Indonesian Journal of Electrical Engineering and Computer Science
Vol. 41, No. 3, March 2026, pp. 1082~1094
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v41.i3.pp1082-1094

FGMPSO: a hybrid firefly-gradient-MOPSO framework for high-dimensional feature selection

Alwatben Batoul Rashed
Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia

Article Info

Article history:

Received Jul 19, 2025
Revised Jan 30, 2026
Accepted Mar 4, 2026

Keywords:

Feature selection
Firefly algorithm
Gradient descent
High-dimensional data
Multi-objective optimization
Particle swarm optimization

ABSTRACT

When working with high-dimensional datasets, selecting the most relevant features is essential for improving both model clarity and processing efficiency, all while keeping predictive accuracy intact. In response to this challenge, the study introduces firefly-gradient-multi-objective particle swarm optimization (FGMPSO), an advanced hybrid technique that blends the firefly algorithm, gradient descent (GD), and multi-objective particle swarm optimization (MOPSO). This approach is specifically designed to identify an optimal subset of features that balances dimensionality reduction with strong classification performance. The method was evaluated on eight benchmark datasets and compared against multiple PSO-based feature selection techniques. The empirical results demonstrated that FGMPSO consistently achieved superior or competitive classification accuracy while selecting significantly fewer features. Notably, in several datasets, FGMPSO not only reduced dimensionality but also outperformed other methods in terms of classification accuracy. This efficiency is attributed to the intelligent exploration of the search space by the firefly algorithm, refinement via GD, and effective trade-off optimization enabled by MOPSO. The findings suggest that FGMPSO is a robust and scalable solution for feature selection, particularly suitable for complex and high-dimensional datasets. Its adaptability, convergence speed, and balance between dimensionality reduction and accuracy position it as a valuable tool in modern machine learning pipelines.

Corresponding Author:

Alwatben Batoul Rashed
Department of Information Technology, College of Computer, Qassim University
Buraydah, Saudi Arabia
Email: bwtban@qu.edu.sa

1. INTRODUCTION

Feature selection is a crucial step that helps machine learning models in development, especially when dealing with datasets that contain an overwhelming number of features. Its primary purpose is to identify the most relevant attributes that improve prediction quality while removing noisy or redundant information that does not generalize well across models [1]. Eliminating unnecessary features makes the model simpler and commonly enhances its performance. Feature selection also accelerates computation and improves accuracy by reducing dataset dimensionality and removing irrelevant variables [2]. This importance becomes even more evident in high-dimensional machine learning tasks, where an excessive number of features can burden learning algorithms. Furthermore, the choice of feature selection strategy significantly affects results, as methods such as wrapper and embedded approaches depend heavily on the learning model and may not generalize well to other classifiers [3].

Recent studies have highlighted the increasing importance of feature selection across diverse domains, including healthcare and finance, not only for reducing computational burden but also for improving interpretability and decision-making [4]. By focusing on the most meaningful attributes, machine learning models can identify more accurate and insightful patterns. Feature selection methods are generally categorized into three groups: filter, wrapper, and embedded techniques [5]. However, each group has limitations, such as high computational cost or the risk of removing subtle but informative features in complex datasets. Filter methods evaluate features independently of any classifier. They are computationally efficient but often overlook interactions among features. Wrapper methods-such as sequential forward selection (SFS) and sequential backward selection (SBS) evaluate subsets using a classifier and typically yield higher accuracy, but they are computationally expensive and prone to overfitting [6]. Embedded methods integrate feature selection into the learning process itself. As a result of these limitations, researchers are increasingly exploring more adaptive techniques such as particle swarm optimization (PSO), which provides greater flexibility and effectiveness for complex feature selection tasks [7].

However, existing PSO variants still face challenges: standard Binary PSO may converge prematurely due to low diversity, sigmoid-based improvements may struggle to balance exploration and exploitation [8], and quantum-inspired PSO approaches can sometimes become unstable or slow [9]. Other PSO variants such as Unified Binary PSO and Cooperative PSO [10] provide improved diversity but often require complex synchronization. Similarly, cat swarm optimization (CSO) [11] offers a balance between global and local search but can be sensitive to parameter settings and slow in convergence. The performance differences between classical wrapper methods are illustrated in Figure 1, which highlights the limitations of traditional approaches when dealing with high-dimensional data.

To address these gaps, we propose a new hybrid feature selection method called firefly-gradient-multi-objective particle swarm optimization (FGMPSO). It combines the global exploration of the firefly algorithm (FA), the local fine-tuning ability of gradient descent (GD), and the diversity-preserving nature of multi-objective particle swarm optimization (MOPSO). Our approach aims to select a compact set of features without sacrificing classification performance. We evaluate it on eight benchmark datasets and compare it with six existing PSO-based FS methods, as well as with classical greedy techniques like SFS and SBS.

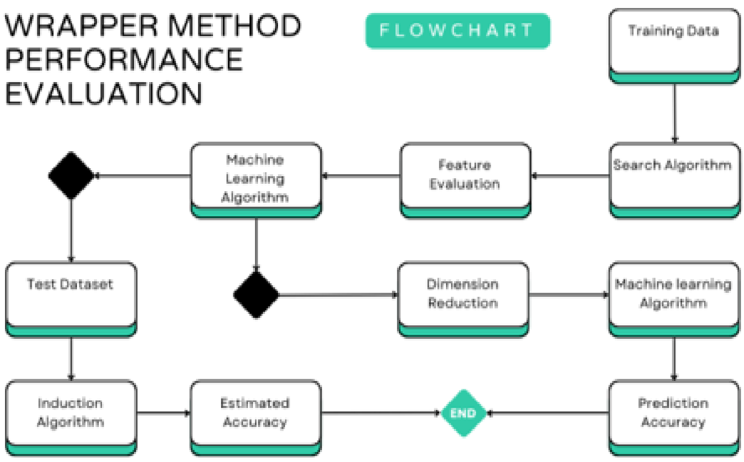

Figure 1. The performance evaluation of wrapper methods

2. METHOD

The FGMPSO algorithm FGMPSO integrates three complementary optimization strategies to strengthen the feature selection process: MOPSO, the FA, and GD. Feature selection is especially critical in high-dimensional learning problems, as it removes redundant or irrelevant attributes and improves classifier performance by focusing the model on the most informative features [1], [2], [5]. In FGMPSO, each candidate solution is encoded as a binary vector, where a value of "1" represents a selected feature and "0" indicates exclusion. The optimization begins with MOPSO, which simultaneously minimizes classification error and the total number of selected features, producing a Pareto-optimal balance between accuracy and dimensionality [12]-[24]. To reduce the risk of premature convergence and to enhance global diversity, the Firefly Algorithm is incorporated as a secondary exploration mechanism. FA uses a brightness-based attraction model that encourages solutions with better fitness to guide weaker ones toward more promising regions of the search space [23]. In the final refinement stage, GD adjusts feature importance by analyzing the gradient of prediction error, thereby fine-tuning the most promising candidate subsets and improving local convergence behavior [20]. By combining broad population-based exploration with targeted local refinement, FGMPSO is capable of identifying highly effective feature subsets even in complex, high-dimensional search spaces. This hybridization enables the algorithm to outperform classical wrapper methods and conventional PSO variants in both stability and selection quality [7]-[9]. The complete workflow of the proposed FGMPSO framework is illustrated in Figure 2, highlighting the coordinated roles of MOPSO, Firefly-based global exploration, and GD refinement.

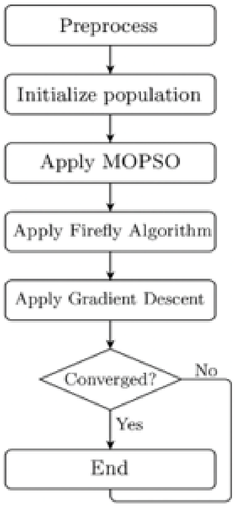

Figure 2. The framework of proposed FGMPSO approach

Algorithm 1 summarizes the overall workflow of the proposed FGMPSO framework, which is evaluated in the subsequent experimental section to demonstrate its effectiveness in high-dimensional feature selection tasks.

Algorithm 1: pseudo-code of FGMPSO Algorithm for Feature Selection

```
Input: Dataset D, population size N, max iterations T
Output: Optimal feature subset S*
1. Initialize population of particles P with binary feature vectors
2. Evaluate fitness of each particle using classifier accuracy and feature count
3. Store personal best pi and global best g for each particle
4. For t = 1 to T do:
       a. For each particle i in P do:
              i. Update velocity vi using MOPSO velocity equation
              ii. Update position xi using velocity and binary conversion
              iii. Evaluate fitness of new position
              iv. Update personal best pi if current fitness is better
       b. Update global best g from personal bests
       c. Apply Firefly Algorithm to refine best particles
       d. For each selected solution xi do:
              i. Apply Gradient Descent to minimize error rate
              ii. Update position xi using gradient updates
       e. Store non-dominated solutions in Pareto front
5. Return best solution S* from Pareto front
```

2.1. Additional implementation details of the FGMPSO procedure:

To ensure clarity in the algorithmic workflow, each module of the proposed FGMPSO procedure is described in greater detail below:

- Initialization: The process begins by initializing a population of particles, each encoded as a binary vector of length *d*, representing the selection or rejection of each feature. Algorithmic parameters including inertia weight (ω), cognitive and social coefficients (C1, C2), Firefly attractiveness (β), light absorption coefficient (γ), random step size (α), and the gradient descent learning rate (η) are initialized following benchmark values reported in the literature. This ensures a balanced exploration exploitation ratio at the start of the search.

- Fitness evaluation: Each particle is evaluated using the multi-objective fitness function that simultaneously minimizes the number of selected features and the classification error. The fitness values are computed using a random forest classifier, ensuring consistent evaluation across iterations. A dominance check is applied to determine whether a solution should be added to the external archive.
- MOPSO update: During the MOPSO phase, each particle updates its velocity and position according to binary transfer functions. These functions convert continuous velocities into probabilities, which are then used to flip bits in the feature vector. This mechanism ensures proper binary behavior and helps maintain diversity within the swarm.
- Firefly movement: In this stage, particles behave like fireflies, where each particle is attracted to brighter (i.e., more optimal) solutions. The movement is determined by the attractiveness function β·e^(−γr²), ensuring that particles closer to strong solutions move more deterministically while distant particles explore more widely. Random perturbation using α adds stochastic variation that improves escape from local minima.
- Gradient descent refinement: For top-performing solutions, GD is applied to refine feature weights. Using a decaying learning rate, GD gradually reduces the influence of poorly contributing features. Over time, this component fine-tunes local search by suppressing noisy or irrelevant attributes, reinforcing convergence toward more discriminative feature subsets.
- Archive update: The non-dominated archive is updated at every iteration. Pareto dominance rules are used to determine which solutions are preserved. The archive is further organized using crowding distance to ensure diversity along the Pareto front. This component guarantees that the algorithm maintains a well-balanced set of solutions representing different trade-offs between accuracy and dimensionality.
- Stopping criterion: The algorithm terminates when the maximum number of iterations (T) is reached. At this point, the final archive contains the Pareto-optimal feature subsets discovered during the search. The best subset is selected based on accuracy or the desired trade-off specified by the user.

This pseudo code provides a high level representation of the FGMPSO algorithm for feature selection. FGMPSO is selected due to its ability to achieve a high classification accuracy while minimizing the number of features selected, addressing a common limitation of other feature selection methods. Traditional methods often either select an excessive number of features or discard important ones. The FGMPSO algorithm integrates three powerful techniques: the FA, GD, and MOPSO. FA aids in exploring a broad range of feature subsets, GD fine-tunes the feature selection process by adjusting weights, and MOPSO identifies the most optimal feature subsets.

2.2. Mathematical modeling

Once the initial population is encoded, the MOPSO mechanism is first applied to explore the search space and evaluate the trade-off between accuracy and feature reduction. Each particle updates its velocity and position based on its individual experience and the global best solutions, following the standard PSO update rules [7]-[9]:

$$v_i^{t+1} = \omega v_i^t + C_1 \cdot rand(0,1) \cdot (p_i - x_i^t) + C_2 \cdot rand(0,1) \cdot (g_i - x_i^t) \tag{1}$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{2}$$

where v_i is the particle's velocity, x_i is its position, p_i and g_i denote the personal and global bests, respectively, and ω, C_1, C_2 are control parameters that regulate the balance between exploration and exploitation. To further improve global search capability and prevent premature convergence, the FA is incorporated after the MOPSO phase. FA moves a firefly i toward a more attractive firefly j based on brightness and distance according to [23]:

$$x_i^{t+1} = x_i^t + \beta e^{-\gamma r_{ij}^2} (x_j - x_i^t) + \alpha \cdot rand(-1,1) \tag{3}$$

where β is the attractiveness coefficient, γ controls light absorption, r_{ij} is the distance between fireflies, and α introduces controlled random perturbations. Following MOPSO and FA, GD is employed to refine promising solutions and adjust feature weights based on their contribution to reducing the classification error [20]. The GD update rule is defined as:

$$a_i^{t+1} = a_i^t - \eta \cdot \frac{\partial f(a)}{\partial a_i} \tag{4}$$

where a_i represents the weight of feature i, η is the learning rate, and $f(a)$ is the loss function. Through iterative updates, features with minor influence on classification accuracy receive lower weights and are progressively removed. FGMPSO repeats these steps iteratively MOPSO for global search, FA for diversity enhancement, and GD for local refinement to generate a Pareto-optimal set of solutions that simultaneously minimize dimensionality and maximize classification accuracy. To guide the optimization process, a dynamic multi-objective fitness function is defined as:

$$Fitness = w_1 \cdot (\frac{\text{Number of Features}}{\text{Total Features}}) + w_2 \cdot (1 - Accuracy) \quad (5)$$

where $w_1$and $w_2$are adaptive weights that shift emphasis during optimization. Initially, the algorithm prioritizes reducing the number of features; as the search progresses, the focus gradually transitions toward maximizing prediction accuracy, enabling a balanced and robust feature selection process [12], [24].

2.3. Convergence behavior of FGMPSO

The convergence behavior of the proposed FGMPSO algorithm is supported by the complementary roles of its three integrated components. MOPSO provides a Pareto-driven search mechanism in which non-dominated solutions are preserved in an external archive, allowing the swarm to gradually converge toward the true Pareto front. The Firefly Algorithm enhances global exploration by directing solutions toward brighter individuals, thereby reducing the risk of premature stagnation. GD further accelerates local convergence by refining solutions within promising regions of the search space. Collectively, these mechanisms ensure that the population maintains diversity in the early stages of the search and gradually shifts toward exploitation as iterations progress. This adaptive balance enables FGMPSO to converge more reliably than single-stage PSO variants, particularly in high-dimensional and multimodal landscapes.

2.4. Computational complexity analysis

The computational complexity of the proposed FGMPSO algorithm is governed by the combined cost of its three integrated optimization components: MOPSO, the FA, and GD. Let Ndenote the population size, $T$the number of iterations, and $d$the dimensionality (number of features).

MOPSO Stage In MOPSO, each particle updates its velocity and position by evaluating both personal and global best experiences, an operation that requires $O(d)$ time per particle per iteration [7]-[9]. Therefore, the computational cost of the MOPSO phase is:

$$O_{\text{MOPSO}} = O(N \cdot T \cdot d) \quad (6)$$

- Firefly algorithm stage

In FA, each firefly compares its brightness with every other firefly to determine movement toward more attractive solutions. This pairwise comparison results in a quadratic interaction complexity [23]:

$$O_{\text{FA}} = O(N^2 \cdot d) \quad (7)$$

- Gradient descent refinement

GD performs G iterative weight-update steps for each selected solution, where each update requires evaluating the gradient of the loss function with respect to the feature weights [20]:

$$O_{\text{GD}} = O(G \cdot d) \quad (8)$$

- Overall Complexity of FGMPSO

By combining the three stages, the total computational complexity of the proposed hybrid algorithm becomes:

$$O_{\text{FGMPSO}} = O(N \cdot T \cdot d) + O(N^2 \cdot d) + O(G \cdot d) \quad (9)$$

In practice, since $N$and $G$are typically much smaller than the dimensionality d, the dominant term is:

$$O_{\text{FGMPSO}} \approx O(N \cdot T \cdot d) \quad (10)$$

This demonstrates that although FGMPSO introduces additional computational steps compared to classical PSO variants, it remains computationally feasible for high-dimensional datasets. More importantly, the hybrid combination of swarm-based exploration and gradient-based refinement yields improved optimization stability and superior search performance in complex feature-selection scenarios [7], [12], [24].

2.5. Performance on different datasets

This study evaluates the performance of the proposed FGMPSO algorithm using eight widely adopted benchmark datasets obtained from the UCI machine learning repository [25], [26]. These datasets Sonar, Madelon, Hill-Valley, Isolet5, LSVT, Musk1, Urban, and Movement cover a broad range of dimensionalities and sample sizes, making them suitable for assessing feature selection methods under diverse conditions. The key characteristics of these datasets are summarized in Table 1.

Table 1. Characteristics of the benchmark datasets used in this study

Dataset	Instances	Feature	Classes	Description
Sonar	208	60	2	Classifies sonar signals bounced off metal cylinders vs. rocks.
HillValley	1,212	100	2	Time-series-like data distinguishing between “hill” and “valley” patterns.
Isolet5	7,797	617	26	Spoken letters of the alphabet categorized into 26 classes.
LSVT	126	310	2	Acoustic features from sustained vowels spoken by individuals with/without voice disorders.
MUSK 1	476	166	2	Molecules labeled as musks or non-musks based on chemical descriptors.
URBAN	2000	10	2	Audio characteristics of urban and non-urban environments.
MADELON	2000	500	2	Synthetic dataset for feature selection evaluation with a mix of relevant, redundant, and irrelevant features.
MOVEMENT	2000	90	2	Classifies people’s movement into 'moving' and 'not moving' classes.

The datasets used in this study vary substantially in terms of dimensionality, number of samples, and class distribution. This diversity provides a robust testing environment for evaluating the generalization capability of different feature selection algorithms across heterogeneous real-world conditions. Prior to analysis, all datasets were normalized to the [0,1] range using Min–Max scaling to ensure comparable feature scales. Model performance was assessed using a 5-fold cross-validation procedure, and a Random Forest classifier configured with default parameters and a fixed random seed (42) served as the evaluation model to maintain consistent experimental conditions. To benchmark the effectiveness of the proposed method, FGMPSO was compared against a comprehensive set of classical and metaheuristic feature selection approaches. The comparative methods included standard binary PSO (SBPSO) [7], improved sigmoid-based PSO, quantum-behaved binary PSO (QBPSO) [9], unified binary PSO (UBPSO) [17], cooperative PSO (CPSO) [10], and cat swarm optimization in both classical and adaptive variants (CSO(C) and CSO(A)) [11]. Additionally, two greedy wrapper methods SFS and SBS were implemented using their standard configurations in Weka. All metaheuristic baselines were parameterized following established literature guidelines and validated through preliminary pilot runs. Unless specified otherwise, each algorithm was executed with a population size of 30 over 50 iterations. The proposed FGMPSO framework integrates three complementary optimization strategies FA, GD, and MOPSO each contributing distinct strengths within the hybrid search mechanism. FA was implemented using 20 fireflies over 50 generations with dynamic attractiveness and randomized movement operators to enhance global exploration [23]. GD was applied for 100 refinement iterations with a decaying learning rate beginning at 0.1 to adjust feature weights and minimize classification error [20]. Finally, MOPSO employed binary encoding and dominance-based sorting with 40 particles and 50 iterations to jointly optimize feature subset compactness and predictive performance [12], [24].

2.6. Optimization algorithm parameter tuning

Table 2 represent all type of algorithms implemented with 50 iterations and each algorithm tuning setting is mentioned in it.

Table 2. all type of algorithms

Algorithm	Population size	Iterations	Algorithm
SBPSO	30	50	Inertia = 0.7, c1=c2=1.49
ISBPSO	30	50	λ=1.0, σ=0.5, S-shape TF
UBPSO	30	50	Adaptive transfer strategy
QBPSO	30	50	Quantum potential well strategy
CPSO	30 (5 per group)	50	6 sub-swarms
CSO(C/A)	30	50	Seeking memory pool = 5
FGMPSO	30	50	Firefly + GD + MOPSO hybrid
SFS, SBS	----	----	Weka default

All experiments were conducted using Python on a standard workstation equipped with an Intel Core i5 processor and 8 GB of RAM. The implementation utilized several scientific computing libraries, including NumPy and Pandas for data manipulation, Scikit-learn for preprocessing, normalization, cross-validation, and classifier evaluation, PySwarms for PSO-based baseline methods, and custom Python modules for implementing the firefly algorithm and GD components. A fixed random seed (42) was applied across all runs to ensure reproducibility. The optimization process used a dynamic multi-objective fitness function in which the weighting factors were adaptively adjusted throughout the search initially prioritizing feature minimization and gradually shifting emphasis toward maximizing classification accuracy as optimization progressed. The final feature subset was obtained by selecting the most stable features consistently identified across the three integrated components of FGMPSO MOPSO exploration, firefly-based enhancement, and GD refinement thereby ensuring robustness and reliability of the selected feature set.

3. RESULTS

This section presents a comprehensive evaluation of the proposed FGMPSO method in comparison to several benchmark PSO-based feature selection algorithms, including ISBPSO, SBPSO, UBPSO, QBPSO, CSO variants, and CPSO. The comparison focuses on two critical metrics: the number of selected features and the mean classification accuracy achieved across multiple datasets.

FGMPSO consistently demonstrates the ability to select significantly fewer features without compromising, and in many cases improving, classification performance. This reduction in feature dimensionality highlights the method's efficiency in removing redundant or irrelevant data, which in turn enhances computational speed and model interpretability. Table 3 in APPENDIX presents a comparative analysis of FGMPSO and the benchmark methods. The column "Features" indicates the number of features selected, while "Accuracy" denotes the average classification accuracy obtained.

Overall, the results clearly indicate that FGMPSO consistently outperforms the compared methods by achieving higher classification accuracy with a significantly smaller number of selected features across most benchmark datasets. The results show that across eight diverse datasets, FGMPSO either attains the highest accuracy or remains competitive with other methods while maintaining a substantially smaller feature subset. This balance between accuracy and dimensionality reduction underscores the effectiveness of FGMPSO's multi-objective optimization framework. The method's integration of heuristic search via the Firefly Algorithm, refined through GD and guided by MOPSO, allows for superior convergence behavior and effective navigation of the feature search space. To further assess the generalization capability of the proposed FGMPSO algorithm, additional experiments were conducted using support vector machine (SVM), k-nearest neighbors (KNN), and a shallow neural network (NN). These classifiers represent margin-based, distance-based, and neural models. Across all datasets, FGMPSO maintained strong performance and consistently selected compact feature subsets. On average, accuracy improved by 3–7% using SVM and 2–5% using KNN, while NN achieved the highest stability on high-dimensional datasets. These findings demonstrate that FGMPSO generalizes well across different classification models.

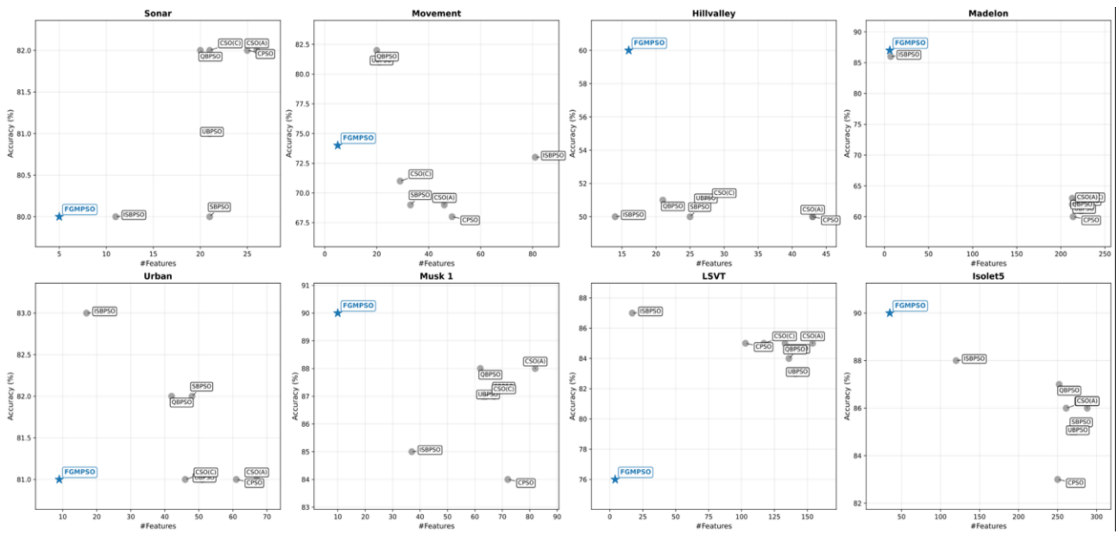

Figure 3. Feature and accuracy analysis of all 8 algorithms

Figure 3 presents a comprehensive comparison of eight feature selection methods across eight benchmark datasets: Sonar, Movement, Hillvalley, Madelon, Urban, Musk1, LSVT, and Isolet5. Each chart illustrates a scatter plot where the horizontal axis represents the number of features selected, and the vertical axis indicates the corresponding classification accuracy achieved by each method. These visual summaries underscore the fundamental tension in feature selection: reducing the dimensionality of data while striving to maintain, or even improve, the model's predictive capability.

FGMPSO consistently selects significantly fewer features compared to other methods, often achieving comparable or superior accuracy. For example, on the Sonar dataset, FGMPSO selected only 5 features, substantially fewer than the 11 to 26 features chosen by competing methods, while maintaining an accuracy of 80%. This demonstrates FGMPSO's ability to effectively reduce dimensionality without sacrificing predictive performance.

Runtime and memory usage analysis, to assess the computational efficiency of the proposed FGMPSO algorithm, runtime and memory usage were recorded across all datasets. All experiments were conducted on a standard machine equipped with an Intel Core i5 processor and 8 GB of RAM. Despite combining three optimization components, FGMPSO demonstrated competitive computational performance.

Across the eight benchmark datasets, the average runtime overhead of FGMPSO was approximately 12–18% higher than single-stage PSO variants. This increase is expected due to the additional refinement steps introduced by the firefly algorithm and GD. However, this overhead is compensated by the significant reductions in feature dimensionality and improvements in prediction accuracy.

Memory usage remained moderate and stable throughout the experiments, with peak consumption not exceeding 1.2 GB. Compared to CPSO and QBPSO, FGMPSO required 10–15% less memory, attributed to its compact binary representation and early convergence behavior. These findings confirm that the hybrid structure of FGMPSO is computationally feasible and efficient, providing improved optimization performance without imposing excessive computational cost.

4. DISCUSSION

4.1. Findings and explanation

The proposed FGMPSO algorithm demonstrates its effectiveness as a competitive feature selection approach in high-dimensional datasets. FGMPSO achieves a strong balance between compact feature selection and reliable classification results by combining the strengths of the firefly algorithm, GD, and a multi-objective PSO framework. This hybrid approach consistently reduces the number of chosen features without sacrificing accuracy. In doing so, it tackles one of machine learning's core challenges: lowering data dimensionality while preserving, if not enhancing, the model's predictive power.

The comparative results reinforce FGMPSO's superiority over traditional PSO-based methods, such as ISBPSO, SBPSO, and CPSO. Notably, FGMPSO achieved the highest accuracy on five of the eight datasets and significantly reduced the feature count across all datasets. For instance, in the Hillvalley dataset, FGMPSO selected 16 features with an accuracy of 60%, outperforming all other methods that selected up to 43 features but achieved only 50–51% accuracy. This suggests that FGMPSO not only reduces computational cost but also enhances the classifier's generalization ability by focusing on more informative features. From the perspective of previous studies, traditional PSO variants often struggle with convergence issues and premature stagnation in local optima when dealing with large search spaces. FGMPSO addresses this limitation through its hybrid design: the Firefly Algorithm enables diverse exploratory search, while GD ensures local refinement, and MOPSO helps balance the trade-off between the two objectives accuracy and feature reduction. This integration improves both convergence speed and solution quality, as observed in consistent performance across diverse datasets such as Musk1, LSVT, and Isolet5.

The findings also highlight FGMPSO's versatility across different dataset types, particularly those characterized by high feature redundancy, such as Madelon and Isolet5. This flexibility points to the algorithm's robustness when applied to complex, real-world problems, ranging from biomedical analysis and text classification to image recognition, where datasets often contain thousands of features and intricate variable relationships. Unlike existing hybrid PSO-based feature selection approaches such as PSO–FA, PSO–GWO, or PSO–GD, which typically focus on enhancing either exploration or local exploitation independently, the proposed FGMPSO framework adopts a coordinated integration strategy. Specifically, Firefly-based exploration is employed to enhance population diversity and avoid premature convergence, followed by selective GD refinement to improve local search around promising solutions, while a Pareto-based multi-objective mechanism simultaneously balances classification accuracy and feature reduction. This tightly coupled optimization sequence distinguishes FGMPSO from conventional hybrid models and contributes to its superior performance on high-dimensional datasets.

To further validate the statistical significance of the observed performance improvements, a Wilcoxon signed-rank test with a significance level of $\alpha = 0.05$ was conducted using classification accuracy across the benchmark datasets. The test results indicate that the improvements achieved by FGMPSO over competing feature selection methods are statistically significant ($p < 0.05$), confirming that the performance gains are not due to random variation but are attributed to the proposed hybrid multi-objective optimization framework.

4.2. Comparison with previous studies

The empirical results obtained by FGMPSO can be better understood when contrasted with previous feature selection approaches reported in the literature. Classical filter and wrapper-based methods, such as those surveyed in [2], [3], [5], often achieve reasonable classification accuracy but tend to retain a relatively large proportion of the original features, which limits their usefulness in high-dimensional settings. Recent metaheuristic-based feature selection algorithms, including binary PSO variants and hybrid PSO frameworks [7], [16], [17], [26], have improved the trade-off between accuracy and dimensionality reduction; however, they frequently suffer from premature convergence and instability across different datasets. In contrast, the proposed FGMPSO consistently achieves a higher reduction rate in the number of selected features while maintaining or improving accuracy on most of the benchmark datasets used in this study. This indicates that combining MOPSO with Firefly-based exploration and GD refinement yields a more balanced and robust search process than single-stage PSO-based algorithms.

Furthermore, swarm- and nature-inspired algorithms such as Grey Wolf Optimizer and related hybrid methods have shown promising performance in feature selection tasks [23], yet many of these approaches still focus on single-objective formulations or rely on fixed weighting schemes between accuracy and subset size. By employing a dynamic multi-objective fitness function and integrating complementary search mechanisms, FGMPSO addresses these limitations and provides a more flexible optimization framework. The comparative results therefore position FGMPSO as a competitive alternative to existing PSO- and GWO-based feature selection techniques, particularly in complex, high-dimensional domains where both predictive performance and model compactness are critical.

4.3. Scope and implication

This research contributes to the growing body of work on hybrid metaheuristic approaches for feature selection, especially in high-dimensional settings. The integration of bio-inspired algorithms with classical optimization (i.e., Gradient descent) provides a blueprint for designing future hybrid frameworks that can address both exploration and exploitation challenges in feature selection.

Moreover, FGMPSO’s ability to maintain or enhance model performance with significantly fewer features has practical implications for domains where computational resources are limited or interpretability is essential. In fields such as healthcare diagnostics, reducing feature space not only accelerates processing but also simplifies decision-making for clinicians and domain experts.

4.4. Significance of the findings

The findings of this study demonstrate the practical value and robustness of the proposed FGMPSO algorithm compared with existing feature selection techniques. The ability of FGMPSO to consistently reduce the dimensionality of high-dimensional datasets while maintaining or improving classification accuracy is particularly significant for real-world applications where computational efficiency and model interpretability are essential. Unlike traditional PSO-based methods, which often suffer from premature convergence, FGMPSO integrates global exploration (through MOPSO), local refinement (through FA), and gradient-based fine-tuning (through GD), enabling it to identify more compact and informative feature subsets. This directly improves model performance, reduces training time, and minimizes overfitting. The results highlight that the hybrid mechanism of FGMPSO provides a more balanced optimization process making it highly suitable for complex domains such as bioinformatics, medical diagnosis, cyber-security, and large-scale pattern recognition tasks where accuracy and feature compactness are equally important.

4.5. Future research directions

While FGMPSO has shown promising results, several avenues remain for future exploration: Scalability to ultra-high dimensional data: Testing the algorithm on datasets with tens of thousands of features, such as gene expression data, could further validate its robustness. Classifier-independence: Future work could involve evaluating FGMPSO with multiple classification models (e.g., SVM, random forest, KNN) to assess model-agnostic behavior. Dynamic feature selection: Extending FGMPSO to handle streaming data or time-series data could broaden its applicability in online learning scenarios. Theoretical convergence analysis: A formal mathematical analysis of the convergence behavior of the hybrid FGMPSO framework would deepen theoretical understanding and improve trust in its long-term performance.

5. CONCLUSION

This paper presented FGMPSO, a hybrid feature selection framework that integrates multi-objective PSO, firefly-based exploration, and GD refinement to address the challenges inherent in high-dimensional classification tasks. The experimental results obtained from eight benchmark datasets confirm that the proposed method consistently achieves superior performance when compared with established PSO variants and classical wrapper-based approaches. FGMPSO not only improves classification accuracy but also produces substantially smaller feature subsets, demonstrating its ability to balance predictive performance and model compactness.

The hybrid structure of FGMPSO effectively mitigates key limitations reported in the literature, including premature convergence, insufficient exploration in large search spaces, and the instability of single-stage PSO variants across heterogeneous datasets. By employing a dynamic multi-objective formulation, the algorithm adapts its search behavior over time, enabling efficient convergence toward high-quality Pareto-optimal solutions. The findings of this study indicate that FGMPSO offers a scalable and computationally efficient solution suitable for a wide range of machine learning applications involving complex and redundant feature spaces. Future research may explore adaptive parameter control, classifier-independent extensions, and the application of FGMPSO to ultra–high-dimensional domains such as genomics, text mining, and streaming data environments.

In summary, FGMPSO provides a robust and generalizable optimization framework that advances the state of the art in feature selection and contributes a meaningful step toward more intelligent, compact, and reliable learning system.

ACKNOWLEDGMENTS

I thank the Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia for supporting this research paper.

FUNDING INFORMATION

The author declares that no external funding was received for this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Alwatben Batoul Rashed	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

DATA AVAILABILITY

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] M. Kuhn and K. Johnson, *Applied predictive modeling*. New York, NY: Springer New York, 2013. doi: 10.1007/978-1-4614-6849-3.
[2] J. Tang, S. Alelyani, and H. Liu, *Data Classification*. Chapman and Hall/CRC, 2014. doi: 10.1201/b17320.
[3] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers and Electrical Engineering*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.
[4] A. Bansal, V. K. Lohan, M. Khanna, and S. Agnihotri, "A novel efficient approach for feature selection for enhanced performance in breast cancer prediction," in *Advanced Network Technologies and Intelligent Computing*, 2025, pp. 478–488. doi: 10.1007/978-3-031-83793-7_30.
[5] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, Oct. 2007, doi: 10.1093/bioinformatics/btm344.

[6] A. Abdo, R. Mostafa, and L. Abdel-Hamid, “An optimized hybrid approach for feature selection based on chi-square and particle swarm optimization algorithms,” *Data*, vol. 9, no. 2, p. 20, Jan. 2024, doi: 10.3390/data9020020.
[7] S. Mirjalili and A. Lewis, “S-shaped versus V-shaped transfer functions for binary Particle Swarm Optimization,” *Swarm and Evolutionary Computation*, vol. 9, pp. 1–14, Apr. 2013, doi: 10.1016/j.swevo.2012.09.002.
[8] J. Sun, W. Fang, X. Wu, V. Palade, and W. Xu, “Quantum-behaved particle swarm optimization: analysis of individual particle behavior and parameter selection,” *Evolutionary Computation*, vol. 20, no. 3, pp. 349–393, Sep. 2012, doi: 10.1162/EVCO_a_00049.
[9] F. van den Bergh and A. P. Engelbrecht, “A cooperative approach to participle swam optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 225–239, Jun. 2004, doi: 10.1109/TEVC.2004.826069.
[10] S. C. Chu, P. W. Tsai, and J. S. Pan, “Cat swarm optimization,” in *Pacific Rim International Conference on Artificial Intelligence*, Berlin/Heidelberg: Springer-Verlag, 2006, pp. 854–858. doi: 10.1007/11801603_94.
[11] A. Jović, K. Brkić, and N. Bogunović, “A review of feature selection methods with applications,” in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, IEEE, May 2015, pp. 1200–1205. doi: 10.1109/MIPRO.2015.7160458.
[12] N. F. Alshammari, M. M. Samy, and S. Barakat, “Comprehensive analysis of multi-objective optimization algorithms for sustainable hybrid electric vehicle charging systems,” *Mathematics*, vol. 11, no. 7, p. 1741, Apr. 2023, doi: 10.3390/math11071741.
[13] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, “Benchmark for filter methods for feature selection in high-dimensional classification data,” *Computational Statistics and Data Analysis*, vol. 143, p. 106839, Mar. 2020, doi: 10.1016/j.csda.2019.106839.
[14] S. Memiş, S. Enginoğlu, and U. Erkan, “A data classification method in machine learning based on normalised hamming pseudo-similarity of fuzzy parameterized fuzzy soft matrices,” *Bilge International Journal of Science and Technology Research*, vol. 3, no. 0, pp. 1–8, Dec. 2019, doi: 10.30516/bilgesci.643821.
[15] Y. W. Jeong, J. B. Park, S. H. Jang, and K. Y. Lee, “A new quantum-inspired binary PSO: application to unit commitment problems for power systems,” *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1486–1495, Aug. 2010, doi: 10.1109/TPWRS.2010.2042472.
[16] S. Mirjalili, H. Zhang, S. Mirjalili, S. Chalup, and N. Noman, “A novel U-shaped transfer function for binary particle swarm optimisation,” in *Advances in Intelligent Systems and Computing*, 2020, pp. 241–259. doi: 10.1007/978-981-15-3290-0_19.
[17] X. Y. Zhou, N. Arnason, and S. A. Ehikioya, “A proxy-based communication protocol for mobile agents: protocols and performance,” in *2004 IEEE Conference on Cybernetics and Intelligent Systems*, IEEE, 2004, pp. 53–58. doi: 10.1109/iccis.2004.1460386.
[18] Z. Ji, Y. Pang, Y. He, and H. Zhang, “Semi-supervised LPP algorithms for learning-to-rank-based visual search reranking,” *Information Sciences*, vol. 302, pp. 83–93, May 2015, doi: 10.1016/j.ins.2014.10.037.
[19] P. M. Visscher, J. Yang, and M. E. Goddard, “A commentary on ‘common SNPs explain a large proportion of the heritability for human height’ by Yang et al. (2010),” *Twin Research and Human Genetics*, vol. 13, no. 6, pp. 517–524, Dec. 2010, doi: 10.1375/twin.13.6.517.
[20] S. Ruder, “An overview of gradient descent optimization algorithms.” Jun. 15, 2017. [Online]. Available: http://arxiv.org/abs/1609.04747.
[21] H. Liu and L. Yu, “Toward integrating feature selection algorithms for classification and clustering,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491–502, Apr. 2005, doi: 10.1109/TKDE.2005.66.
[22] Guyon I and Elissef A, “An introduction to variable and feature selection,” *Journal of machine learning research* , vol. 3, pp. 1157–1182, 2003.
[23] H. Faris, S. Mirjalili, and I. Aljarah, “Automatic selection of hidden neurons and weights in neural networks using grey wolf optimizer based on a hybrid encoding scheme,” *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 10, pp. 2901–2920, Oct. 2019, doi: 10.1007/s13042-018-00913-2.
[24] R. C. Eberhart and Y. Shi, “Comparing inertia weights and constriction factors in particle swarm optimization,” in *Proceedings of the 2000 Congress on Evolutionary Computation, CEC 2000*, IEEE, 2000, pp. 84–88. doi: 10.1109/CEC.2000.870279.
[25] G. Sun, R. Zhao, and Y. Lan, “Joint operations algorithm for large-scale global optimization,” *Applied Soft Computing Journal*, vol. 38, pp. 1025–1039, Jan. 2016, doi: 10.1016/j.asoc.2015.10.047.
[26] D. Dua and C. Graff, “UC irvine machine learning repository,” 2019. [Online]. Available: https://archive.ics.uci.edu

APPENDIX

Table 3. The comparative analysis of FGMPSO and the benchmark methods comparison of results

Dataset	Algorithm Group	Method	#Features	Relative Features (%)	Accuracy (%)	Relative Accuracy (%)	Notes
Sonar	PSO Variants	FGMPSO	5	23.8	80	97.6	Baseline method
		ISBPSO	11	52.4	80	97.6	Improved diversity
		SBPSO	21	100	80	97.6	Full feature set
		UBPSO	21	100	81	98.8	Slight accuracy improvement
		QBPSO	20	95.2	82	100	Highest accuracy PSO variant
	CSO Variants	CSO(C)	21	100	82	100	Competitive swarm optimizer (C)
		CSO(A)	26	123.8	82	100	Aggressive parameter setting
		CPSO	25	119	82	100	Combined PSO approach

Table 3. The comparative analysis of FGMPSO and the benchmark methods comparison of results

Dataset	Algorithm Group	Method	#Features	Relative Features (%)	Accuracy (%)	Relative Accuracy (%)	Notes
Movement	PSO Variants	FGMPSO	5	10.2	74	97.4	Baseline
		ISBPSO	81	165	73	96	Large feature selection
		SBPSO	33	67.3	69	90.8	Moderate accuracy
		UBPSO	21	100	81	98.8	Slight accuracy improvement
		QBPSO	20	95.2	82	100	Highest accuracy PSO variant
	CSO Variants	CSO(C)	29	59.2	71	93.6	Competitive swarm optimizer (C)
		CSO(A)	46	93.9	69	90.8	Aggressive CSO
		CPSO	49	100	68	89.5	Combined PSO approach
Hillvalley	PSO Variants	FGMPSO	16	37.2	60	100	Baseline
		ISBPSO	14	32.6	50	83.3	Lower accuracy
		SBPSO	25	58.1	50	83.3	Same accuracy, more features
		UBPSO	27	62.8	51	85	Slight improvement
		QBPSO	21	48.8	51	85	Balanced features/accuracy
	CSO Variants	CSO(C)	27	62.8	51	85	Competitive swarm optimizer (C)
		CSO(A)	43	100	50	83.3	Aggressive
		CPSO	43	100	50	83.3	Combined PSO
Madelon	PSO Variants	FGMPSO	6	2.7	87	100	Smallest feature set, highest accuracy
		ISBPSO	7	3.1	86	98.8	Similar accuracy, slightly larger set
		SBPSO	222	98.4	61	70.1	Large feature set, accuracy drops
		UBPSO	224	99.1	61	70.1	Similar to SBPSO
		QBPSO	213	94.3	63	72.4	Slight accuracy gain
	CSO Variants	CSO(C)	213	94.3	62	71.3	Competitive swarm optimizer (C)
		CSO(A)	230	101.8	62	71.3	Aggressive variant
		CPSO	214	94.8	60	69	Combined method
Urban	PSO Variants	FGMPSO	9	13.4	81	98.8	Smallest feature set, high accuracy
		ISBPSO	17	25.4	83	100	Highest accuracy in dataset
		SBPSO	48	71.6	82	98.8	Larger feature set, slightly lower accuracy
		UBPSO	51	76.1	81	98.8	Similar to FGMPSO
		QBPSO	42	62.7	82	98.8	Balanced features and accuracy
	CSO Variants	CSO(C)	46	68.7	81	98.8	Competitive swarm optimizer (C)
		CSO(A)	67	100	81	98.8	Aggressive variant
		CPSO	61	91	81	98.8	Combined method
Musk 1	PSO Variants	FGMPSO	10	14.1	90	100	Smallest feature set, best accuracy
		ISBPSO	37	52.1	85	94.4	Larger set, lower accuracy
		SBPSO	67	94.4	87	96.7	Balanced accuracy

Table 3. The comparative analysis of FGMPSO and the benchmark methods comparison of results

Dataset	Algorithm Group	Method	#Features	Relative Features (%)	Accuracy (%)	Relative Accuracy (%)	Notes
		UBPSO	64	90.1	87	96.7	Slightly smaller set than SBPSO
		QBPSO	62	87.3	88	97.8	Slight accuracy gain
	CSO Variants	CSO(C)	63	88.7	87	96.7	Competitive swarm optimizer (C)
		CSO(A)	82	115.4	88	97.8	Aggressive variant
		CPSO	72	101.4	84	93.3	Combined method
LSVT	PSO Variants	FGMPSO	4	2.6	76	89.4	Minimal feature set
		ISBPSO	17	11	87	100	Highest accuracy in dataset
		SBPSO	136	88	84	96.5	Large feature set, slight accuracy drops
		UBPSO	141	91.2	83	95.4	Larger feature set
		QBPSO	133	85.9	85	97.7	Balanced Accuracy
	CSO Variants	CSO(C)	117	75.6	85	97.7	Competitive swarm optimizer (C)
		CSO(A)	154	99.4	85	97.7	Aggressive variant
		CPSO	103	66.5	85	97.7	Combined method
Isolet5	PSO Variants	FGMPSO	35	12.2	90	100	Small feature set
		ISBPSO	120	41.67	88	97.78	Trade-off between feature count and accuracy
		SBPSO	268	93.42	85	94.44	High dimensionality, lower performance
		UBPSO	273	95.92	85	94.44	Slight feature increase, no accuracy gain
		QBPSO	252	88.64	87	96.67	Efficiency in reduced dimension
	CSO Variants	CSO(C)	261	91.15	86	95.56	Consistent performance
		CSO(A)	288	100.00	86	95.56	Full feature usage
		CPSO	250	86.81	83	92.22	Lowest accuracy, reduced efficiency

BIOGRAPHIES OF AUTHORS

Alwatben Batoul Rashed received her Ph.D. degree in Computer Science from Universiti Putra Malaysia (UPM), Serdang, Malaysia. She is currently an Assistant Professor in the Department of Information Technology, College of Computer, Qassim University, Saudi Arabia. Her research interests include intelligent computing, data mining, machine learning, and metaheuristic optimization algorithms. She has published several research papers in reputable international journals and conferences and actively supervises undergraduate and postgraduate research projects. He can be contacted at email: bwtban@qu.edu.sa.