Invasive Weed Optimization K-Means Performance Robust Operations (IWOKM PRO) in High-Dimensional Datasets

Ni Luh Gede Pivin Suwirmayanti
Faculty of Engineering, Udayana University, Bali, Indonesia | Department of Computer System, Institut Teknologi and Bisnis STIKOM Bali, Indonesia
pivin@stikom-bali.ac.id (corresponding author)

I. Ketut Gede Darma Putra
Department of Information Technology, Faculty of Engineering, Udayana University, Bali, Indonesia
ikgdarmaputra@unud.ac.id

Made Sudarma
Department of Electrical Engineering, Faculty of Engineering, Udayana University, Bali, Indonesia
msudarma@unud.ac.id

I. Made Sukarsa
Department of Information Technology, Faculty of Engineering, Udayana University, Bali, Indonesia
sukarsa@unud.ac.id

Emy Setyaningsih
Department of Computer Systems Engineering, Universitas AKPRIND Indonesia, Yogyakarta, Indonesia
emysetyaningsih@akprind.ac.id

Ricky Aurelius Nurtanto Diaz
Department of Computer Systems, Institut Teknologi and Bisnis STIKOM Bali, Bali, Indonesia
ricky@stikom-bali.ac.id

Received: 24 March 2025 | Revised: 29 April 2025 | Accepted: 4 May 2025

 DOI: https://doi.org/10.48084/etasr.11112

ABSTRACT

This study presents a novel clustering approach called Invasive Weed Optimization K-Means Performance Robust Operations (IWOKM PRO) to improve clustering performance on high-dimensional datasets. Unlike previous IWOKM implementations, IWOKM PRO focuses on optimizing parameter efficiency to conserve computational resources and applies centroid selection techniques to accelerate convergence and enhance clustering results. To evaluate its effectiveness, IWOKM PRO was tested on stock data collected from the Indonesia Stock Exchange (IDX), comprising 604 stocks with adjusted closing price features from January 2019 to December 2023. The experimental results demonstrate that IWOKM PRO outperforms the original IWOKM method in both convergence speed and clustering accuracy. Specifically, in the three-cluster scenario, IWOKM PRO achieved the best fitness value in 1.37 s with a Sum of Squared Errors (SSE) of 973.6434. In the five-cluster scenario, IWOKM PRO reached an average convergence time of 6.45 s with an SSE of 443.8437. Compared to IWOKM, these results significantly improve computational efficiency and clustering performance. In general, this study shows that IWOKM PRO is an effective solution to improve the efficiency and accuracy of clustering, particularly for high-dimensional financial datasets.

Keywords-clustering; k-means; IWO; IWOKM PRO; high-dimensional dataset; hybrid metaheuristic; Indonesia stock exchange; stocks

I. INTRODUCTION

Research on the exploration of high-dimensional datasets is growing rapidly in line with the development of computing technology and artificial intelligence [1-3]. High-dimensional datasets can be handled with various techniques, one of which is data clustering [3-4]. High-dimensional data clustering is used in a variety of applications, including financial sectors such as banking and stocks. In these sectors, clustering methods play an important role in real-world implementations, such as customer segmentation, credit risk analysis, and the detection of suspicious transactions in the prevention of money laundering [5-7]. Furthermore, in stock market analysis, clustering methods can help classify stocks based on price movement patterns, volatility, and other fundamental factors. Recent studies have shown that the use of a hybrid approach in clustering can improve prediction accuracy and provide deeper insights into stock market dynamics [8-9].

K-means is one of the most popular traditional clustering algorithms. The advantage of this algorithm is its ease of adaptation, as K-means is easy to apply in various data mining processing, making it one of the most widely used clustering algorithms [10-11]. However, in its application, this algorithm often faces difficulties in handling high-dimensional data due to its sensitivity to centroid initialization, convergence problems, and a tendency to get stuck in local optima [12-13]. To overcome these limitations, hybrid metaheuristic approaches have been developed by integrating optimization techniques to improve performance and efficiency.

Hybrid metaheuristic approaches combine the advantages of various optimization methods to overcome the limitations of traditional ones. Various optimization algorithms have been used to improve the performance of K-means, such as Adaptive K-means Clustering Undersampling (AKCUS) [13], Random Forest classifier [10], Particle Swarm Optimization (PSO) [11, 14], Artificial Neural Network, ABC [15], Improved Bat Algorithm [16], Genetic Algorithm (GA), and Invasive Weed Optimization (IWO) [17]. Hybrid techniques that incorporate these methods have been shown to provide more accurate and stable results, especially in clustering high-dimensional datasets used in a variety of applications, including financial sectors such as banking and stocks.

Some of the following methods have been hybridized to manage high-dimensional datasets [18]. Newly developed metaheuristic models have shown strong performance; for example, the WWO algorithm improved accuracy and F score rates by an average of 4% and 7%, respectively, compared to existing clustering algorithms [19]. An adaptive metaheuristic approach has also been introduced to detect similarities between two melodies, combining two distinct methods: one utilizing a textual representation and the other employing a vectorial representation of them [20]. Extensive experimental results indicate that CFOA delivers outstanding optimization performance across various optimization tasks. Furthermore, when applied to data clustering problems, CFOA achieves an overall error rate of less than 20%, leading to improved clustering outcomes [21]. Additionally, the Puma Optimizer (PO), a novel metaheuristic optimization algorithm inspired by the intelligence and hunting strategies of pumas, effectively balances exploration and exploitation, facilitating a more efficient search for optimal solutions [22].

Among hybrid approaches, Invasive Weed Optimization K-Means (IWOKM) emerged as a promising technique that leverages the global exploration capabilities of Invasive Weed Optimization (IWO) while maintaining the efficient clustering mechanism of K-means. This combination allows IWOKM to overcome some of the limitations of traditional K-means by improving centroid selection, reducing sensitivity to initialization, and improving overall clustering accuracy. IWOKM has shown significant advantages in big data analysis, such as financial data clustering [23]. Recent studies have investigated hybrid models that integrate IWO with K-means to improve clustering accuracy and efficiency. The incorporation of enhanced IWO demonstrated promising results in optimizing intra-cluster similarity and inter-cluster separation. Additionally, a combination of Improved IWO and the Bird Swarm Algorithm (BSA) was applied to select features in diabetes prediction, achieving high accuracy rates of up to 90% [24]. Furthermore, the hybrid NSGA II-IWO method is a valuable approach to enhance the performance of FDSHX and advance optimization design theories for heat exchangers [25]. The Neuro-IWO model exhibited a deviation of less than 5%, with experimental results confirming its effectiveness in complex environments. Moreover, the hybrid route planning algorithm leveraging Neuro-IWO showed superior accuracy and achieved the shortest route compared to Neural Networks (NN) [26].

However, despite its various benefits, IWOKM still faces several challenges, including high computational complexity, which leads to increased execution time, complex parameter tuning, which requires many experiments to obtain optimal configurations, and susceptibility to noisy and outlier data, which can affect herd resilience [17]. To overcome these weaknesses, IWOKM PRO (Invasive Weed Optimization K-Means Performance Robust Operations) was developed as a better version of IWOKM. IWOKM PRO integrates a variety of performance-based improvements, where the focus is on parameter efficiency optimization to save computing resources and the application of centroid selection techniques to accelerate convergence. The introduction of IWOKM PRO seeks to improve clustering more effectively, resulting in better accuracy and scalability in high-dimensional datasets. This study aimed to evaluate the effectiveness and performance improvement of IWOKM PRO by comparing it with the previous clustering method [23]. Using advanced hybrid metaheuristics, this research is expected to contribute to the development of better clustering techniques to handle high-dimensional data environments.

The proposed IWOKM PRO builds on these prior works by introducing an advanced hybrid clustering approach specifically designed for high-dimensional datasets. Unlike conventional optimization-based clustering models, IWOKM PRO integrates enhanced Performance Robust Operations (PRO) to achieve:

- Higher clustering accuracy by improving centroid optimization through IWO.

- Better cluster compactness and separation, reducing intra-cluster variance.
- Improved computational efficiency, making the algorithm scalable for large-scale high-dimensional datasets.

This study focuses on clustering high-dimensional financial datasets, specifically the Indonesia Stock Exchange (IDX) dataset, to evaluate the effectiveness of IWOKM PRO. Using the strengths of IWO in global search and K-means in partitioning, IWOKM PRO offers a novel approach that addresses existing challenges in high-dimensional clustering, particularly in financial data analysis.

II. PROPOSED METHOD

The development of IWOKM PRO to manage high-dimensional data involves several stages. The following is an overview of the process of the proposed IWOKM PRO method.

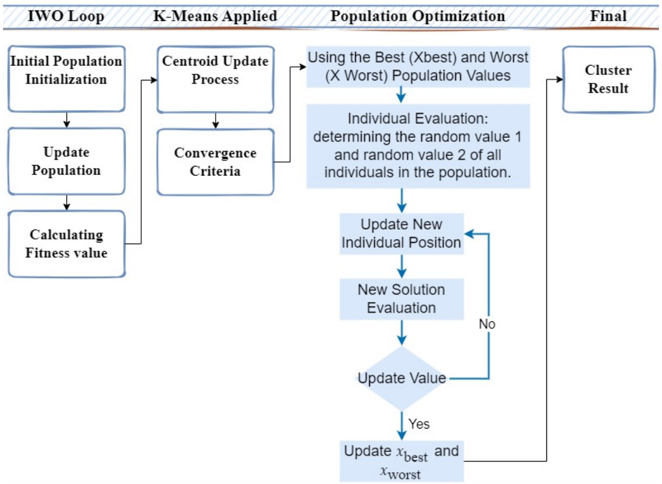

Fig. 1. IWOKM PRO method.

As shown in Figure 1, the IWOKM PRO method builds on the standard IWOKM process by introducing an additional population optimization stage. The conventional IWOKM workflow consists of three main processes: population initialization, population update, and fitness value calculation, followed by the application of K-means clustering through centroid updates and convergence checks. In IWOKM PRO, after the standard IWOKM steps are completed, the process continues with a dedicated population optimization phase. This phase aims to further refine the clustering solutions obtained from the K-means process by optimizing individuals within the population. In this stage, the best (X_{best}) and worst (X_{worst}) individuals are identified, and each individual's position is updated based on random values to enhance the diversity and quality of solutions. New solutions are iteratively evaluated and updates are made until a better solution or convergence is achieved.

Thus, IWOKM PRO represents an improvement over the normal IWOKM flow by integrating a systematic optimization mechanism after initial clustering, ensuring that the final cluster results are more accurate, stable, and computationally efficient. Adding the population optimization phase is the key innovation that differentiates IWOKM PRO from the traditional IWOKM model.

III. RESULT AND DISCUSSION

A. Dataset

A high-dimensional dataset was obtained from the IDX data using scrapping techniques obtained from the Yahoo Finance API, using 604 IHSG stock close price data from January 2019 to December 2023. The results obtained from this data acquisition process are 604 rows (in rows that are issuer data) and 1231 columns. Figure 2 shows a sample of the dataset used in this study.

	Emiten	2019-01-01	2019-01-02	2019-01-03	2019-01-04	2019-01-07	2019-01-08	2019-01-09	2019-01-10	2019-01-11	...
0	ABBA.JK	0.001053	0.001053	0.001053	0.001053	0.001053	0.001053	0.001052	0.001053	0.001052	...
1	ABDA.JK	0.002339	0.002339	0.002339	0.002339	0.002339	0.002339	0.002339	0.002339	0.002339	...
2	AKKU.JK	0.001048	0.001048	0.001048	0.001048	0.001048	0.001048	0.001048	0.001048	0.001048	...
3	ALMI.JK	0.001118	0.001120	0.001121	0.001121	0.001122	0.001125	0.001126	0.001127	0.001131	...
4	ALTO.JK	0.001118	0.001118	0.001118	0.001118	0.001118	0.001118	0.001109	0.001111	0.001114	...
...	...	...	...	...	...	...	...	...	...	...	...

Fig. 2. Sample of the private dataset.

B. Preprocessing

The dataset consists of adjusted closing price data from various issuers listed on the IDX, where stock prices vary significantly across different companies. The min-max scaler technique was applied to the data to standardize these variations before further processing. Min-max normalization is recognized as one of the most effective techniques to enhance the performance of machine learning models [27]. This method is particularly well-suited for stock price data as it maintains proportional increases and decreases between stocks while rescaling the data into a consistent range. As a result, each stock contributes equally to distance-based calculations, avoiding bias caused by stocks with higher nominal values dominating the clustering results. By preserving relative price movements among issuers, min-max normalization plays a vital role in achieving accurate, balanced, and interpretable clustering outcomes for financial datasets.

C. IWOKMPRO Results

The optimization process using the IWOKM PRO method consists of the following steps:

Initialization of the initial position of each individual in the population. Determine the individuals who have the best (x_{best}) and worst (x_{worst}) positions. Determine $random_1$ and $random_2$ values in the range of 0 to 1 for each individual generated to regulate movement toward the best position and away from the worst position. The new position update for each individual is calculated using the following formula to produce a new, more optimal position :

$$x_a' = x_a + (random_1 \cdot (x_{best} - x_a) - random_2 \cdot (x_{worst} - x_a))$$

where x_a' is a new solution for a, x_a is the individual position or solution on index a in the current population, $random_1$ and

$random_2$ are random values that result in variations in the movement of individuals toward the best solution and away from the worst solution, and x_{best} and x_{worst} are the best and worst individuals in the population, respectively.

After the calculation, the new solution is evaluated, and if it is better than the previous position, then x_{best} and x_{worst} are updated. This process continues to iterate until a stop condition is reached, such as the maximum number of iterations or very small solution changes. When the stop condition is met, the optimization process is complete, and the last best position found becomes the final result of optimization.

The performance of the model was evaluated using computational time and SSE. This test aimed to determine the most optimal combination of parameters in producing the best clustering with efficient computing time. The test parameters of the developed model were 10, 50, and 100 iterations and 10, 60, and 100 population sizes.

Based on these parameters, the proposed model was tested with three and five clusters, with the results shown in Tables I and II, respectively.

TABLE I. IWOKM PRO 3 CLUSTER PERFORMANCE

Parameters	Time to reach the best fitness	SSE value
Itr10Pop10	1.2 s	973.643475346027913
Itr10Pop60	1.2 s	973.643475346027913
Itr10Pop100	1.5 s	1292.646719908244904
Itr50Pop10	1.1 s	973.643475346027913
Itr50Pop60	1.7 s	973.643475346027913
Itr50Pop100	2.1 s	973.643475346027913
Itr100Pop10	1.0 s	973.643475346027913
Itr100Pop60	1.0 s	973.643475346027913
Itr100Pop100	1.6 s	973.643475346027913

TABLE II. IWOKM PRO 5 CLUSTER PERFORMANCE

Parameters	Time to reach the best fitness	SSE value
Itr10Pop10	1.3 s	473.507170798511311
Itr10Pop60	3.7 s	473.507170798511311
Itr10Pop100	6.9 s	480.006359429873555
Itr50Pop10	0.7 s	443.843722695888118
Itr50Pop60	5.3 s	443.843722695888118
Itr50Pop100	15.5 s	443.843722695888118
Itr100Pop10	2.3 s	443.843722695888118
Itr100Pop60	6.2 s	443.843722695888118
Itr100Pop100	16.2 s	443.843722695888118

The results of the IWOKM PRO model for three clusters show consistent SSE values in the range of 973.6434 for eight different tests, except for one test on a combination of population 100 with 10 iterations (Itr10Pop100) that resulted in an SSE value of 1292.6467. In terms of computational time, the average was under 2 s, with only one test with a combination of population of 100 with 50 iterations that resulted in an SSE score of 973.6434 but with a time to achieve the best fitness score of 2.1 s. Figure 4 illustrates how the IWOKM PRO model performed in terms of the SSE values generated with the time it took to achieve the best fitness value. The best SSE values with the lowest computation times were obtained when testing using a combination of population 10 with 100 iterations and population 60 with 100 iterations.

In the experiment with five clusters, IWOKM PRO produced consistent SSE values for eight tests, except for the test with a population of 100 and 10 iterations that produced the highest SSE score and a fairly high time to achieve it.

D. Comparison with IWO

To determine the capabilities of the IWOKM PRO model, tests were also carried out with the same test parameters but with the IWOKM model without optimization. Table III shows the performance results of IWOKM for a target of three clusters, where the SSE value results varied between 973.6434, 1116.8577, and 1292.6467. The lowest time to achieve the best fitness score was obtained with a combination of a population of 10 with 100 iterations, which took 0.7 s to reach the best SSE value of 1292.6467.

TABLE III. IWOKM 3 CLUSTER PERFORMANCE

Parameters	Time to reach the best fitness	SSE value
Itr10Pop10	2.6 s	1116.857758847424293
Itr10Pop60	1.8 s	973.643475346027913
Itr10Pop100	2.4 s	1292.646719908244904
Itr50Pop10	1.7 s	1292.646719908244904
Itr50Pop60	22.2 s	973.643475346027913
Itr50Pop100	5.8 s	973.643475346027913
Itr100Pop10	0.7 s	1292.646719908244904
Itr100Pop60	35.8 s	973.643475346027913
Itr100Pop100	30.7 s	973.643475346027913

Table IV shows the performance results of IWOKM for five clusters. The distribution of SSE values in this experiment ranged from 473.5071 to 538.9385. The lowest SSE value with the lowest time to reach the best fitness was obtained with a population of 100 with 50 iterations. The lowest time was 1.4 s when using a population of 10 with 10 iterations but resulted in a higher SSE value of 480.0063.

TABLE IV. IWOKM 5 CLUSTER PERFORMANCE

Parameters	Time to reach the best fitness	SSE Value
Itr10Pop10	1.4 second	480.006359429873555
Itr10Pop60	2.5 second	538.938597237626595
Itr10Pop100	10.5 second	473.507170798511311
Itr50Pop10	8.9 second	473.507170798511311
Itr50Pop60	28.6 second	473.507170798511311
Itr50Pop100	2.2 second	473.507170798511311
Itr100Pop10	13.5 second	473.507170798511311
Itr100Pop60	25.9 second	473.507170798511311
Itr100Pop100	43.4 second	473.507170798511311

Figures 3 and 4 compare IWOKM PRO with IWOKM in tests for three and five clusters, respectively. In Figure 3, it can be observed that the computation time to achieve the best fitness value was consistently lower in IWOKM PRO, with an average time of 1.37 s and a best SSE value of 973.6434. As shown in Figure 4, for the 5-cluster target trial, IWOKM PRO was also superior to IWOKM in producing the best SSE value in less time. The average time to achieve the best fitness score in IWOKM PRO was 6.45 s with the lowest SSE value of 443.8437, while the average time to achieve the best fitness score for the IWOKM model was 15.21 s with the lowest SSE value of 473.5071.

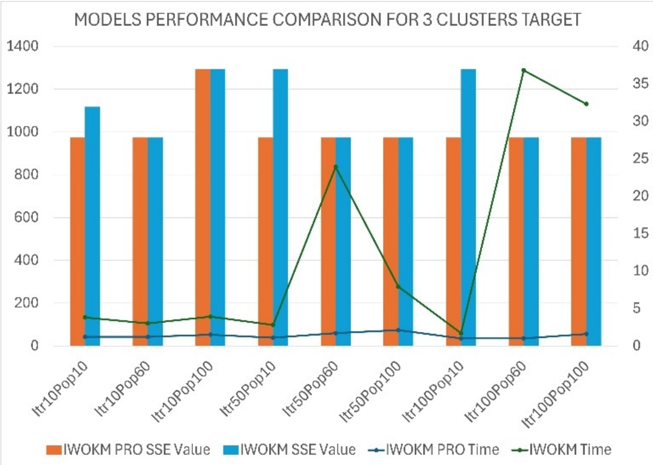

Fig. 3. Performance comparison for three clusters.

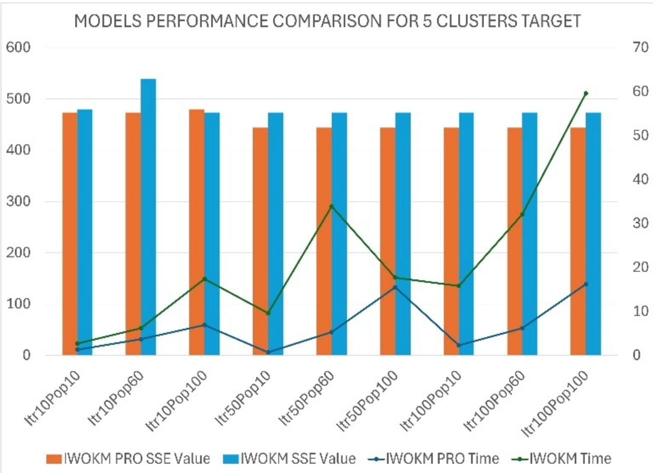

Fig. 4. Performance comparison for five clusters.

IV. CONCLUSION

This study introduced IWOKM PRO as an enhancement of IWOKM with robust operations, focusing on optimizing parameter efficiency to save computing resources and applying centroid selection techniques to accelerate convergence, thereby improving clustering performance on high-dimensional datasets. The proposed IWOKM PRO model was evaluated on data independently collected from the IDX, consisting of adjusted closing price data of 604 rows (representing issuers) and 1231 columns. The tests involved different iterations, populations, and number of clusters, and the results showed that IWOKM PRO was superior to IWOKM. For a target of three clusters, IWOKM PRO reached the best fitness score with an average time of 1.37 s, with the best SSE value of 973.6434. For the test with a target of five clusters, IWOKM PRO was also superior to IWOKM in terms of the best SSE value and the time to achieve the best fitness score, requiring 6.45 s for the lowest SSE of 443.8437. In contrast, the average time for the IWOKM model to reach the best fitness value was 15.21 s, with the lowest SSE being 473.5071.

According to previous studies, hybrid models based on IWO have demonstrated significant potential in enhancing clustering performance; however, computational challenges persist, particularly when applied to high-dimensional datasets. IWOKM PRO, which integrates Performance Robust Operations, was shown to improve clustering efficiency and accuracy, especially in the analysis of financial data, with a specific focus on stock data obtained from IDX. In the future, the IWOKM PRO model will be tested on different high-dimensional datasets and compared with other clustering optimization models.

REFERENCES

[1] A. Hartebrodt, R. Röttger, and D. B. Blumenthal, "Federated singular value decomposition for high-dimensional data," *Data Mining and Knowledge Discovery*, vol. 38, no. 3, pp. 938–975, May 2024, https://doi.org/10.1007/s10618-023-00983-z.

[2] J. Q. Yang *et al.*, "Bi-Directional Feature Fixation-Based Particle Swarm Optimization for Large-Scale Feature Selection," *IEEE Transactions on Big Data*, vol. 9, no. 3, pp. 1004–1017, Jun. 2023, https://doi.org/10.1109/TBDATA.2022.3232761.

[3] F. Cheng, J. Cui, Q. Wang, and L. Zhang, "A Variable Granularity Search-Based Multiobjective Feature Selection Algorithm for High-Dimensional Data Classification," *IEEE Transactions on Evolutionary Computation*, vol. 27, no. 2, pp. 266–280, Apr. 2023, https://doi.org/10.1109/TEVC.2022.3160458.

[4] X. F. Song, Y. Zhang, D. W. Gong, and X. Z. Gao, "A Fast Hybrid Feature Selection Based on Correlation-Guided Clustering and Particle Swarm Optimization for High-Dimensional Data," *IEEE Transactions on Cybernetics*, vol. 52, no. 9, pp. 9573–9586, Sep. 2022, https://doi.org/10.1109/TCYB.2021.3061152.

[5] P. Sood, C. Sharma, S. Nijjer, and S. Sakhuja, "Review the role of artificial intelligence in detecting and preventing financial fraud using natural language processing," *International Journal of System Assurance Engineering and Management*, vol. 14, no. 6, pp. 2120–2135, Dec. 2023, https://doi.org/10.1007/s13198-023-02043-7.

[6] M. Tao, E. Silva, M. S. Sheng, L. Wen, and L. Qi, "How financial clustering influences China's green development: Mechanism investigation and empirical discussion," *Journal of Environmental Management*, vol. 347, Dec. 2023, Art. no. 119081, https://doi.org/10.1016/j.jenvman.2023.119081.

[7] V. Sinap, "Comparative analysis of machine learning techniques for credit card fraud detection: Dealing with imbalanced datasets," *Turkish Journal of Engineering*, vol. 8, no. 2, pp. 196–208, Apr. 2024, https://doi.org/10.31127/tuje.1386127.

[8] R. Hurriyati, A. A., S. Sulastri, L. Lisnawati, and T. Sawangsang, "Stock Market Trend Analysis and Machine Learning-based Predictive Evaluation," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 14, no. 3, pp. 267–281, Sep. 2023, https://doi.org/10.58346/JOWUA.2023.I3.020.

[9] M. Ashrafzadeh, H. M. Taheri, M. Gharehgozlou, and S. Hashemkhani Zolfani, "Clustering-based return prediction model for stock pre-selection in portfolio optimization using PSO-CNN+MVF," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 9, Oct. 2023, Art. no. 101737, https://doi.org/10.1016/j.jksuci.2023.101737.

[10] A. Bhattacharjee, R. Murugan, and T. Goel, "A hybrid approach for lung cancer diagnosis using optimized random forest classification and K-means visualization algorithm," *Health and Technology*, vol. 12, no. 4, pp. 787–800, Jul. 2022, https://doi.org/10.1007/s12553-022-00679-2.

[11] M. Behera *et al.*, "Automatic Data Clustering by Hybrid Enhanced Firefly and Particle Swarm Optimization Algorithms," *Mathematics*, vol. 10, no. 19, Sep. 2022, Art. no. 3532, https://doi.org/10.3390/math10193532.

[12] A. M. Ikotun and A. E. Ezugwu, "Boosting k-means clustering with symbiotic organisms search for automatic clustering problems," *PLOS*

ONE, vol. 17, no. 8, Aug. 2022, Art. no. e0272861, https://doi.org/10.1371/journal.pone.0272861.

[13] R. M. Mathew and R. Gunasundari, "A Cluster-based Undersampling Technique for Multiclass Skewed Datasets," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10785–10790, Jun. 2023, https://doi.org/10.48084/etasr.5844.

[14] Y. Li, X. Zhou, J. Gu, K. Guo, and W. Deng, "A Novel K-Means Clustering Method for Locating Urban Hotspots Based on Hybrid Heuristic Initialization," *Applied Sciences*, vol. 12, no. 16, Aug. 2022, Art. no. 8047, https://doi.org/10.3390/app12168047.

[15] Q. Pu, J. Gan, L. Qiu, J. Duan, and H. Wang, "An efficient hybrid approach based on PSO, ABC and k-means for cluster analysis," *Multimedia Tools and Applications*, vol. 81, no. 14, pp. 19321–19339, Jun. 2022, https://doi.org/10.1007/s11042-021-11016-6.

[16] A. Kaur and Y. Kumar, "Neighborhood search based improved bat algorithm for data clustering," *Applied Intelligence*, vol. 52, no. 9, pp. 10541–10575, Jul. 2022, https://doi.org/10.1007/s10489-021-02934-x.

[17] S. Saleem, F. Hussain, and N. K. Baloch, "IWO-IGA—A Hybrid Whale Optimization Algorithm Featuring Improved Genetic Characteristics for Mapping Real-Time Applications onto 2D Network on Chip," *Algorithms*, vol. 17, no. 3, Mar. 2024, Art. no. 115, https://doi.org/10.3390/a17030115.

[18] A. M. Ikotun and A. E. Ezugwu, "Enhanced Firefly-K-Means Clustering with Adaptive Mutation and Central Limit Theorem for Automatic Clustering of High-Dimensional Datasets," *Applied Sciences*, vol. 12, no. 23, Nov. 2022, Art. no. 12275, https://doi.org/10.3390/app122312275.

[19] A. Kaur and Y. Kumar, "A new metaheuristic algorithm based on water wave optimization for data clustering," *Evolutionary Intelligence*, vol. 15, no. 1, pp. 759–783, Mar. 2022, https://doi.org/10.1007/s12065-020-00562-x.

[20] D. Malandrino, R. De Prisco, M. Ianulardo, and R. Zaccagnino, "An adaptive meta-heuristic for music plagiarism detection based on text similarity and clustering," *Data Mining and Knowledge Discovery*, vol. 36, no. 4, pp. 1301–1334, Jul. 2022, https://doi.org/10.1007/s10618-022-00835-2.

[21] H. Jia, Q. Wen, Y. Wang, and S. Mirjalili, "Catch fish optimization algorithm: a new human behavior algorithm for solving clustering problems," *Cluster Computing*, vol. 27, no. 9, pp. 13295–13332, Dec. 2024, https://doi.org/10.1007/s10586-024-04618-w.

[22] B. Abdollahzadeh *et al.*, "Puma optimizer (PO): a novel metaheuristic optimization algorithm and its application in machine learning," *Cluster Computing*, vol. 27, no. 4, pp. 5235–5283, Jul. 2024, https://doi.org/10.1007/s10586-023-04221-5.

[23] N. L. G. P. Suwirmayanti, E. Setyaningsih, R. A. N. Diaz, and K. Budiarta, "Optimization of the K-Means Method for Clustering Banking Data Using the Hybrid Model of Invasive Weed Optimization and K-Means (IWOKM)." ICIC International, 2024, https://doi.org/10.24507/icicel.18.04.413.

[24] C. N. Aher and A. K. Jena, "Improved invasive weed bird swarm optimization algorithm (IWBSOA) enabled hybrid deep learning classifier for diabetic prediction," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 4, pp. 3929–3945, Apr. 2023, https://doi.org/10.1007/s12652-022-04462-z.

[25] Z. Xu, Z. Yu, X. Ning, X. Wan, Z. Qu, and C. Zhao, "Hybrid of non-dominated sorting genetic algorithm II and invasive weed optimization for fan duct surface heat exchanger configuration optimization design," *Case Studies in Thermal Engineering*, vol. 61, Sep. 2024, Art. no. 105149, https://doi.org/10.1016/j.csite.2024.105149.

[26] B. Sahoo, D. Das, K. C. Pujhari, and Vikas, "Optimization of route planning for the mobile robot using a hybrid Neuro-IWO technique," *International Journal of Information Technology*, vol. 17, no. 3, pp. 1431–1439, Apr. 2025, https://doi.org/10.1007/s41870-024-02231-z.

[27] M. Shantal, Z. Othman, and A. A. Bakar, "A Novel Approach for Data Feature Weighting Using Correlation Coefficients and Min–Max Normalization," *Symmetry*, vol. 15, no. 12, Dec. 2023, Art. no. 2185, https://doi.org/10.3390/sym15122185.