliability of the model [10][8]. These metrics are defined as follows:

$$\text{Detection Rate} = \frac{\text{TP}}{\text{TP} + \text{TN}} \text{X } 100 \quad .....(7)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad ....(8)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad .....(9)$$

After using the KDD99 training file to train HCAII The results obtained from the training phase are summarized in the table below. Then the best solution that resulted from the training process was entered into the testing process with testing data and it has been getting the results described in Table 4.

Table (4) Results of HCA-II Training and Testing for Anomaly Detection

DR – Normal	DR – R2L	DR – U2R	DR – PROBE	DR – DOS	Exec. Time
96.21%	97.91%	99.62%	99.71%	96.01%	5s

Table (5) Results of HCA-II Testing Process for Misuse-Detection

	FP	FN	TP	TN	FNR	FPR	TNR	TPR	Precision	Accuracy	Exec. Time
Training	97%	15,454	0	412,197	81,823	0%	16%	84%	100%	96%	14:58
Testing	98%	0	4,837	245,599	65,430	2%	0%	100%	98%	98%	1:00

Using the same methodology, the system was then applied to classify attacks into their specific categories: DOS, PROBE, U2R, and R2L. When five clusters were used, the classification performance declined significantly. However, using two clusters yielded excellent results. For example, when detecting DOS attacks, the first cluster was designated for DOS and the second cluster for all other categories—including normal traffic and other attack types.

The experimental results demonstrate that the proposed HCAII framework achieves consistently high detection rates across multiple attack categories while maintaining a low false-alarm rate. In particular, the model shows strong performance in detecting high-frequency attack types such as DoS and Probe attacks, which can be attributed to the adaptive centroid optimization mechanism and the joint evolutionary–swarm search strategy. The integration of GA and PSO enables effective exploration of the feature space while preserving convergence stability, leading to improved clustering quality and classification consistency. These results indicate that HCAII effectively balances detection accuracy and robustness, even when operating in an unsupervised learning environment. One limitation of the proposed method is reflected in the false-alarm rate, particularly for U2R and R2L attacks. This limitation is reflected in the results summarized in Table 5 following the training phase. The best solutions obtained from the training process were used to evaluate misuse detection on the KDD99 testing data

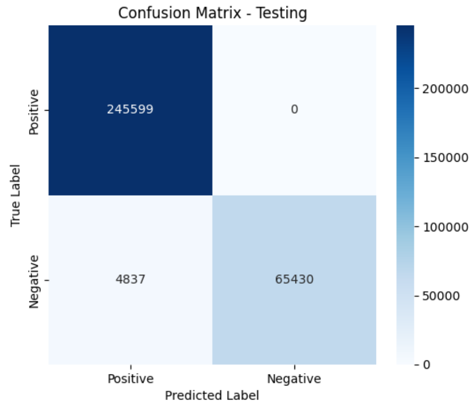

Figure 4 Confusion matrix obtained from the testing phase

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a10

As shown in fig.4, the confusion matrix illustrates the performance of the classifier on the testing dataset, revealing a high level of accuracy in predicting Positive and Negative classes. The model correctly identified 245,599 instances as Positive, as no False Negatives were recorded. This indicates that all actual Positive instances were accurately classified, minimizing the risk of overlooking critical cases. On the other side, the model produced 4,837 False Positives, where Negative instances were categorized as Positive faulty. Despite this, the high count of True Negatives (65,430) indicates robust precision in distinguishing Negative cases. More detailed confusion matrix is introduced in fig.5:.

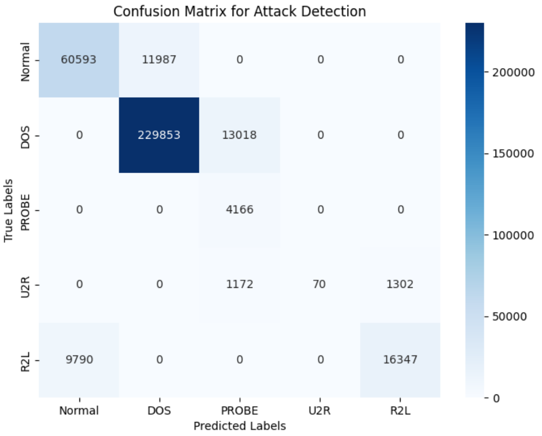

Figure 5 Detailed confusion matrix showing classification performance across all attack categories in the testing dataset

For further comparison, Fig. 6 shows the difference between the predicted and actual packet counts for each attack type:

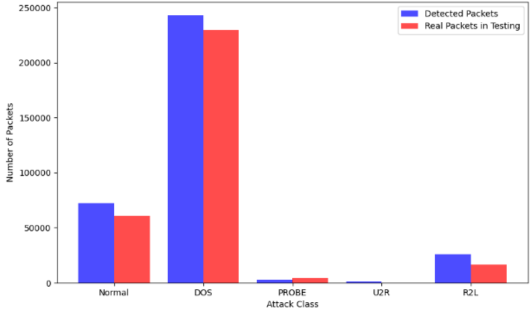

Figure 6 Comparison between detected and actual packet counts for each attack category in the testing dataset.

As shown in Fig. 6, the detection rate differs across attack categories; the detection percentages are illustrated in Fig. 7.:

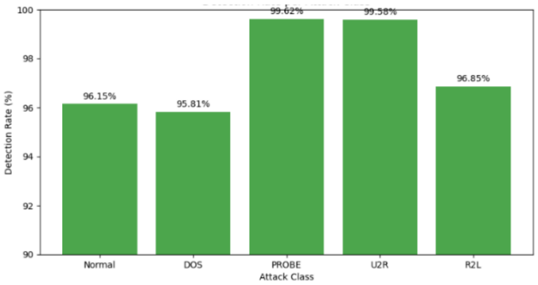

Figure 7 Detection rate achieved by the proposed HCAII model for each attack category.

To further assess the effectiveness of the proposed approach, the obtained results were compared with those reported in previous intrusion detection studies that employed the KDDCUP99 dataset under similar experimental settings. Compared to conventional K-means-based IDS and standalone GA- or PSO-driven models, HCAII achieves higher detection rates and improved accuracy while maintaining competitive execution time. In particular, prior studies typically report detection rates ranging between 90% and 95% for hybrid or optimization-based IDS models, whereas the proposed HCAII framework attains detection rates exceeding 96% across major attack categories. This quantitative improvement underscores the advantage of the unified evolutionary–swarm optimization strategy adopted in this work.

6. CONCLUSION

The objective of this study was to address key limitations of existing clustering-based intrusion detection systems, particularly premature convergence, unstable centroid adaptation, and reduced detection reliability under imbalanced attack distributions. To this end, a hybrid intrusion detection framework, referred to as HCAII, was proposed by integrating K-means clustering with Genetic Algorithms and Particle Swarm Optimization within a unified evolutionary–swarm optimization process.

The experimental results demonstrate that the proposed HCAII framework achieves high detection accuracy across major attack categories while maintaining competitive execution time. The adaptive centroid optimization mechanism enables improved clustering stability and contributes to consistent detection performance in an unsupervised learning environment. These findings confirm that the proposed framework effectively meets the research objectives outlined in this study, particularly in enhancing detection robustness and classification consistency.

Despite the promising results, The evaluation was conducted exclusively on the KDDCUP99 dataset, In addition, although the hybrid optimization strategy improves sensitivity to low-frequency attacks, false-alarm rates for minority classes such as U2R and R2L remain a challenge. These limitations highlight the

ISSN: 3078-5367 **DOI:** https://doi.org/10.69513/jncs.v2.i2.a10

inherent difficulty of unsupervised intrusion detection in highly imbalanced datasets.

Future work will focus on extending the proposed HCAII framework to more recent intrusion detection datasets, including NSL-KDD, UNSW-NB15, and CIC-IDS2017. Furthermore, additional optimization strategies and adaptive parameter tuning mechanisms will be explored to further enhance detection performance for rare attack categories and improve scalability in real-world network environments.

REFERENCES

[1] Huang P., Yang Ch., Basilico N., Nam Ahn T., "Design and Implementation of a Distributed Early Warning System Combined with Intrusion Detection System and Honeypot," *Proceedings of the International Conference on Convergence and Hybrid Information Technology*, 2009.

[2] K. M. M. Salih and N. B. Ibrahim, "Enhancing IoT forensics through deep learning: Investigating cyber-attacks and analyzing big data for improved security measures," *2023 4th International Conference on Big Data Analytics and Practices (IBDAP)*, Bangkok, Thailand, Aug. 2023, pp. 1–6, doi: 10.1109/IBDAP58581.2023.10271950.

[3] Hussein M. K., ALkahla L. T., Alqassab A., "Hyperspectral Image Classification Using Hybrid Swarm Feature Selection and Ensemble Classifier," *Ingénierie des Systèmes d'Information*, vol. 29, no. 6, pp. 2367–2375, 2024.

[4] K. M. M. Salih and N. B. Ibraheem, "Alpha-FedAvg: Safeguarding Privacy and Enhancing Forensic Analysis in Federated Learning on Edge Devices," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 1391–1403, 2024.

[5] Sabhnani M., Serpen G., "Application of machine learning algorithms to KDD intrusion detection dataset within misuse detection context," *Proceedings of the International Conference on Machine Learning: Models, Technologies, and Applications (MLMTA)*, Las Vegas, NV, USA, 2003, pp. 209–215.

[6] G. Stein, B. Chen, A. S. Wu, and K. A. Hua, "Decision tree classifier for network intrusion detection with GA-based feature selection," *Proceedings of the 43rd Annual Southeast Regional Conference – Volume 2*, Kennesaw, GA, USA, 2005, pp. 136–141, doi: 10.1145/1167253.1167288.

[7] M. Omran, A. P. Engelbrecht, and A. Salman, "Particle swarm optimization method for image clustering," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 19, no. 3, pp. 297–321, 2005, doi: 10.1142/S0218001405003987.

[8] Y. Liu, Y. Li, and L. M. Ni, "A hybrid anomaly detection technique based on PCA and support vector machines," *Proceedings of the 2006 International Conference on Advances in Intelligent Computing*, Kunming, China, 2006, pp. 1130–1135, doi: 10.1007/11816171_144.

[9] A. Al-Yaseen, Z. A. Othman, and M. Z. A. Nazri, "Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system," *Expert Systems with Applications*, vol. 67, pp. 296–303, 2017, doi: 10.1016/j.eswa.2016.09.041.

[10] A. M. Ambusaidi, X. He, and Z. Tang, "Building an intrusion detection system using a filter-based feature selection algorithm," *IEEE Transactions on Computers*, vol. 65, no. 10, pp. 2986–2998, Oct. 2016, doi: 10.1109/TC.2016.2519914.

[11] S. S. Chouhan, S. Kaushik, and S. K. Gupta, "An intelligent intrusion detection system using fuzzy logic and particle swarm optimization," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 4297–4308, 2019, doi: 10.3233/JIFS-169580.

[12] A. Aljanabi, A. A. Abd, and M. A. Mohammed, "A hybrid intrusion detection system based on swarm intelligence and machine learning for network security," *Expert Systems with Applications*, vol. 209, 2023, Art. no. 118259.

[13] H. Liu, B. Lang, M. Liu, and Y. Yan, "PSO-optimized intrusion detection for imbalanced network traffic," *Applied Soft Computing*, vol. 132, 2023, Art. no. 109876.

[14] M. A. Ferrag, L. Shu, O. Friha, and X. Yang, "Deep learning and hybrid intelligence for intrusion detection systems: Recent advances and challenges," *IEEE Network*, vol. 37, no. 2, pp. 48–55, 2023.

[15] A. K. Shamsaldin, A. H. Ali, and R. S. Khairy, "An efficient hybrid intrusion detection system using genetic algorithms and ensemble learning for cyberattack detection," *Computers & Security*, vol. 131, 2024, Art. no. 103189.

[16] S. A. Albahri, J. K. Alwan, and Z. A. Othman, "Metaheuristic-based hybrid intrusion detection systems for modern network environments," *Future Generation Computer Systems*, vol. 148, pp. 215–228, 2024.

[17] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: Techniques, datasets, and future directions," *Cybersecurity*, vol. 5, no. 1, 2022.

[18] M. Al-Hawawreh, N. Sitnikova, and M. Slay, "Evaluating intrusion detection systems against contemporary cyber threats," *Future Generation Computer Systems*, vol. 137, pp. 188–203, 2022.

[19] K. M. M. Salih, *FedForensics-IoT: A non-IID Federated Learning Framework for Forensic Investigation of IoT Attacks on Edge Devices, AIP Conference Proceedings*, vol. 3211, no. 1, Art. no. 010001, May 2025.

[20] Riikka P., Aki S., "Real-coded genetic algorithms and nonlinear parameter identification," *University of Oulu Control Engineering Laboratory Report A*, no. 34, Apr. 2008.

[21] Thanoon ALkahla L., Salahaldeen Alneamy J., "Improving the ability of persons identification in video files based on hybrid intelligence techniques," *Next Generation of Internet of Things: Proceedings of ICNGIoT 2022*, Springer Nature Singapore, pp. 509–518, 2022.

[22] Soares C., Gilbert J., "Predicting cross-country results using feature selection and evolutionary computation," *ACM Conference Proceedings*, Portland, OR, USA, 2009.

[23] Wang H., Wu Z., Liu Y., "Space transformation search: A new evolutionary technique," *ACM Conference Proceedings*, Shanghai, China, 2009.