International Journal of Basic and Applied Sciences, 14 (5) (2025) 282-288

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/gcjafw37
Research paper

PCA-Based Model to Enhance The Performance of Data Clustering Using A Metaheuristic Algorithm

Shailendra Singh Yadav *, Kamal Sutaria

Department of Computer Science & Engineering, Parul Institute of Engineering & Technology, Faculty of Engineering and Technology, Parul University,Waghodia Vadodara, India
**Corresponding author E-mail: 2123004136009@paruluniversity.ac.in*

Received: July 15, 2025, Accepted: August 25, 2025, Published: September 9, 2025

Abstract

This paper explores the enhancement of clustering techniques through the integration of preprocessing methods and nature-inspired algorithms. Achieving perfection in making clusters from the given dataset is a crucial task. Many algorithms have been proposed in the literature to make useful clusters of raw data. But due to the day-by-day changing properties and complexity of data, like multidimensional data values, increasing size, and many more parameters, they still require lots of enhancements in clustering techniques. By addressing the limitations of traditional clustering approaches, the research emphasizes the importance of pre-processing methodologies like PCA and KNN to improve clustering outcomes. The limitation of handling high-dimensional data by an optimizer algorithm is addressed by PCA, which improves feature examination of the dataset at a low level for further processes. Additionally, it examines various nature-inspired algorithms that can be applied to clustering tasks, demonstrating their efficacy in optimizing and refining clustering results. Overall, the study presents a framework combining KNN+PCA+EGWO that combines these elements to achieve superior clustering performance, i.e., increasing the result performance by 45-50 % and the effects of preprocessing steps are also shown with low error and other great parameters present in the paper.

Keywords: *Clustering; Preprocessing; Enhanced Grey Wolf Optimizer; PCA, Homogeneity Level.*

1. Introduction

Optimality of the result is one of the major issues in today's research scenario, and applying nature-inspired algorithms is one of the most popular and advanced ways to achieve it. Many algorithms, like Genetic Algorithm [1], Particle Swarm Optimization [2], Differential Evolution [3], Ant Colony Optimization [4], and many more, have evolved so far. Each of these algorithms has a different set of working principles. Researchers were applying these metaheuristic approaches to optimize their results over existing real-world problems. But none of them were found to be perfect for one type of problem. Every new paper tries to enhance the given result with newly invented approaches. So, the evolution of new nature-inspired algorithms is still going on, to enhance the results of different research areas, which include clustering [5], classification [6], robotics [7], and many more.

One of the widely classified research areas is clustering. Clustering is also treated as a multi-objective optimization problem. It is an existing real-world problem that still needs researchers' attention to make effective clusters from raw data. Well, the formation of perfect clusters helps real-world problems in different fields, and it can be understood by studying the following importance of clustering, such as:

- Organizing data to show internal structure
- Partitioning the data to make segments
- Making of a knowledge discovery system

These things help in the field of medical diagnosis, the educational sector, in making business strategies for decision-making, and many more.

Therefore, several classical clustering algorithms are presented in the literature. Among them, K-means [8-9] is one of the classical clustering algorithms, known for its simplicity, but lagging just because of a few problems, like initialization with value k, local optima trapping, and balancing the cluster size, whereas EGWO addresses this issue by optimize the initial value of dataset with population (swarm intelligence) approach to achieve optimal number of clusters.

In this paper, we applied the recently developed algorithm EGWO by Shail. G. et.al.[10] to achieve clustering of data for some well-defined datasets like Dermatology, Glass, etc. It is one of the recently introduced population-based metaheuristic approaches.

Some advantages of EGWO over existing nature-inspired algorithms will be expressed by the following application to which it has been applied: the algorithm has a good convergence rate and shows better results for standard benchmark multimodal mathematical functions, and it has also been applied to other engineering problems. The detailed explanation of EGWO has been done in a further section (3).

The paper is organized in the following manner: Section 2 consists of a literature survey, Section 3 describes a key algorithm, i.e., Enhancing the grey wolf optimizer, Section 4 deals with the result analysis and proposed model, and lastly, Section 5 concludes the work.

2. Literature Survey

Clustering means forming groups with similar properties or features. It is an unsupervised learning process, which can be said to be learning from observations. Clustering aims to group the elements such that they acquire maximum intra-cluster and minimum inter-cluster distance. Clusters can be formed by applying any one of the following methods: partitioning *method*, *Hierarchical method, Density-based method, Grid-based method, etc.*

So, here in this paper, the partitioning method has been used to evaluate the performance of the EGWO algorithm. Optimization algorithms are gaining interest throughout the research world, as researchers are always trying to find a better solution than the previous one, commonly known as the fittest value. Optimization algorithms can be categorized into population-based and non-population-based algorithms, but both have different ways of working. In a population-based algorithm, the social behaviour of other species affects the nature of the algorithm, whereas in a non-population-based algorithm, there is no social interaction between other species. To understand the concept, some GWO-based papers have been discussed below to understand the importance of the algorithm. M. Seyedeli et al. [11-12] discovered the grey wolf optimizer (GWO), which mimics the leadership hierarchy with the help of four types of wolves, such as alpha, beta, delta, and omega.

In this [13] paper, the author applied the grey wolf optimizer (GWO) for TCP, referred to as TCP-GWO. The issue of splitting text documents into homogeneous clusters using GWO is sufficiently precise and functional.

Another application of GWO [14] is implemented for topology shaping, and its shaping involves two aspects, i.e., the optimal mapping and the optimization of shaping position.

The author proposed [15] a modified version of GWO for Data clustering analysis to optimize results and compare with existing methods in many fields, such as pattern recognition and image processing.

In [16], this paper, the author proposes a strategy inspired by the Grey Wolf optimization algorithm to select features.

In this paper [17], the author uses Meta-heuristic algorithms (MHA), which can track the maximum power point in a power-voltage (P-V) curve with multiple peaks by improving the performance of the existing grey wolf optimizer, named as MGWO. In this paper [18], an improved binary grey wolf optimizer (IBGWO) is proposed to resolve the problem of dependent task scheduling in edge computing.

In this [19] paper, the author proposes an improved version of the binary grey-wolf optimizer (BGWO) to resolve the problem of feature selection (FS) related to high-dimensional data, duplicate data, noise, and irrelevant data, which helps in machine learning to improve classification/ clustering accuracy with less training time.

3. Enhance The Grey Wolf Optimizer

The EGWO is a population-based metaheuristic approach, which is an enhanced version proposed by the author [10]. The algorithm "simulates the social leadership and hunting behaviour of grey wolves in nature". The basic steps used by a grey wolf for hunting were implemented, and they are searching for prey, encircling prey, and attacking prey.

Grey wolves hunt by forming groups (population), so the population is categorized into four groups: Alpha (α), *Beta* (β), Delta (δ), and Omega (ω)

The fittest solution is achieved by the position of alpha (α) wolve*s,* the second fittest solution is given by beta (β) and delta (δ), and the rest of the solution is taken by omega (ω) wolves, as these wolves only follow other wolves. The mathematical model of encircling prey and the concept of position updating are shown below, in equations (1) and (2).

$$\vec{X}(t+1) = \vec{X}p - \vec{A}.\vec{D} \tag{1}$$

$$\vec{D} = \left|\vec{C}.\vec{X}_p(t) - \vec{X}\right| \tag{2}$$

The values of the vector $\vec{A}$ & $\vec{C}$, is calculated by equations (3) and (4), which were random parameters that allow wolves to move, to and fro, around the prey.

$$\vec{A} = 2\vec{a}\vec{r}_1 - \vec{a} \tag{3}$$

$$\vec{C} = 2.\vec{r}_2 \tag{4}$$

The parameter t indicates the current iteration, the value of $\vec{a}$ will linearly decrease from 2 to 0, $\vec{r}_1$, $\vec{r}_2$ is a random vector between [0, 1].

During the choice of fitness value, the first three best solutions were considered as α, β, and δ, respectively. The other wolves, i.e., ω, can update their position with respect to α, β, and δ. The mathematical model proposed for updating the positions of ω wolves is shown by the equations (5), (6), and (7), [6] which define the step size of the ω wolf toward α, β, and δ, respectively.

$$\vec{D}_\alpha = \left|\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}\right| \tag{5}$$

$$\vec{D}_\beta = \left|\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}\right| \tag{6}$$

$$\vec{D}_{\delta} = \left|\vec{C}_3 . \vec{X}_{\delta} - \vec{X}\right| \quad (7)$$

The final position of the ω wolves is shown in equations (8), (9), (10), and (11).

$$\vec{X}_1 = \vec{X}_{\alpha} - \vec{A}_1 .(\vec{D}_{\alpha}) \quad (8)$$

$$\vec{X}_2 = \vec{X}_{\beta} - \vec{A}_2 .(\vec{D}_{\beta}) \quad (9)$$

$$\vec{X}_3 = \vec{X}_{\delta} - \vec{A}_3 .(\vec{D}_{\delta}) \quad (10)$$

Equations (11), (12), and (13) have been incorporated into the conventional GWO framework in the Enhanced version of the Grey Wolf Optimizer (GWO), which enhances the exploration capabilities of the original algorithm. By skilfully striking a balance between exploration and exploitation, this integration strengthens the algorithm's resilience. To greatly increase the diversity among the population members during the exploration phase, the author has incorporated a differential perturbation equation into the classical GWO algorithm. Equations (11), (12), (13), and (14) provide the mathematical expressions for the suggested changes.

Z1=xi+f*(X1-Xrand1) (11)

Z2=xi+f*(X2-Xrand2) (12)

Z3=xi+f*(X3-Xrand3) (13)

Where the ith omega wolf, represented by xi, is represented by the modified alpha, beta, and delta wolves, Z1, Z2, and Z3. The three randomly selected omega wolves (xrand1, xrand2, and xrand3) help the leader wolves locate more varied solutions around the most promising region of the search space. This increases the exploration capability of the algorithm. Here, the traditional GWO algorithm is replaced with Eq. (17) in the EGWO approach.

$$X_i(t+1)=(Z_1+Z_2+Z_3)/3 \quad (14)$$

The detailed description of all variables used in the GWO algorithm is listed in Table 1.

Table 1: Meaning of Variables Used in EGWO

S. No.	Symbol	Meaning
1	$\vec{X}(t+1)$	Final position of the current solution
2	$\vec{X}_1, \vec{X}_2, \vec{X}_3$	Final position of the ω wolves.
3	$\vec{D}$	Distance vector
4	$\vec{A}, \vec{C}$	Coefficient vectors
5	$\vec{r}_1, \vec{r}_2$	Random vectors in [0,1]
6	$\vec{a}$	Component linearly decreased from 2 to 0
7	$\vec{X}_p$	Current position vector of the prey
8	t	Current iteration
9	$\vec{D}_{\alpha}, \vec{D}_{\beta}, \vec{D}_{\delta}$	Distance between the current solution and alpha, beta, and delta
10	$\vec{X}_{\alpha}$	Best search agent position
11	$\vec{C}_1, \vec{C}_2, \vec{C}_3$	Random vectors
12	$\vec{X}$	Current position vector of a grey wolf
13	$\vec{X}_{\beta}$	Second-best search agent position
14	$\vec{X}_{\delta}$	Third-best search agent position
15	$\vec{A}_1, \vec{A}_2, \vec{A}_3$	Random vectors

4. Proposed Model and Result Analysis

This section describes some pre-defined datasets for data clustering, with some description of the SSE Homogeneity level, Intra-cluster distance and Inter-cluster distance, followed by result analysis. A detailed working flow is depicted in Fig. 1. With the help of an enhanced version of GWO integrated with KNN+PCA, a model in Fig. 1 has been proposed in the paper. Various papers use the concept to make effective clusters. A similar approach has been used in this paper to make effective clusters from raw data.

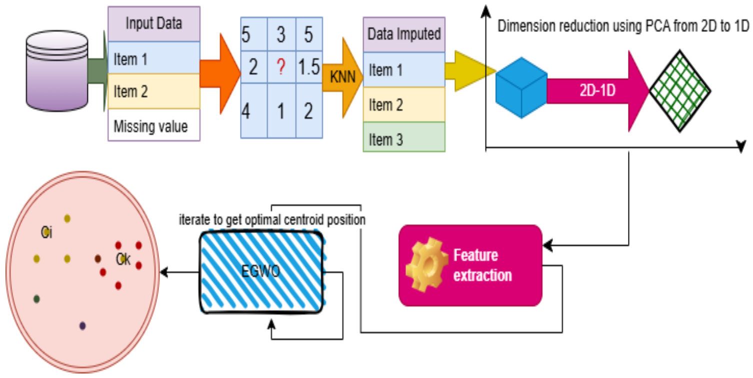

Fig. 1: Proposed Model EGWO+PCA+KNN.

4.1. Preprocessing and Its Importance

Preprocessing [20-21] is one of the basic steps that can be embedded in the model, to be applied over raw data to make it effective for further steps. In the literature, several methods have been given, but for missing value imputation, KNN [22] has been used here, and for dimensional reduction, PCA [23] has been implemented, because it is one of the finest algorithms to get optimal results. Handcrafted features are extracted and implemented in the work by using a feature extraction tool [24]. To show the importance of preprocessing, an experimental result is given in Table 2, which shows how the result of the dermatology dataset values is enhanced by 30%-40% by applying preprocessing steps in each model.
The computational cost of the model is much less compared to existing ones and takes much less time to get an optimal result at standard parameters.

Table 2: Results with and without Preprocessing Step

PARAMETER	Without preprocessing		With preprocessing	
	EGWO	PSO	EGWO	PSO
SSE	0.85107	1.8282	0.6451	1.3168
Homogeneity Level	0.73453	0.72514	0.92938	0.8104
Inter Cluster Distance	10.3017	4.0178	10.8864	4.9597
Intra-Cluster Distance	0.079545	0.49154	0.013588	0.26887

4.2. Data Source

To experiment, a well-defined dataset from the repository [25] has been taken. The detailed description of the dataset is given in Table 3. All data sets considered in our experiments are standard databases normally used by researchers to apply their algorithms for clustering purposes.
The detailed description of these datasets is given below:

- Dermatology dataset contains 34 attributes of a skin disease for experimental purposes.
- Glass data set having 214 instances with 6 classes of glass containing 9 features.
- The thyroid is a diagnosis of thyroid and is used to classify 3 classes of thyroid function as
- Over function
- Normal function
- Under function

The data set contains 215 patterns.

- In teaching assistant evaluation, the data consist of evaluations of teaching performance with three attributes("low", "medium", and "high")

Table 3: Characteristics for Clustering Data Set

S. No	Dataset	Instances	Clusters	Features
1.	Dermatology	366	6	34
2.	Glass	214	6	9
3.	Thyroid	215	3	5
4.	Teaching Assistant Dataset	151	3	5

4.3. Sum of Squared Error

Cluster similarity is calculated as the mean value of an object in a cluster, viewed as the centroid or center of gravity for a given cluster. One of the well-known methods adopted is SSE [26], i.e., sum of squared error, to measure the quality of clustering, which is shown below in equation (15)

$$SSE = \sum_{i=1}^{k} \sum_{p \in C_i} \|p - m_i\|^2 \tag{15}$$

Where p represents the point for a given object, m_i is the mean value of the cluster C_i, and k represents the number of clusters
In this Euclidean distance of the closest centroid has been measured, and calculates SSE. The error said above is basically the nearest distance to the cluster centroid and is targeted to minimize the error rate.

4.4. Inter-cluster distance

The inter-cluster distance [27] based on the centroids can be defined as

$$Inter = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} D(ci, cj) \tag{16}$$

Where D(ci,cj,) is the distance between the centroid ci and cj of k clusters.

4.5. Intra-Cluster Distance

$$Intra = \sum_{i=1}^{k} \sum_{j=1}^{q} D(xj, ci) \tag{17}$$

Where D (xi, ci) is the distance between x_j and centroid ci for all k clusters and q items for x_j data items..

4.6. Homogeneity Level

It is bounded between 0 and 1, with low values indicating a low homogeneity [28] between true and predicted entropy.

$$1 - \frac{H(C \mid K)}{H(C)} \tag{18}$$

4.7. Parameter Setting

It is one of the major issues in computing any optimization technique. , following are the parameters that were taken to compute the PSO and EGWO algorithms:
PSO, we use 30 particles, and set w = 0.1298 and $C_1 = C_2 = 1.4968$. The population size is taken as 30 for both algorithms, with other default parameters.

4.8. Results Analysis

Each algorithm has been performed several times to make an experimental judgment. The mean value of 100 iterations is used to compare the outcomes by taking an average of five times. Similarly, while calculating the distance between data points, the Euclidean distance is considered.
Table 4 displays the results of the sum of square errors, and the error rate by the EGWO communication is less than PSO by 50 percent in some cases, even though the standard deviation of the proposed model is less in comparison with PSO. Hence, it proves the stability of the optimization algorithm also.
Similarly, by calculating Inter-cluster distance and Intra-cluster distance in Table 4, it shows that the results for EGWO are 60%-70% more effective than PSO.
In the case of the homogeneity level of data, where it should be higher, it was shown that EGWO performs better than PSO in most cases.

Table 4: Sum of Squared Error Inter Cluster Distance Intra Cluster Distance Homogeneity Level

Algorithm→ Data sets	Values	EGWO SSE	PSO	EGWO Inter Cluster Distance	PSO	EGWO Intra-Cluster Distance	PSO	EGWO Homogeneity Level	PSO
Dermatology	Avg.	*0.51501	1.47058	10.60082	4.9106	0.064025	0.240589	0.88689	0.84137
	Min.	0.37152	1.0751	10.0502	4.1174	0.039359	0.020442	0.84376	0.7367
	Max	0.77836	1.9155	10.9578	0.334564	0.093335	0.4711	0.95183	0.95522
	std	0.171551	0.311515	0.385292	4.55568	0.026156	0.202453	0.042791	0.079
Glass	Avg.	*0.733052	1.22285	10.50282	4.3624	0.04378	0.143526	0.847106	0.7951
	Min.	0.25873	0.9533	10.0303	4.0775	0.016391	0.033602	0.77291	0.71862
	Max	1.0336	1.5879	10.7418	4.5957	0.085172	0.25289	0.94834	0.028384
	std	0.303945	0.267007	0.30612	0.18522	0.027496	0.097707	0.073416	0.765374
Thyroid	Avg.	*0.460718	1.28784	10.47566	4.38588	0.071559	0.270086	0.937802	0.787176
	Min.	0.14915	1.0543	10.1516	4.1946	0.025953	0.10621	0.85607	0.68543
	Max	0.65317	1.4535	10.832	4.7178	0.10462	0.4166	0.937802	0.83148
	std	0.199039	0.174979	0.258562	0.209974	0.031517	0.141658	0.06646	0.063722
Teaching assistant evaluation	Avg.	*0.505589	1.34476	9.367399	4.061464	0.021471	0.287833	0.750591	0.715332
	Min.	0.25355	0.95661	10.5253	4.1539	0	0.21384	0.79362	0.74423
	Max	0.83678	1.9046	10.8099	4.9623	0.049674	0.42882	0.89189	0.86703
	std	0.274051	0.39606	0.13009	4.1539	0.021612	0.091574	0.042866	0.051593

The sum of squares errors is one of the traditional ways to compute the result of data clustering in terms of its compaction value. Within-cluster distance means the lower the value, the better the DC components. Here, the given model increases the optimal capacity of the Enhanced Grey wolf optimizer algorithm, depicted by Figure 2, i.e., statistical representation of the mean value of SSE, which shows that the error rate in the EGWO-based model is 50 percent less than the other optimizer algorithms.

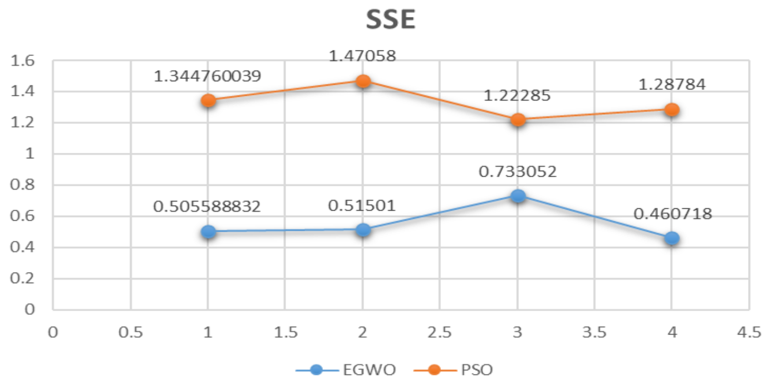

Fig. 2: Statistics of the Mean Value of the Sum of Squared Error.

Inter-cluster distance is one of the measurement techniques to show the distance between different cluster centroids, i.e. higher the distance between the centroids of clusters, the better the clustering result. It is shown in Fig. 3, which depicts the statistical representation of the mean value of inter-cluster distance, which shows that the centroid value calculated by the EGWO-based model is 45 percent more than the other optimizer algorithms.

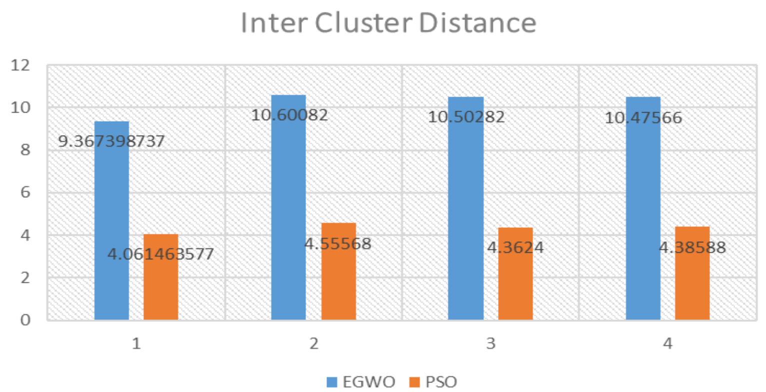

Fig. 3: Mean Value of Inter-Cluster Distance.

Intra-cluster distance makes the cluster stronger as more similar levels of data are accumulated at one centroid point, which has been taken as the alpha position of wolves, so the lower the distance between the centroid and cluster points, the better the clustering result in Figure 4. It depicts the statistical representation of the mean value of intra-cluster distance, which shows that the centroid value calculated by the EGWO-based model is 30 percent more than the other optimizer algorithms. Also shows the compactness in clustering formation.

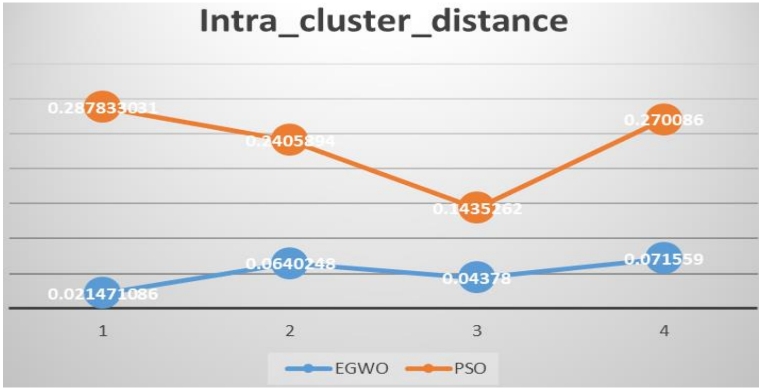

Fig. 4: Mean Value of Intra-Cluster Distance.

5. Conclusion

Several approaches have been applied so far for data clustering, but this field still needs some enhancement. The paper used a recently invented metaheuristic approach, i.e., Enhanced Grey Wolf Optimizer, for clustering, and the results show that the integrated model KNN+PCA+EGWO performs better than existing optimization techniques, i.e., Particle Swarm Optimization. The paper also focuses on the importance of preprocessing steps by showing the results with and without preprocessing steps. The paper also checks the quality of clustering, with the help of, which shows an excellent result, for a predefined dataset for data clustering, for Sum of Square Error,

Homogeneity level, Intra-cluster distance, and Inter-cluster distance. The result shows that new metaheuristic approaches perform well and give healthy competition to existing optimization techniques.
In the future, we can apply the proposed model to various other applications and over more datasets, by increasing the volume and instances, including big data based on the Hadoop model, to get experimental judgment of the model.

References

[1] Raman R, Kumar V, Pillai BG, Rabadiya D, Patre S, Meenakshi R. The impact of enhancing the k-means algorithm through genetic algorithm optimization on high-dimensional data clustering outcomes. In: 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS). IEEE; 2024. p. 1–5. https://doi.org/10.1109/ICKECS61492.2024.10617268.

[2] Zhang X, Lin Q, Mao W, Liu S, Dou Z, Liu G. Hybrid Particle Swarm and Grey Wolf Optimizer and its application to clustering optimization. Appl Soft Comput. 2021;101(107061):107061. Available from: https://doi.org/10.1016/j.asoc.2020.107061.

[3] Song X, Zhang X, Zhao M. Improved artificial bee colony algorithm embedded with differential evolution operator. In: 2024 9th International Conference on Electronic Technology and Information Science (ICETIS). IEEE, 2024. https://doi.org/10.1109/ICETIS61828.2024.10593697.

[4] Guo H, Liu Q, Dang Z. Optimization of a two-dimensional path optimization algorithm based on Dijkstra ant colony optimization algorithm. In: 2024 2nd International Conference on Signal Processing and Intelligent Computing (SPIC) IEEE; 2024. p. 399–402. https://doi.org/10.1109/SPIC62469.2024.10691583.

[5] Yue J, Arimuzha. Big data optimization clustering algorithm for power grid CPS based on PSO. In: 2024 Second International Conference on Data Science and Information System (ICDSIS). IEEE, 2024. https://doi.org/10.1109/ICDSIS61070.2024.10594360.

[6] Suganya D, Sugumar R. PSO-optimized CNN for feature extraction and accurate classification of satellite images using machine learning. In: 2024 International Conference on Computing and Data Science (ICCDS). IEEE; 2024. https://doi.org/10.1109/ICCDS60734.2024.10560453.

[7] Liang Z, Wang Z, Wang Y, Yan X. A study of trajectory planning of medical robots based on particle swarm optimization (PSO) algorithm. In: 2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI). IEEE; 2024. https://doi.org/10.1109/ICETCI61221.2024.10594551.

[8] Ikotun AM, Ezugwu AE, Abualigah L, Abuhaija B, Heming J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. Inf Sci (Ny)]. 2023;622:178–210. Available from: https://doi.org/10.1016/j.ins.2022.11.139.

[9] Wang C, Yang N, Xu W, Wang J, Sun J, Chen X. Research on a text data preprocessing method suitable for clustering algorithm. In: 2022 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS). IEEE; 2022. https://doi.org/10.1109/ISPDS56360.2022.9874172.

[10] Shial G, Sahoo S, Panigrahi S. An enhanced GWO algorithm with improved explorative search capability for global optimization and data clustering. Appl Artif Intell]. 2023;37(1). Available from: https://doi.org/10.1080/08839514.2023.2166232.

[11] Rahmani AM, Haider A, Ali S, Mohammadi M, Mehranzadeh A, Khoshvaght P, et al. A routing approach based on combination of gray wolf clustering and fuzzy clustering and using multi-criteria decision making approaches for WSN-IoT. Comput Electr Eng]. 2025;122(109946):109946. Available from: https://doi.org/10.1016/j.compeleceng.2024.109946.

[12] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Adv Eng Softw]. 2014;69:46–61. Available from: https://doi.org/10.1016/j.advengsoft.2013.12.007.

[13] Rashaideh H, Sawaie A, Al-Betar MA, Abualigah LM, Al-laham MM, Al-Khatib RM, et al. A grey wolf optimizer for text document clustering. J Intell Syst]. 2019;29(1):814–30. Available from: https://doi.org/10.1515/jisys-2018-0194.

[14] Yang Y, Zhang X, Li B, Qin K. A grey wolf optimizer-based topology shaping method for UAV swarm. In: 2022 IEEE 5th International Conference on Electronics Technology (ICET). IEEE; 2022. https://doi.org/10.1109/ICET55676.2022.9824250.

[15] Ahmadi R, Ekbatanifard G, Bayat P. A modified grey wolf optimizer based data clustering algorithm. Appl Artif Intell]. 2021;35(1):63–79. Available from: https://doi.org/10.1080/08839514.2020.1842109.

[16] Kihel BK, Chouraqui S. A Novel Genetic Grey Wolf optimizer for Global optimization and Feature Selection. In: 2020 Second International Conference on Embedded & Distributed Systems (EDiS). IEEE; 2020. https://doi.org/10.1109/EDiS49545.2020.9296449.

[17] Millah IS, Chang PC, Teshome DF, Subroto RK, Lian KL, Lin J-F. An enhanced grey wolf optimization algorithm for photovoltaic maximum power point tracking control under partial shading conditions. IEEE Open J Ind Electron Soc]. 2022;3:392–408. Available from: https://doi.org/10.1109/OJIES.2022.3179284.

[18] Jiang K, Ni H, Sun P, Han R. An improved binary grey wolf optimizer for dependent task scheduling in edge computing. In: 2019 21st International Conference on Advanced Communication Technology (ICACT). IEEE; 2019. https://doi.org/10.23919/ICACT.2019.8702018.

[19] Xu H, Liu X, Su J. An improved grey wolf optimizer algorithm integrated with Cuckoo Search. In: 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). IEEE; 2017. https://doi.org/10.1109/IDAACS.2017.8095129.

[20] Wang W, Ye L, Zhang Y, Li Y. Manual bidirectional dislocation flip chip alignment technology based on image preprocessing. In: 2024 25th International Conference on Electronic Packaging Technology (ICEPT). IEEE; 2024. p. 1–5. https://doi.org/10.1109/ICEPT63120.2024.10668765.

[21] Kyriaki K, Koukopoulos D, Fidas CA. A comprehensive survey of EEG preprocessing methods for cognitive load assessment. IEEE Access 2024;12:23466–89. Available from: https://doi.org/10.1109/ACCESS.2024.3360328.

[22] Murti DMP, Pujianto U, Wibawa AP, Akbar MI. K-nearest neighbor (K-NN) based missing data imputation. In: 2019 5th International Conference on Science in Information Technology (ICSITech). IEEE; 2019. https://doi.org/10.1109/ICSITech46713.2019.8987530.

[23] Salem N, Hussein S. Data dimensional reduction and principal omponents analysis. Procedia Computer Science. 2019;163:292–9. https://doi.org/10.1016/j.procs.2019.12.111.

[24] Nanni L, Ghidoni S, Brahnam S. Handcrafted vs. non-handcrafted features for computer vision classification. Pattern Recognit 2017;71:158–72. Available from: https://doi.org/10.1016/j.patcog.2017.05.025.

[25] UCI machine learning repository Uci.edu. [cited 2025 Apr 15]. Available from: https://archive.ics.uci.edu/.

[26] Nainggolan R, Perangin-angin R, Simarmata E, Tarigan AF. Improved the performance of the K-means cluster using the Sum of Squared Error (SSE) optimized by using the elbow method. J Phys Conf Ser. 2019;1361(1):012015. Available from: https://doi.org/10.1088/1742-6596/1361/1/012015.

[27] Binu Jose A, Das P. A multi-objective approach for inter-cluster and intra-cluster distance analysis for numeric data. In: Lecture Notes in Networks and Systems. Singapore; Singapore: Springer; 2022. p. 319–32. https://doi.org/10.1007/978-981-19-0707-4_30.

[28] Li C, Wu T. A boosted clustering algorithm for distributed homogeneous data mining. In: 2006 6th World Congress on Intelligent . control and automation 2006.